Alteration of Intensity Resolution and Loudness by Sounds Presented in Temporal Proximity

vorgelegt von

Diplom-Psychologe
Daniel Oberfeld-Twistel

Von der Fakultät V - Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Philosophie
- Dr. phil. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. H. Jungermann
Berichter:	Prof. Dr. M. Thüring
Berichter:	Dr. R. P. Carlyon

Tag der wissenschaftlichen Aussprache: 6. Mai 2005

Berlin 2005
D 83

Abstract

A sound presented temporally proximal to a target sound has effects on intensity resolution and loudness. The aim of this work was to develop a model for three phenomena. First, an intense non-simultaneous masker causes a large increase in intensity difference limens for a midlevel standard, but has only a small effect at low and high standard levels. Second, a non-simultaneous masker introduces a systematic shift in target loudness towards masker loudness. For example, loudness of the target is enhanced by a masker higher in level than the target. Third, in rhythmic equitone sequences constructed as pairs of sounds separated by a longer interval, an accent is perceived on the second sounds of the pairs, i.e., loudness of the second sounds exceeds loudness of the first sounds.

Previous studies indicated a close relation between the masker-induced changes in loudness and intensity resolution. The model formulated in this thesis attributes the empirical similarities between the first two phenomena to a common mechanism. Above that, the question was studied whether the interval-produced accent can be understood as loudness of the second tone of a pair being enhanced by the first tone.

Based on a review of data and models, a new model for loudness and intensity resolution in non-simultaneous masking was developed, termed the Similarity Model. It assumes that the memory representation of target loudness used in a loudness matching task is a weighted average between target and masker loudness [Elmasian, Galambos, & Bernheim (1980), J. Acoust. Soc. Am. 67, 601-607]. The important new feature of the model is an effect of the target-distractor similarity. The weight assigned to masker loudness is assumed to be a function of the perceptual similarity between masker and target, decreasing with the perceptual difference between the two sounds. Based on an idea by Carlyon and Beveridge (1993, J. Acoust. Soc. Am. 93, 2886-2895) the model predicts the elevation of the just-noticeable difference (jnd) to be monotonically related to the loudness change caused by a non-simultaneous masker.

Unlike previous models, the similarity model predicts a mid-difference hump, that is, at each target level the effects of the masker are expected to be most pronounced at intermediate masker-target level differences. The model was tested in three experiments. In Experiment 1, intensity difference limens were measured for 25, 55, and 85-dB SPL standards, while the masker-target level difference was varied across a range of −60 to +60 dB. Compatible with the predictions of the similarity model, listeners produced a mid-difference hump and the masker caused significant jnd elevations at the lowest and the highest standard level. Previous models predict no or only very small effects of the masker at low and high levels, and at midlevels a monotonic increase of the jnd with masker level. The data are incompatible with these models. In Experiment 2, loudness matches were obtained in the same conditions as in Experiment 1 and for the same listeners. Mid-difference humps were observed again. The masker-induced loudness change was smallest at the lowest target level. In Experiment 3, similarity on the duration dimension was varied. For masker and target additionally differing in duration, the similarity model predicts the same pattern (i.e., a mid-difference hump), but a generally smaller effect of the masker. Results were compatible with this prediction.

According to the similarity model, the common mechanism underlying both loudness enhancement and the jnd elevation is the inclusion of masker intensity information in the decisions made in the respective experiments. Compatible with this assumption, the correlation between loudness enhancement and the difference limen was significant. The correlation was rather weak, however, partly because loudness decrement corresponded to small jnd's.

It was demonstrated that the similarity model can quantitatively predict the loudness change caused by a non-simultaneous masker. The mathematical modelling combines properties of the auditory periphery, represented by the loudness function for pure tones, with the presumably more centrally located influence of masker loudness on the memory representation of target loudness. With only two free parameters, the model produced reasonable to excellent fits to previous data sets and to the data from Experiment 2. The smaller amounts of loudness enhancement found at the lowest target level were correctly predicted, due to the steeper slope of the loudness function at low levels. It is important to note, however, that recently, Scharf, Buus, and Nieder (2002) suggested that the loudness matches obtained in a three-tone matching task reflect "loudness recalibration" (that is, a reduction in loudness, cf. Marks, 1994) of the comparison tone rather than loudness enhancement of the target tone. Supporting evidence was provided by Arieh and Marks (2003). It remains to be shown whether the entire set of "loudness enhancement" effects can be explained by loudness recalibation. The strong dependence of loudness matches on the masker-target level difference found in Experiment 2 of this work and the observation of loudness decrement seem incompatible with the loudness recalibration data.

To model the effects the inclusion of masker intensity information has on discrimination performance, a signal detection approach was used. The predicted patterns were not compatible with the data, however.

To answer the question of whether the accent can be attributed to loudness enhancement, Experiment 4 provided an estimate of the temporal limit of the accent. The accent was observed for intra-pair intervals as long as 800 ms which is longer than the 400-ms masker-target interval reported to represent the temporal limit of loudness enhancement. In Experiment 5, presentation level was varied. If loudness enhancement were the cause of the accent, it should be less pronounced at the lowest target level, just as loudness enhancement in Experiment 2. Contrary to this prediction, an accent was present at low and intermediate levels, but virtually absent at the highest level.

It was demonstrated that the alteration of intensity resolution and loudness caused by a non-simultaneous masker can be attributed to a common mechanism. There was no evidence for the hypothesis that the accent represents a special case of loudness enhancement.

Danksagung

Ich danke Prof. Dr. Manfred Thüring für die Betreuung meiner Arbeit, für wichtige Anregungen und Kommentare und für seine Hilfe auch in organisatorischen Problemen.

Prof. Dr. Helmut Jungermann danke ich für seine stete und großzügige Unterstützung während meiner Zeit an der Technischen Universität Berlin, Rat in ‚strategischen' Fragen, hilfreiche Anmerkungen zur Struktur der Arbeit und sein in mich gesetztes Vertrauen.

Dr. Robert Carlyon danke ich für das Interesse an meiner Arbeit, wertvolle Anregungen und die Übernahme des Gutachtens.

Bei allen Kolleginnen und Kollegen der Abteilung Allgemeine Psychologie an der TU Berlin möchte ich mich für die schöne Zeit, die angenehme Arbeitsatmosphäre und die freundschaftliche Unterstützung bedanken.

Wolfram Rollett war immer bereit, über ‚denkwürdige' Modelle zu diskutieren, ich bedanke mich für viele Anregungen.

Prof. Dr. Stefanie Kelter danke ich für ihr Interesse an meiner Arbeit und für viele anregende und ermutigende Diskussionen.

Bei PD Dr. Carsten Henkel möchte ich mich für kritische Anmerkungen zum mathematischen Teil der Arbeit bedanken.

Ich bedanke mich bei Prof. Dr. Heiko Hecht und den Kolleginnen und Kollegen an der Universität Mainz, die mich während der Abschlussphase meiner Arbeit unterstützt haben.

Dem Institut für Technische Akustik der TU Berlin, besonders Kai Johannsen, danke ich für die technische Unterstützung.

Schließlich danke ich besonders meiner Frau Gabi Twistel, die mich während der oft anstrengenden Diss-Phase ermutigt und unterstützt hat.

Contents

1 General Introduction

We constantly perceive and judge the loudness of auditory events and use it as a cue for a wide variety of behaviors such as judging the distance of a car approaching as we want to cross the road (cf. Button and Davids, 2004), grasping the meaning of a spoken sentence, or dancing to the beat of a pop song. In all of these situations, our auditory system provides us with sufficient intensity resolution over a surprisingly large dynamic range.

From the early days of psychophysical and psychological experimentation and theorizing, as evident in the work of Fechner (1860) and Helmholtz (1863), the question of how the physical characteristics of a sound relate to its perceived intensity (loudness) and its discriminability has received considerable interest. Many early experiments and theories were restricted to perception and processing of sounds presented in isolation, however (e.g., Stevens, 1955).

Now maybe some of us spend far too much time of their lives in sound proof chambers -- but it is obvious that in everyday life, we only seldom encounter sounds "in quiet", i.e., presented in isolation and in a very background. Therefore, to gain an understanding of auditory intensity processing, it is important to have data and models about how sounds presented simultaneously or sequentially interact.

This work studies three phenomena in intensity perception and intensity resolution of auditory stimuli that can all be described as sequential effects because loudness and/or discriminability of a brief sound are altered by another sound presented in temporal proximity. In some respects, the stimuli represent just the smallest possible increase in complexity, as we move from a single sound presented in isolation to two sounds presented within a short temporal window. Nevertheless, it will become clear that the effects caused by introduction of the additional sound are rather complex, some of them are highly counterintuitive, and the formulation of models accounting for the effects is far from trivial.

All of the three phenomena discussed in this thesis are observed in situations where a brief sound is preceded or followed by another sound:

1. Zeng, Turner, and Relkin (1991) first reported the "*midlevel hump*" in non-simultaneously masked intensity discrimination. In the presence of an intense forward masker, just-noticeable differences (jnd's) were elevated relative to the jnd in quiet at intermediate standard levels. At low and high standard levels, however, the effect of the masker was very small. This pattern is clearly different from the level dependence of the jnd in quiet, which decreases monotonically with standard level (the "near miss to Weber's law"; e.g., McGill and Goldberg, 1968 a, b).

2. The second phenomenon is *loudness enhancement*. The presence of an intense tone (termed "conditioner") preceding or following a target tone causes loudness of the target tone to be larger than in quiet (e.g., Zwislocki and Sokolich, 1974). Before the background of the usual physiological and psychophysical effects of masking, one would have expected the presence of an intense forward masker to *reduce* the loudness of the following sound just as in simultaneous partial masking (e.g., Hellman and Zwislocki, 1964). Conditioners presented at a level below target tone level result in a reduction of perceived intensity (loudness decrement; e.g., Elmasian and Galambos, 1975).

3. The third phenomenon has been discussed mainly in the context of rhythm perception up to now. Vos (1977) found that in equitone sequences (where all sounds are physically identical) constructed as an alternation of short and long inter-onset-intervals (IOIs), listeners perceived an *accent* on the second sound of each pair of sounds (i.e., the sounds initiating the long IOIs). Because in music, accents are mainly produced by making a sound more intense (cf. Cooper and Meyer, 1964; Fraisse, 1982), this can be taken as evidence for rhythmic structure influencing intensity perception. In fact, in an experiment by Povel and Okkerman (1981), the "interval-produced" accent on the second sounds of the pairs was balanced by increasing the level of the first sounds by approximately 4 dB.

In Chapter 2, the three phenomena are described in greater detail.

It is evident that the settings used in the corresponding experiments (sounds, temporal configuration, and experimental paradigms) closely resemble each other (Fig. 1). Loudness and/or discriminability (intensity resolution) of a brief tone are altered by a sound presented in temporal proximity (inter-stimulus interval shorter than 1 s).

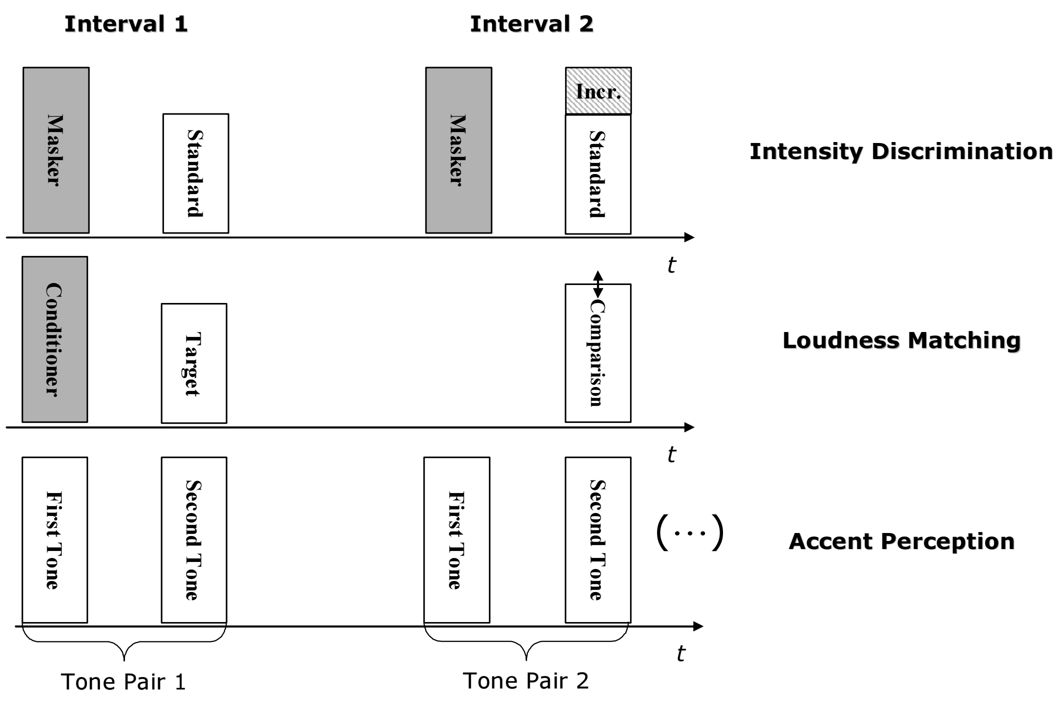

Fig. 1: Terminology used for the stimuli presented in experiments studying the three phenomena.

The objective of this work is to identify the mechanisms underlying the three phenomena. An important question arising in this context is whether the phenomena might be caused by a *common mechanism*. It has been pointed out previously that the effects of various parameters on loudness enhancement and intensity discrimination are similar (Carlyon and Beveridge, 1993; Plack, 1996 a). For example, both loudness enhancement and the elevation of intensity jnd's caused by an intense non-simultaneous masker are most pronounced at intermediate standard levels (Zeng, 1994). The multitude of similarities led Carlyon and Beveridge (1993) to suggest that "[...] *the two types of experiments are tapping the same process*" (p. 2894). Phenomenological, the interval-produced accent can be described as loudness of the second tone of a pair being enhanced by the first tone.

Consequently, the question arises whether the same mechanism that results in an intense masker enhancing loudness of a target tone also underlies the interval-produced accent.

In Chapter 3, the empirical evidence available to date is reviewed. The similarities between the phenomena support the idea of a common mechanism.

Above that, and possibly more important, the discussion of the data provides the basis for a systematic and critical review of explanations and models proposed for the three phenomena. In Chapter 4, it will become clear that the existing models can not account for the complete range of findings and that some of them are subject to serious limitations. Moreover, despite the empirical similarities, most models are restricted to only one phenomenon. The notable exception is the "loudness enhancement hypothesis" brought forward by Carlyon and Beveridge (1993). This model assumes that loudness enhancement causes loudness variability, which in turn results in increased intensity difference limens.

In Chapter 5, a new model of the effects of non-simultaneous masking on intensity discrimination and loudness is proposed. In line with the hypothesis of a common mechanism, the model assumes both loudness enhancement and the impairment of intensity resolution to be due to the inclusion of masker information at a central location of processing. The model combines the "mergence hypothesis" proposed by Elmasian, Galambos and Bernheim (1980) and the loudness enhancement hypothesis by Carlyon and Beveridge (1993). According to the mergence hypothesis, loudness enhancement and decrement are due to listeners using a weighted average between masker and target loudness in a loudness matching task. This model is extended by introducing an effect of the perceptual similarity between masker and target on the weight assigned to masker loudness. The similarity effect results in the model correctly predicting the midlevel hump in loudness enhancement. The loudness enhancement hypothesis assumes enhancement to introduce loudness variability, so that intensity resolution is impaired in situations where loudness is enhanced.

In Chapter 6, this "similarity model" is tested in experiments measuring intensity discrimination (Experiment 1) and loudness enhancement (Experiment 2) as a function of the level difference between masker and target tone. Although theoretical considerations suggest this level difference to be an important factor, especially for loudness enhancement, previous experiments did not vary the masker-target level difference independent of target level. The predictions of the similarity model concerning the effects of the masker-target level difference significantly differ from the predictions of the existing models, so that the data collected in Experiments 1 and 2 can provide a critical test. Experiment 3 studies the additional effect of using equal and different tone durations for masker and target. As the data are collected for the same stimuli and listeners in Experiments 1, 2, and 3, they can be used as a test for the correlation between loudness enhancement and difference limens predicted by the similarity model.

In the first part of Chapter 7, a *quantitative version* of the 'loudness part' of the similarity model is formulated. It predicts the effects of a non-simultaneous masker on the loudness of a pure tone. The mathematical modeling combines well-established properties of the auditory periphery, represented by the loudness function for pure tones, with the presumably more centrally located mechanism of the influence of masker loudness on the memory representation of target loudness. The capability of the quantitative model to account for previous data and for the data collected in Experiments 1-3 is discussed.

The second part of Chapter 7 discusses several approaches to formulating a signal-detection model of the effects the inclusion of masker intensity information has on performance in an intensity discrimination experiment.

The discussion of potential models of the interval-produced accent (Chapter 4) provides some arguments against the intuitive hypothesis that the accent is simply a special case of loudness enhancement. Instead, the (seriously limited) data seem to be compatible with a model proposed by Povel and Okkerman (1981), according to which processing of the first tone of a pair in auditory sensory memory is interrupted by the second tone, resulting in reduced loudness of the first tone. As only two studies on the interval-produced accent have been published, however, the empirical evidence for or against the two models is rather weak. Above that, from an accent reported on the second sounds of the pairs, it is principally impossible to infer whether the first tone *enhanced* loudness of the second tone or whether the second tone *reduced* loudness of the first tone. It is therefore necessary to examine whether, e.g., the time course or the level dependence of the accent is compatible with either loudness enhancement, or interrupted processing, or both. In this line of thinking, Chapter 8 describes two experiments designed to provide a critical test between the models. Concerning the effects of temporal structure of the rhythmic sequences on accent perception, data by Povel and Okkerman (1981) seem at first sight to indicate that the temporal limits of the interval-produced accent are incompatible with the temporal limits of loudness enhancement. Their estimate is likely to be biased, however, due to a problem in their experimental design. A preliminary experiment and Experiment 4 avoid this problem. Additionally, a quantitative model based on the concept of interrupted-processing is used to derive a more reliable estimate of the time constant from the data of Experiment 4 and the data by Povel and Okkerman (1981). Finally, in Experiment 5, the effect of presentation level on the interval-produced accent is studied. A midlevel hump in accent perception would present strong evidence for a relation between the interval-produced accent and loudness enhancement, while presentation level should have no effect according to the interrupted processing model.

2 Effects of adjacent sounds on intensity resolution and loudness

The phenomena observed in the presence of an additional sound presented in temporal proximity to the target sound are best discussed in comparison to the characteristics of intensity processing in quiet. For this reason, some of the basic findings are introduced in a short excursus, before we turn to the three phenomena.

2.1 Excursus: Basics of intensity perception

The perception of intensity (loudness) is one of the issues studied most extensively in psychoacoustics. This is not surprising if one considers that loudness is one of the most salient sensations associated with auditory sounds and is of high importance for interaction with our environment (cf. Scharf and Houtsma, 1986). Frequently, loud sounds are an indication of danger (and if only for the auditory system itself). Loudness is involved in the spatial localization of sounds, and also in the communication of meaning in speech. The alarm function of loudness might be regarded as closely related to the facts that we can not close our ears, and that there is much less adaptation in the perception of loudness than in the perception of most other stimulus dimensions (cf. Scharf, 2001).

For sounds in quiet, there is extensive evidence about the influences of sound pressure level, frequency, bandwidth, duration, and several other parameters on intensity resolution and loudness (for recent reviews cf. Scharf and Houtsma, 1986; Plack and Carlyon, 1995; Zwicker and Fastl, 1999).

2.1.1 Intensity resolution in quiet

For tones in quiet, discrimination performance is related to standard intensity by what is termed the "near-miss to Weber's law" (e.g., McGill and Goldberg, 1968 a, b; Jesteadt, Wier, and Green, 1977). At a given standard intensity I, we define ΔI_{DL} to be the 'just noticeable increment' if a comparison tone with intensity $I + \Delta I_{DL}$ is correctly identified as more intense than the standard in, e.g., 70.7% of all trials. The just-noticeable difference (jnd) decreases monotonically with standard level (cf. the 'Quiet' curve in Fig. 3). This effect can be attributed to spread of excitation on the basilar membrane (e.g., Plack and Carlyon, 1995): with increasing intensity, the regions on the basilar membrane that are excited by a pure tone become wider. Consequently, auditory neurons with characteristic frequencies outside the critical band centered at tone frequency provide additional information. This model is in very good accordance with the finding that intensity discrimination for wide-band noise follows Weber's law (e.g., Houtsma, Durlach, and Braida, 1980), as for this signal, the number of excited auditory neurons does not increase with intensity. A similar argumentation applies to the observation that jnd's obtained for high frequency tones more closely follow Weber's law than for low-frequency tones (Florentine, 1983), and that band-gap masking noise restricting off-frequency listening removes the near-miss (e.g., Moore and Raab, 1974).

The neural basis of the psychophysical data has been a continuous topic of debate. Most models assume the neural code for intensity and/or loudness to be the number of neural spikes produced by auditory nerve neurons in response to a sound. Only few models are based on temporal properties of the neural responses, like for example the interarrival times of neural spikes (Luce and Green, 1974). Efforts have been directed towards accounting for the near miss to Weber's law (e.g. McGill and Goldberg, 1968 a, b; Teich and Lachs, 1979), the extremely large dynamic range of human intensity discrimination opposed to the narrow dynamic range of single auditory nerve fibers (e.g., Delgutte, 1987; Viemeister, 1988; Winslow and Sachs, 1988), intensity discrimination under masking and for wideband stimuli (e.g., Florentine and Buus, 1981), the shape of the psychometric function (Green and Swets, 1966; Laming, 1986), or the relation between intensity discrimination and loudness (e.g., Lachs and Teich, 1981; Zwislocki and Jordan, 1986; Hellman and Hellman, 1990).

2.1.2 Loudness in quiet

While the study of intensity discrimination can rely on objective methods, loudness is, by definition, a subjective quality. It is measured by different methods and on different scales. Despite a long controversial discussion, the magnitude estimation methods promoted by Stevens (1955, 1957) are widely accepted. From experiments using such methods, we know that loudness is not a linear function of intensity. Instead, the loudness function relating sound intensity I and loudness N can be described by a power law

$$N = b\, I^k \qquad (2\text{-}1)$$

for sound pressure levels greater than 40 dB SPL (cf. Scharf, 1978).

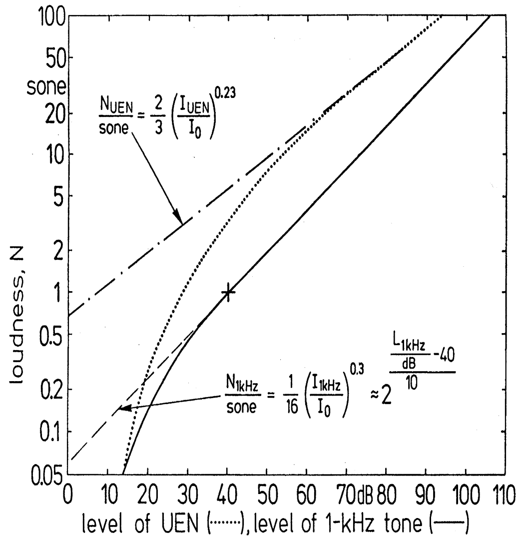

Fig. 2: Loudness in assigned numbers, as measured by magnitude estimation, for a 1-kKz pure tone (solid line) and uniform-exciting noise (dotted line). From Zwicker and Fastl (1999, p. 207).

To give an example, Hellman and Zwislocki (1963) used magnitude estimation procedures with and without a designated standard to measure the loudness of a 1-kHz tone and obtained an exponent k = 0.27. At low levels, the loudness function is steeper than at higher levels (Fig. 2). Hellman (1991) provides a review of a large number of magnitude estimation experiments.

For pure tones, Zwislocki (1965) developed a model that can account for loudness in quiet and under simultaneous masking (see below). This model will be discussed in Chapter 7.1.1. For complex stimuli, Zwicker (1958, 1963, cf. Zwicker and Fastl, 1999, p. 223ff) formulated a model based on excitation patterns. An advanced model recurring to the ideas of Zwicker was proposed by Moore, Glasberg, and Baer (1997).

2.1.3 Intensity discrimination under simultaneous masking

To prevent effects of beating present in tone-on-tone masking, intensity discrimination under simultaneous masking was studied using noise maskers in most experiments. Results from Harris (1963) and Jesteadt, Wier, and Green (1977) suggest that the effect of a masker on intensity discrimination is negligible if the standard is more than 5-10 dB above its masked threshold. Different results were reported by Widin, Viemeister, and Bacon (1986), who measured difference limens for 1-kHz tones of 20 ms duration centered temporally in a 26-ms wide-band noise masker (low-pass filtered at 10 kHz). Pedestal level was 85 dB SPL. Jnd's were slightly elevated even for pedestal levels 30 dB above detection threshold.

Only masker energy at frequencies at or above test tone frequency has a significant effect (Viemeister, 1972; Moore and Raab, 1974), presumably because of mechanical properties of the basilar membrane that result in a limited downward spread of masking but a large upward spread of masking (see Moore, 1995, for a review). The data indicate that if masker level is kept constant relative to pedestal level, the jnd's observed under masking of higher frequencies follow Weber's law. Florentine and Buus (1981) were able to account for these results by a multiband excitation-pattern model in which intensity resolution is represented by an optimum decision rule based on excitation differences from 24 critical bands.

2.1.4 Loudness under simultaneous masking

Intensity perception (loudness) in the presence of simultaneous masking has also been studied extensively (cf. Scharf and Houtsma, 1986). A masker not only produces an elevation of the detection threshold at frequencies present in the spectrum of the masker, but also alters the perception of supra-threshold stimuli. The loudness function becomes steeper than in quiet at levels near masked threshold, while it is nearly unchanged at high levels. The result of this so-called 'loudness recruitment' is that the loudness of sounds at levels near masked threshold is strongly reduced, while masked and unmasked loudness become virtually identical at levels about 30 dB above masked threshold (e.g., Hellman and Zwislocki, 1964).

2.2 Phenomenon 1: Intensity discrimination under non-simultaneous masking

In the last decade, there has been considerable interest in phenomena emerging under **non-simultaneous** masking. Zeng et al. (1991) were the first to report what they termed the “midlevel-hump in intensity discrimination”. In experiments using brief sounds and forward masking by an intense narrow-band noise burst, just-noticeable differences (measured as $\Delta L_{DL} = 10 \cdot \log_{10}[1 + \Delta I_{DL}/I]$, where ΔI_{DL} is the intensity of the just-noticeable increment and I is pedestal intensity) were elevated at intermediate standard levels (squares in Fig. 3). This pattern differs strongly from the ‘near-miss to Weber’s law’ observed in quiet. The pronounced effects of the masker are also somewhat surprising, as the neuronal representation of standard in the auditory periphery can be assumed to unaffected by the masker at the 100-ms inter-stimulus-interval (ISI) used in most studies (Kiang et al., 1965).

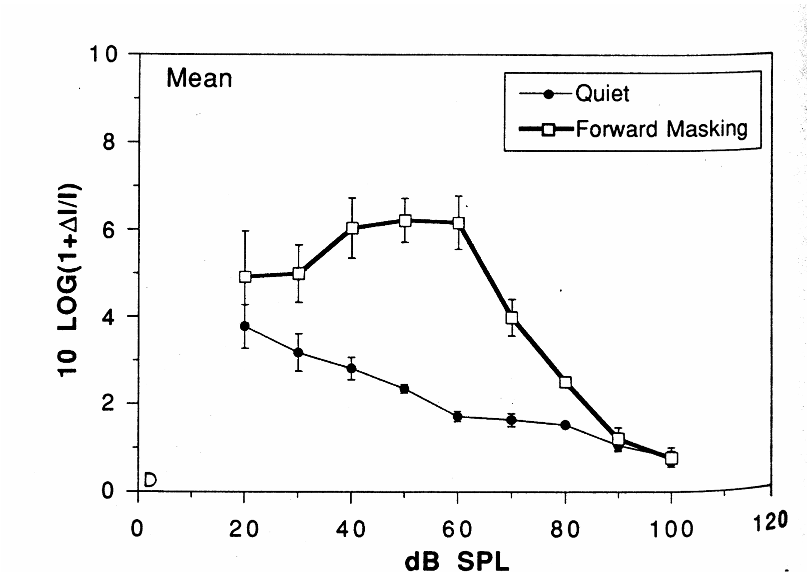

Fig. 3: Mean intensity jnd's obtained by Zeng, Turner, and Relkin (1991) for 25-ms 1-kHz tone bursts in quiet (circles) and in the presence of a 90-dB SPL forward masker (squares).

Fig. 4 depicts the stimulus configuration used in most modern intensity discrimination experiments

To give an example, Schlauch, Lanthier, and Neve (1997) presented 1-kHz pure tones of 10 ms duration. Forward masker level was 90 dB SPL; the standard followed after a silent interval of 100 ms and was presented at 20-80 dB SPL. The inter-standard offset-onset interval was 850 ms. A two-interval, two alternatives forced-choice (2I, 2AFC) adaptive procedure (Levitt, 1971) was used. In one of two observation intervals (selected randomly), an increment was added to the pedestal. Listeners decided whether the increment had been presented in the first or in the second interval. In the first trial, the increment was intense enough to be detected. After three consecutive correct responses, the increment was decreased by 3 dB. After each incorrect response, it was increased by 3 dB (3-down, 1-up rule). This procedure tracks the 79.4%-correct point on the psychometric function.

A considerable number of studies conducted subsequently provided further data on the midlevel hump in intensity discrimination. These findings are reviewed in Chapter 3.

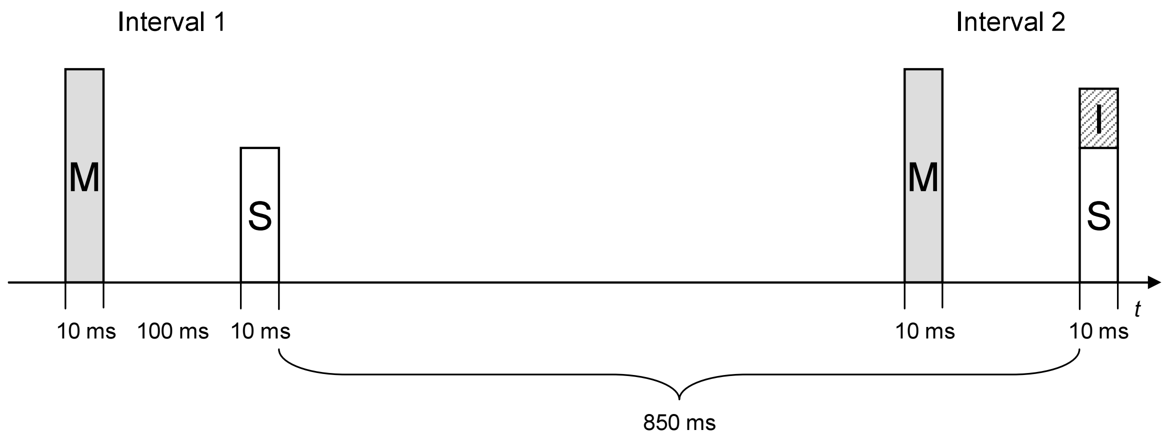

Fig. 4: Example trial configuration used in the two-interval, two-alternative forced-choice (2I, 2AFC) adaptive procedure for intensity discrimination (e.g., Schlauch, Lanthier, and Neve, 1997). In each of the two observation intervals, a forward masker M and the standard S are presented. The intensity increment I is presented in interval 1 or interval 2 with equal a priori probability. The listener responds whether the increment occurred in the first or in the second interval. After three consecutive correct responses, the increment is decreased. After each incorrect response, the increment is increased.

A midlevel-hump has also been found at high frequencies and at short tone durations even without non-simultaneous masking. Carlyon and Moore (1984; 1986 a, b) reported a hump for the detection of tones in noise and for intensity discrimination in quiet at high pedestal frequencies and termed these effects the "severe departure from Weber's law". Carlyon and Beveridge (1993) measured intensity discrimination and detection of tones in noise within the same experiment and reported small humps in quiet for 1-kHz pedestals and larger humps for 6.5-kHz pedestals. A small hump was also present for 6-kHz pedestals in quiet in an experiment by Plack and Viemeister (1992a). Raab and Taub (1969) studied intensity discrimination between clicks (very brief pulses) and found Weber fractions ($\Delta I_{DL}/I$) to be larger at intensities between 30 and 70 dB than at low and high intensities. Nizami, Reimer, and Jesteadt (2001) observed jnd's for short-duration Gaussian-shaped tone pips at 2 kHz to be elevated at intermediate levels. Introduction of additional forward maskers increased these midlevel humps (Nizami, Reimer, and Jesteadt, 2002).

2.3 Phenomenon 2: Loudness enhancement

Galambos et al. (1972) were the first to report that an intense forward masker (presented contralaterally in their experiments) caused the loudness of a following brief sound to be *enhanced*. This phenomenon is rather counterintuitive given the usual physiological and psychophysical effects of masking: one would expect the presence of an intense forward masker to **reduce** the loudness of the following sound just as in simultaneous partial masking (e.g., Hellman and Zwislocki, 1964). The loudness enhancement effect is observed in similar conditions as the midlevel hump in intensity discrimination. In most experiments, a masker (frequently termed "conditioner" in this context) and a target were presented in the first observation interval, separated by a short inter-stimulus interval (ISI < 500 ms; in this thesis, the ISI is defined as the silent interval between offset of the first and onset of the second tone). Listeners adjusted the level of a comparison tone, presented approximately 1000 ms after the target tone, until loudness of the conditioner matched the loudness of the target (Fig. 5).

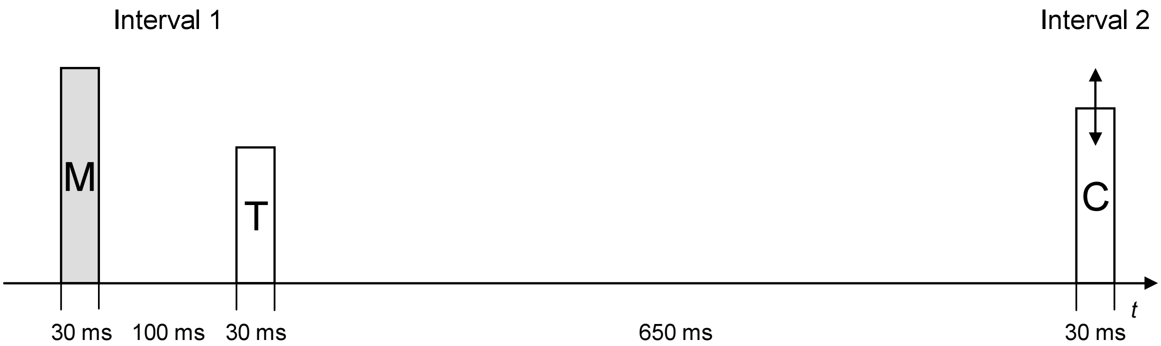

Fig. 5: Example trial configuration used in loudness enhancement experiments (e.g., Elmasian and Galambos, 1975). The listener adjusts the level of the comparison tone C so that it sounds equally loud as the target tone T.

Loudness enhancement is defined either as the difference between target level and the matched level of the comparison burst (e.g., Elmasian and Galambos, 1975), or as the difference between the loudness matches (physical level of the comparison burst) in the presence versus in the absence of the conditioner (e.g., Irwin and Zwislocki, 1971).

Loudness is enhanced if conditioner intensity is greater than target intensity (e.g., Elmasian and Galambos, 1975; Fig. 6, left panel). If the conditioner is less intense than the target, loudness is reduced (Zwislocki and Sokolich, 1974; Elmasian, Galambos, and Bernheim, 1980; Fig. 6, right panel). The amount of this loudness decrement is smaller than the amount of loudness enhancement. Note, that loudness decrement can not be explained by partial masking because the effect is stronger if the conditioner is **less** intense.

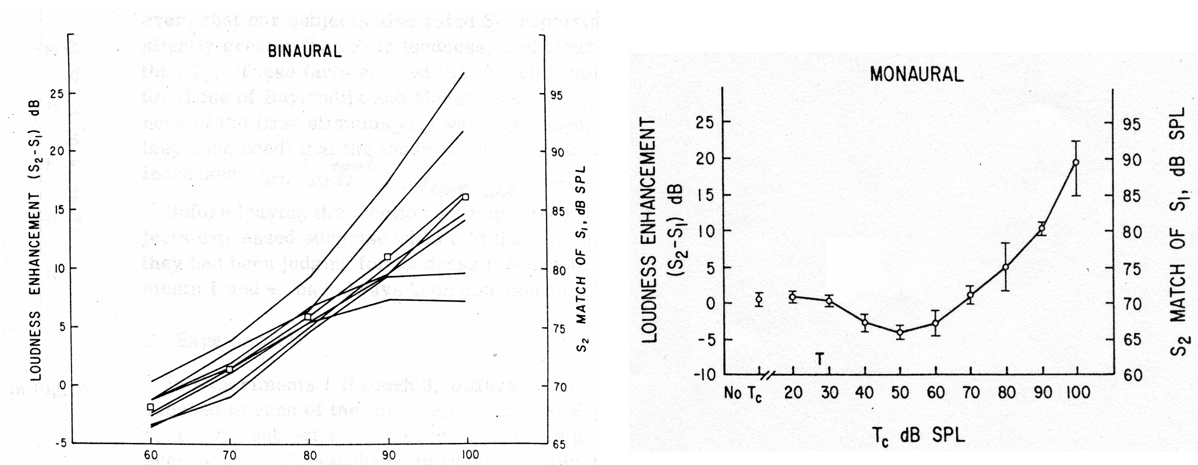

Fig. 6: Left panel: Individual loudness matches (difference between comparison tone level and target tone level, $L_C - L_T$) as a function of conditioner level. Target level was 70 dB SPL, the masker was presented diotically and the target monaurally. Squares denote mean data.
Right panel: Loudness matches for one listener as a function of masker level (plotted on the x-axis). Target level was 70 dB SPL, the masker was presented ipsilaterally, the symbol T denotes detection threshold for the listener. From Elmasian and Galambos (1975).

It should be noted that some early studies (e.g. Rubin, 1960) reported a "facilitation effect" of a forward masker on detection performance that resembles the enhancement effect in loudness. Detection thresholds for brief tones under forward masking were reduced relative to the values in quiet. Miskiewicz, Buus, and Florentine (1994) could demonstrate that such effects are absent if a 2I, 2AFC procedure is used and concluded that auditory facilitation can be understood as a procedural rather than a sensory effect.

2.4 Phenomenon 3: Intensity perception in 'short-long' equitone sequences

In a study by Vos (1977), listeners classified equitone sequences constructed as an alternation of short and long inter-onset-intervals (IOIs, Fig. 7) as either trochaic (each first tone of a pair accented) or iambic (each second tone of a pair accented). The stimuli were 600-Hz tones. Presentation level was not specified in the paper. In one condition, the sequences consisted of pairs of two 80-ms tones separated by an ISI of 80 ms; the interval between the offset of the second tone of one pair and the onset of the first tone of the next pair was 320 ms. In 85% of the trials, the sequence was classified as iambic, i.e., listeners reported hearing accents on the second sounds of the pairs. In other words, loudness of the second tones was larger than loudness of the first tones.

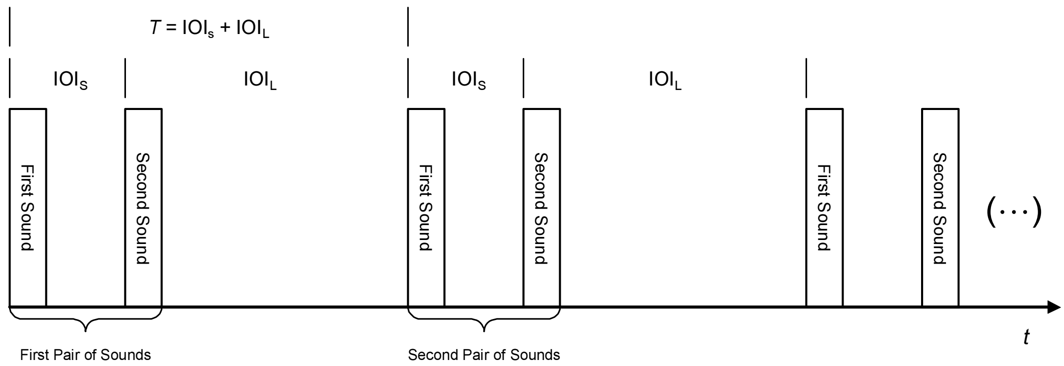

Fig. 7: 'Short-long' equitone sequence: Alternation of short and long IOIs (IOI_S and IOI_L, respectively). All sounds are physically identical. Listeners frequently perceive the second sounds of the pairs (sounds initiating the long IOIs) as more intense if the rate of presentation is sufficiently fast and the difference between the long and the short IOI is sufficiently large (Povel & Okkerman, 1981).

Povel and Okkerman (1981) conducted a series of studies using a comparable stimulus configuration. Rhythmic short-long sequences consisting of physically identical sounds were presented at several values of IOI_S and IOI_L. The stimuli were 1-kHz tones (duration 50 ms) presented at 50 dB SPL. In one set of conditions, the silent interval between a pair of tones was fixed to 50 ms while the interval between the offset of the second tone of a pair and the onset of the first sound of next pair was varied between 70 ms and 430 ms. In the remaining conditions, the short ISI was varied between 80 and 320 ms, while the long ISI remained at 340 ms. Listeners responded whether they had perceived the first sound or the second sound of each pair as accented. The proportion of perceived accents reported on the second sounds of the sound pairs increased with the difference between the short and the long ISI (Fig. 8). In an additional experiment, tone duration was 150 ms. The "interval-produced accent" was less pronounced than for the 50-ms tones, but the same overall pattern was found.

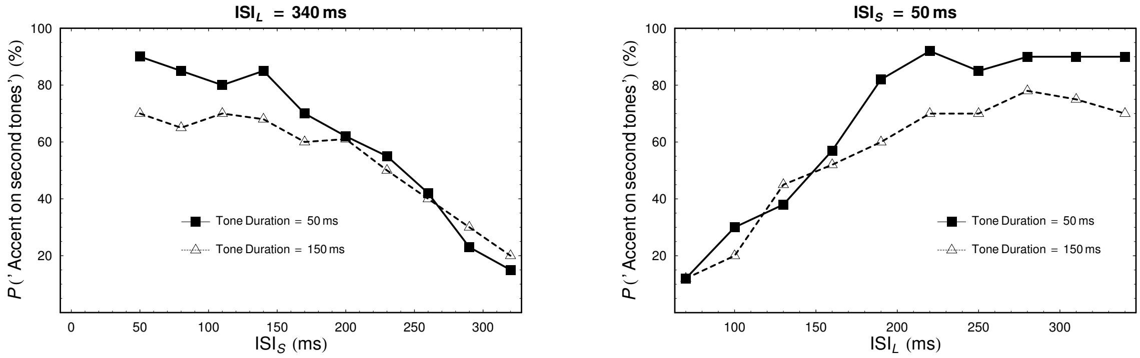

Fig. 8: Mean proportions of trials within which three listeners reported an accent on the second tones of the tone pairs in a study by Povel and Okkerman (1981). Left panel: long ISI fixed (340 ms), short ISI varied. Right panel: short ISI fixed (50 ms), long ISI varied. Squares: tone duration 50 ms. Triangles: tone duration 150 ms.

In experiment 4 of Povel and Okkerman's study, one listener adjusted the intensity of the first sounds within each rhythmic sequence until all tones sounded equally accented. The accent on the second sounds was balanced by increasing the level of the first sounds by about 4 dB SPL, i.e., loudness of the first and second tones was matched at a level difference of 4 dB.

The temporal configuration of the stimuli closely resembles those used in studies on loudness enhancement and intensity discrimination (cf. Fig. 1). Again, the first tone presented in temporal proximity of the second tone influences loudness of the latter. Two important differences, however, are the use of a same-level instead of a more intense 'conditioner' (i.e., the first sound of a pair) and the repeated, rhythmic presentation.

3 Empirical similarities between the three phenomena

In this Chapter, the empirical findings regarding the effects of various parameters on the three phenomena will be discussed. As the terminology used differs between the respective studies, it is important to recall the similarity between the settings. In Fig. 1 it can be seen that the masker presented in interval 1 of a discrimination experiment corresponds to the conditioner in a loudness enhancement experiment. The target tone in a loudness matching procedure corresponds to standard or standard-plus-increment in a discrimination setting. In the following, the terms standard and pedestal are used interchangeably.

For the rhythmic short-long sequences, the first tone of each tone pair could be viewed as the conditioner and the second tone as the target.

Most evidence comes from studies on intensity discrimination and loudness, as only two studies on the interval-produced accent have been published.

3.1 Target tone level

The most prominent effect of non-simultaneous masking on intensity discrimination is the observation that the jnd elevation effected by an intense masker is stronger at intermediate than at low or high standard levels (the midlevel hump). The standard level at which the maximum jnd elevation was found differs from study to study, but also between listeners within a study (e.g., Zeng, Turner, and Relkin, 1991). It is located between 40 and 60 dB SPL if a masker level of 90-100 dB SPL is used.

Most of the exceptions to this pattern occurred in experiments using perceptually similar maskers and standards. One of four listeners in the study by Schlauch, Lanthier, and Neve (1997) produced very large jnd's at low levels with a 10-ms pure-tone forward masker and a 10-ms standard of the same frequency. For two of the four listeners in an experiment by Schlauch, Clement, Ries, and DiGiovanni (1999, 100-ms, 1-kHz masker and standard), the jnd elevations at low and intermediate levels were approximately equal. This was also the case for listener RB of Zeng et al. (1991, 100-ms narrow-band noise masker and 25-ms 1-kHz standard). In a condition presenting 10-ms wide-band noise maskers and standards (Schlauch et al., 1997), jnd's were largest at low levels.

Schlauch et al. (1997) also observed elevated jnd's at high pedestal levels in an experiment using wide-band noise maskers.

For a 100-110-ms, 1-kHz, 90-dB SPL pure tone forward masker combined with 25-30 ms, 1-kHz targets, loudness enhancement was most pronounced at intermediate target levels in experiments by Zeng (1994) and Plack (1996 a).

Contradicting results were presented by Elmasian and Galambos (1975). All stimuli were 5-kHz, 20-ms tones. Masker level was fixed to 90 dB SPL, while target level was varied between 40 to 100 dB SPL. Enhancement increased with decreasing target level, i.e., it was largest at low target levels. Only one listener participated in the experiment, however, and the smallest target level was above the minimum level (20 dB SPL) used by Zeng and Plack.

There are no data on the dependence of the interval-produced accent on presentation level. As the midlevel hump is very important for the discussion of a common mechanism underlying the effects of non-simultaneous masking on loudness perception and intensity discrimination, the effect of presentation level on accent perception will be studied in Experiment 5.

The observation of a midlevel hump in both loudness enhancement and intensity discrimination represents an important similarity and hints to either a causal relation between the two phenomena (Carlyon and Beveridge, 1993), or to a common underlying mechanism (Plack, 1996 a).

3.2 Masker level

In an experiment by Zeng and Turner (1992), intensity jnd's for 40-60 dB SPL, 25-ms, 1-kHz standards were independent of masker level if the latter was lower than pedestal level. If masker level exceeded pedestal level, jnd's increased monotonically with masker level. 100-ms notched-noise maskers (center frequency 1 kHz) were used.

Schlauch, Clement, Ries, and DiGiovanni (1999) measured jnd's using 53-dB SPL forward maskers. There was nearly no jnd elevation at pedestal levels of 40, 50, and 60 dB. All stimuli were 1-kHz pure tones; tone duration was 100 ms for two listeners and 10 ms for another two listeners.

Nizami, Reimer, and Jesteadt (2001) found that shortening the duration of Gaussian-shaped 2-kHz tone-pips and broadband noise pips to values below 10 ms resulted in a midlevel-hump in quiet. In a subsequent study, Nizami, Reimer, and Jesteadt (2002) demonstrated that 200-ms, 2-kHz forward maskers increased the size of the humps at masker-pip intervals of 10 ms and 100 ms. At an ISI of 100 ms, there was **no** effect of masker level (50 dB SPL vs. 70 dB SPL).

Widin, Viemeister, and Bacon (1986) systematically varied standard and masker intensity in an intensity discrimination experiment. They used 1-kHz pure tones. Forward masker duration was 200 ms, pedestal duration 20 ms. The silent interval between masker and pedestal was only 5 ms. The authors presented their results in terms of the level of a tone in quadrature phase with the pedestal that would have resulted in the same sound pressure of standard-plus-increment as the in-phase increments actually used in the experiment. To be able to compare the results to data from other experiments, the increment levels ΔP for each listener and condition were estimated from the figures presented in the paper and converted to ΔL_{DL} using the relation

$$\Delta L_{DL} = 10 \log_{10}(1 + \Delta I/I) = 10 \log_{10}(1 + 10^{\Delta P/10}/10^{P/10}), \qquad (3\text{-}1)$$

where ΔP and P are the sound pressure levels (dB SPL) of the quadrature-phase increment and the pedestal, respectively.

As seen in Fig. 9, jnd's increased monotonically with the masker-standard level difference at each standard level. The slope of the masking function seemed to be slightly greater at low and intermediate than at high standard intensities. For listeners 1 and 2, there was evidence for a 'saturation effect': the jnd elevation appeared to be an s-shaped function of the level difference. As no data were obtained for low standard intensities combined with intense maskers, the presence or absence of a midlevel hump can not be inferred from these data. Above that, at an ISI of 5 ms only, different mechanisms can be expected to be in effect than at ISI of about 100 ms, as significant adaptation in the auditory nerve should be present in the 5-ms condition (Harris and Dallos, 1979).

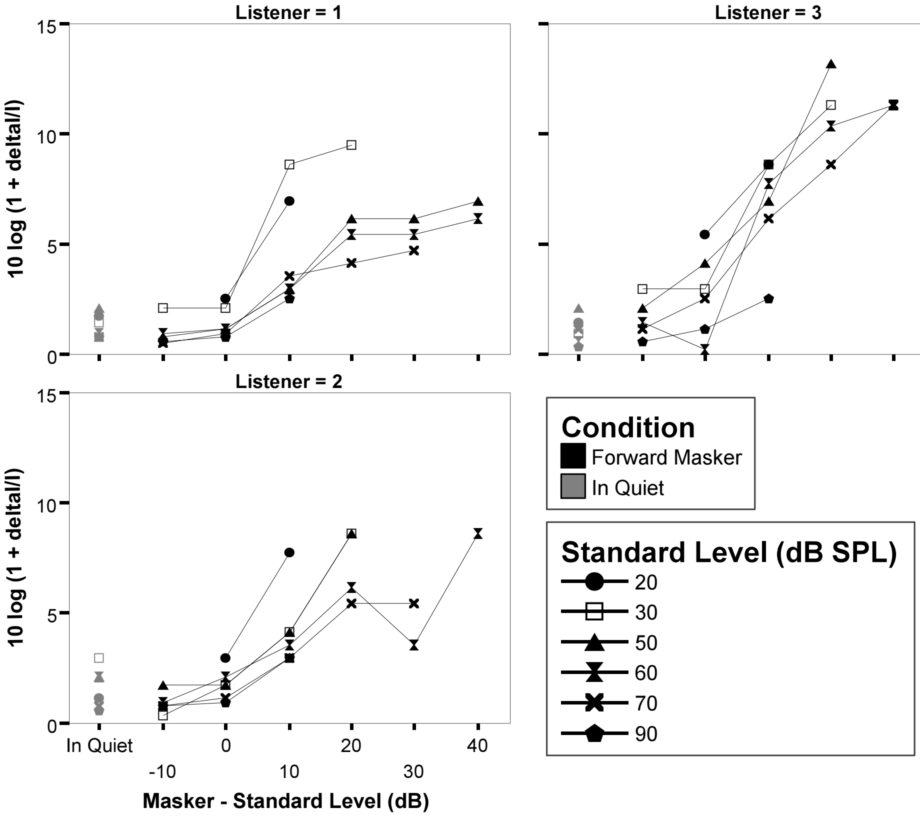

Fig. 9: Individual intensity jnd's *(ΔL_{DL}) as a function of the level difference ($L_M - L_S$) between masker and standard at various standard levels. Data from Widin, Viemeister, and Bacon (1986).*

For conditioner level exceeding target level, loudness enhancement increases monotonically with the level difference between conditioner and target (Elmasian and Galambos, 1975; Elmasian, Galambos, and Bernheim, 1980). For conditioner level smaller than target level, loudness decrement is a non-monotonic function of the level difference because the effect disappears as conditioner level approaches absolute threshold (Zwislocki and Sokolich, 1974; Elmasian, Galambos, and Bernheim, 1980). It should be noted that in these studies target level was fixed at either 70 or 90 dB SPL. Accordingly, the range of level differences presented in these studies was rather small because maximum masker level was 100 dB SPL. Experiments 2 and 3 of this thesis will measure loudness enhancement for a wide range of masker-target level differences at low, intermediate, and high target levels.

Zwislocki and Sokolich (1974) reported 1-3 dB of enhancement if conditioner and target were identical. In other experiments, there was no effect or even loudness decrement in this situation, however (Zeng, 1994; Elmasian and Galambos, 1975).

The interval-produced accent is observed in equitone sequences. This corresponds to masker level equal to target level in a three-tone matching task. For one subject in an experiment by Povel and Okkerman (1981), loudness of the first and the second tones was matched at a level difference of 4 dB. This value is larger than the 1-3 dB of enhancement found by Zwislocki and Sokolich (1974) for identical conditioner and target.

Except for the study of Nizami, Reimer, Jesteadt (2002), both jnd elevation and enhancement increased with the difference between masker and target intensity at a given target level. As the very brief Gaussian-shaped pips used by Nizami et al. also resulted in a midlevel hump in quiet, it is possible that the mechanisms effective for such stimuli differ from the situation where longer tone durations are used.

No experiment independently varied masker and target level over a range large enough to allow comparing the effects of the masker-target level difference at low, intermediate, and high target levels. Experiments 1-3 of this work are designed to provide such data.

The interval-produced accent appears to be stronger than the small amounts of loudness enhancement reported for conditioner and target of the same level, although this conclusion rests on a seriously limited data basis.

3.3 Temporal parameters

3.3.1 ISI between masker and target tone(*s*)

For intensity discrimination, the following effects were observed:

1. The jnd elevation at midlevels decreases as the mask-standard ISI is increased from 100 ms to larger values (Zeng and Turner, 1992; Plack, Carlyon, and Viemeister, 1995). At an ISI of 400 ms, Zeng and Turner (1992) still found a small midlevel hump, and Plack et al. (1995) reported a considerable jnd elevation at a standard level of 50 dB SPL.
2. Zeng and Turner (1992) reported a deviation from the midlevel hump pattern at the shortest ISI in their study (50 ms): jnd's were large at low levels also. They used a 100-ms narrow-band noise masker and a 25-ms, 1-kHz pedestal. On the other hand, Plack, Carlyon, and Viemeister (1995) found that decreasing the ISI to values below 100 ms resulted in **smaller** jnd's for a 50-dB SPL pedestal presented with ipsilateral 80-dB SPL forward maskers, ipsilateral backward maskers, or diotic backward maskers. In the remaining conditions (contralateral backward maskers, contralateral or diotic forward maskers), jnd's increased with decreasing ISI. Masker and standard were 30-ms, 1-kHz pure tones. Nizami, Reimer, and Jesteadt (2002) measured intensity-difference limens for brief Gaussian-shaped 2-kHz tone pips. If the masker-pedestal ISI was 10 ms rather than 100 ms, a 50-dB SPL, 2-kHz forward masker (duration 200 ms) caused a stronger jnd elevation at a standard level of 50 dB SPL for four of the nine listeners. Stellmack and Viemeister (2000), on the other hand, reported the jnd elevation effected by a forward or backward masker presented at masker-standard level differences between −10 and +10 dB to be essentially independent of the ISI. Stimuli were 10-ms, 1-kHz tone pulses, standard level was 75 dB SPL.

The following results were obtained in loudness enhancement experiments:

1. Zwislocki and Sokolich (1974) reported enhancement to drop to zero at a conditioner-target ISI of 500 ms. Stimuli were 10-ms, 1-kHz pure tones. The ipsilateral conditioner was presented at a level of 60 dB SL, target level was 40 dB SL.
   Bauer, Elmasian, Galambos (1975), using 8-kHz signals of only 0.2 ms duration, still found 6-10 dB of enhancement at an ISI of 400 ms. Conditioner level was 100 dB SPL, target level was 75 dB SPL.

2. For some listeners in an experiment by Zwislocki and Sokolich (1974), loudness enhancement of a 40 dB SL target was a non-monotonic function of the ISI between the 60 dB SL conditioner and the target. At an ISI of 10 - 30 ms, loudness was less enhanced than at 70 - 150 ms. In the median data, however, loudness enhancement was strongest at the shortest ISI. For same-level conditioners and targets (both presented at 40 dB SL), a comparable non-monotonic time-course was present in the median data. The authors proposed that at an ISI of 10 ms, conditioner and target appear fused and listeners match the comparison burst to the overall loudness of the pair. In other words, they attributed the strong median loudness enhancement at the shortest ISIs to listeners failing to separate the sensations elicited by conditioner and target, respectively. Zwislocki and Sokolich noted that for a noise conditioner, loudness was depressed rather than enhanced at short time intervals (Zwislocki et al., 1974) and suggested that burst separation should be easier if the conditioner has a different timbre. Results from the second experiment in their study do not support this idea, however, as the non-monotonic time course was present in the median data (target 40 dB SL, conditioner 40 or 60 dB SPL) if both conditioner and target were 10-ms bursts, but absent if conditioner duration was increased to 50 ms. As a different duration should make the distinction between target and conditioner easier, the non-monotonic time course should have been present for the 50-ms rather than the 10-ms conditioner according to the 'burst-separation' hypothesis.
3. Zwislocki, Ketkar, Cannon, and Nodar (1974) presented wide-band noise conditioners combined with 1-kHz pure-tone targets and comparison bursts. Target level was 45 dB SL. The sound pressure level of the noise within a critical band centered at 1 kHz was equal to target level. At ISIs between 100 and 200 ms, there was loudness enhancement (4 dB), while loudness of the target was reduced at smaller ISIs.

In the condition with constant inter-pair interval (340 ms; Fig. 8, left panel), Povel and Okkerman (1981) found that perception of the second tones in short-long sequences as accented occurred only if the intra-pair interval did not exceed roughly 250 ms. To compare this value with the temporal limit found for loudness enhancement and the midlevel hump in intensity discrimination, it is important to note that for the interval-produced accent, the temporal asymmetry between the short intra-pair an the longer inter-pair interval is important. If both intervals are identical, any influence of, e.g., the first sound of a pair on loudness of the second sound is matched by the influence of the second sound on the first sound of the next pair. From this point of view, the interpretation of Povel and Okkerman's data is problematic. In the condition with constant inter-pair interval, the difference between the long and the short ISI *decreased* with short interval duration, so that the estimate of 250 ms can be assumed to considerable underestimate the maximum short-interval duration resulting in an interval-produced accent.

This issue will be discussed in greater detail in Chapter 8, where Experiment 4 will provide data for a condition in which interval difference is not correlated with the duration of the intra-pair interval. Additionally, a quantitative model will be used to derive a more reliable estimate of the time constant from the data by Povel and Okkerman (1981).

The maximum masker-target interval at which both the jnd elevation at midlevels and loudness enhancement can be observed appears to be approximately 500 ms. How can this value be reconciled with the fact that temporal resolution of the auditory system has been found to be better than 3 ms in many experiments (see Eddins and Green, 1995, for a review)? It seems that in the tasks discussed in this thesis, the auditory system 'integrates' information over a substantial time window (cf. Plank and Ellermeier, 2003, for a discussion of the "resolution-integration paradox" in the context of loudness perception).

The maximum value of short ISI duration resulting in an interval-produced accent is 250 ms according to Povel and Okkerman (1981), which would be significantly smaller than the value relevant for the other two phenomena. As there are problems with Povel and Okkerman's estimate, however, the answer to the question of whether the time courses of the three phenomena are compatible or not has to be delayed until Experiment 4 of this work.

A non-monotonic time course was reported for both loudness enhancement and the midlevel hump in intensity discrimination, as the effects were smaller at ISIs below 100 ms than at ISIs greater than 100 ms. Results are not unequivocal, however. We will return to the issue when discussing models of the midlevel hump. Povel and Okkerman (1981) reported the interval-produced accent to be most pronounced at the smallest intra-pair interval (50 ms).

3.3.2 Target tone duration, masker duration

Schlauch, Lanthier, and Neve (1997) varied masker duration and standard duration. They observed only a very small midlevel hump in intensity discrimination if a 250-ms pure tone masker was followed by a 10-ms standard, while a pronounced hump resulted if both masker and standard were 10 ms long, or if masker standard were both 250-ms tones. Essentially the same effect was found for wide-band noise maskers of 10 or 100 ms duration presented before 10-ms pure-tone or wide-band noise standards: midlevel humps were stronger in the equal-duration conditions.

Plack, Carlyon, and Viemeister (1995) presented pure-tone maskers of the same duration as the pedestal and reported ΔL_{DL} to be approximately 15 dB for a 50-dB SPL standard combined with an 80-dB SPL masker. Zeng (1998), using a 100-ms forward masker, found average ΔL_{DL} for a 50-dB SPL standard combined with a 90-dB SPL masker to be only 7-11 dB, although he used a 3-down, 1-up tracking rule rather than the 2-down, 1-up rule employed by Plack et al. In studies in which masker duration was longer than pedestal duration and narrow-band noise maskers were used (e.g., Zeng, Turner, and Relkin, 1991), maximum difference limens were as small as 5 dB. In this case it remains unclear, however, if masker duration, masker waveform, or both produced the effect. In general, the conclusions to be drawn from the comparison between different studies are very limited due to the considerable inter-individual variation.

In the case of equal-duration masker and standard, the size of the hump was smaller for the 250-ms than for the 10-ms stimuli in the experiment by Schlauch, Lanthier, and Neve (1997). As mentioned above, very short tone durations result in a midlevel hump even in quiet (e.g., Raab and Taub, 1969), so that this finding is not surprising.

For loudness enhancement, Zwislocki and Sokolich (1974) found only a small effect of masker duration. A 50-ms conditioner combined with a 10-ms target produced a slightly smaller amount of loudness enhancement than a 10-ms conditioner, but the maximum difference was 2 dB. The duration range used by Zwislocki and Sokolich, however, was much smaller than in Schlauch et al.'s (1997) study (10 ms vs. 50 ms and 10 ms vs. 250 ms, respectively).

Target tone duration has not been varied within the same loudness enhancement experiment. Bauer, Elmasian, and Galambos (1975) presented 8-kHz pips with a duration of 0.2 ms. For a 100-dB SPL masker combined with a 65-dB SPL target, they found approximately 17 dB of loudness enhancement. This value is virtually identical to the 18 dB of loudness enhancement observed by Elmasian and Galambos (1975) for 20-ms, 5-kHz stimuli (100-dB SPL masker, 70 dB SPL target). These results suggest that loudness enhancement of a target of moderate intensity by an intense masker does not increase with decreases in duration, but this conclusion again rests on a comparison between data from different experiments.

Both Vos (1977) and Povel and Okkerman (1981) reported that the interval-produced accent was stronger for shorter tone durations. Yet, varying duration in the short-long sequences also altered the tempo of the sequences (i.e., number of stimuli per time unit) as the silent intervals between the tones remained fixed. Accordingly, interpretation of these results is again problematic.

For the midlevel hump in intensity discrimination, the effect of using a masker differing in duration from the standard appears to be best described as a similarity effect. This finding is of relevance for the discussion of models of the midlevel hump.

The effect of increasing masker duration from 10 to 50 ms was rather small in the loudness enhancement experiment by Zwislocki and Sokolich (1974). To draw a conclusion, however, more data would be required.

The midlevel hump in intensity discrimination is more pronounced at short tone durations. For loudness enhancement, the comparison between two experiments using different durations indicated that target duration has virtually no effect in the very restricted range of conditions studied. For the interval-produced accent, there was a duration effect comparable to the effect found for discrimination (effect stronger for shorter tones). The tempo of the sequence was correlated with tone duration, however.

3.3.3 Forward vs. backward masking

Midlevel humps in intensity discrimination were observed under both forward and backward masking (Plack and Viemeister, 1992 a, b; Plack, Carlyon, and Viemeister, 1995; Plack, 1996a). Plack and Viemeister (1992 b) reported the hump observed under backward masking to be even more pronounced.

Loudness enhancement and decrement were observed both in settings where the conditioner preceded the target (‘forward masking’), and where the conditioner followed the target (‘backward masking’; Elmasian, Galambos, and Bernheim, 1980). In the forward masking condition, slightly more enhancement was found.

In the short-long sequences, forward and backward ‘masking’ effects can only be inferred indirectly (see Chapter 4.3.1).

The observation of jnd elevations caused by backward masking has important consequences for models of the effect, as this finding is strong evidence against the auditory periphery as the locus of the relevant mechanisms (see Chapter 4.1.1.1.5). At the same time, the fact that both loudness enhancement and the jnd-elevation are observed for backward maskers is further evidence for the close relation between the two phenomena.

3.4 Masker laterality

In intensity discrimination, contralaterally presented maskers produced significant midlevel humps, which were smaller in size than for ipsilateral maskers, however (Plack, Carlyon, and Viemeister, 1995; Zeng and Shannon, 1995; Schlauch, Clement, Ries, and DiGiovanni, 1999). In an experiment by Plack et al. (1995), though, the effect of a contralateral masker exceeded the effect of an ipsilateral masker at small ISIs between masker and standard (< 100 ms).

Schlauch et al. (1999) conducted an experiment in order to test the hypothesis that the hump observed under contralateral stimulation was caused by cross hearing (sound conduction from the stimulated to the non-stimulated ear). Assuming an interaural attenuation of at least 40 dB, they compared the effects of a 93-dB SPL contralateral masker and a 53-dB SPL ipsilateral masker. The jnd elevation was considerable greater for the contralateral masker.

Schlauch et al. (1999) also combined an ipsilateral masker (80 dB SPL) and a contralateral masker (93 dB SPL). The interesting feature of this condition was that listeners reported to perceive **only** the contralateral masker. Nevertheless, the size of the midlevel hump was comparable to that produced by the ipsilateral masker only.

Diotic forward maskers combined with a monaural standard resulted in identical or slightly smaller jnd elevations than ipsilateral maskers in two experiments by Plack, Carlyon, and Viemeister (1995). In backward masking and with ISIs smaller than 100 ms, diotic maskers had a stronger effect than ipsilateral maskers, however.

Loudness enhancement and decrement are weaker but still significant for dichotic presentation of conditioner and target (Elmasian and Galambos, 1975; Elmasian, Galambos, and Bernheim, 1980).

A diotic conditioner combined with a monaural target produced a slightly smaller amount of loudness enhancement than an ipsilateral conditioner in an experiment by Elmasian and Galambos (1975).

For the interval-produced accent, no data are available.

The finding of reduced but still significant effects of contralateral maskers applies to discrimination as well as to loudness. This observation is an important similarity between the two phenomena, especially as this finding presents a significant problem to one of the models proposed for the discrimination data. The results restrict the possible physiological location of the underlying mechanism to a point beyond the first binaural interaction (superior olivary complexes, cf. Yost, 2000).

A diotic forward masker presented with a masker-standard ISI of 100 ms slightly decreased the size of the midlevel jnd elevation relative to an ipsilateral masker and resulted in loudness enhancement being slightly less pronounced in a comparable setting. For backward masking and at small ISIs, a diotic masker produced stronger jnd elevations than an ipsilateral masker, however. As no loudness enhancement data are available for these conditions, it remains unclear whether the effects of masker laterality are exactly equivalent for the two phenomena.

3.5 Target tone frequency

The midlevel jnd elevation is greater at higher standard frequencies. This was demonstrated by Carlyon and Beveridge (1993) using 1-kHz and 6.5-kHz pedestals, and by Plack and Viemeister (1992 a) using 1-kHz and 6-kHz pedestals, respectively.

At frequencies exceeding 4 kHz, a midlevel hump is observed even in the absence of masking (the “severe departure from Weber’s law”, Carlyon and Moore, 1984, 1986 a, b; Plack and Viemeister, 1992 a; Carlyon and Beveridge, 1993; Plack, 1998)

To my knowledge, the effect of target tone frequency on loudness enhancement has never been measured within the same experiment. Therefore, it is only possible to compare the results from different studies.

Elmasian, Galambos and colleagues (1975, 1980) obtained loudness matches for 5-kHz, 20-ms conditioners and targets, with an ISI of 100 ms and a target level of 70 dB SPL. With a 90-dB conditioner, mean loudness enhancement was 7 dB with in the backward masking and 12 dB in the forward masking condition. Zeng (1994), using a 1-kHz, 100-ms, 90 dB SPL forward masker and a 1-kHz, 25-ms target, reported that on average a 70 dB target was matched by an 80 dB conditioner (10 dB enhancement). Plack (1996 a) found between 2 and 10 dB of enhancement for a 1-kHz, 110-ms, 90-dB SPL forward masker and a 1-kHz, 30-ms target presented at levels between 60 and 70 dB SPL. These results are weak evidence for loudness enhancement being more pronounced at higher target frequencies. Note, however, the longer conditioner duration and the different experimental procedures used. In the experiment by Elmasian and Galambos, listeners adjusted the level of the third tone (comparison burst) to match the loudness of the target (second tone within the trial, Fig. 5). In the experiments by Zeng and Plack, an adaptive double-staircase procedure was used to obtain loudness matches, and the level of the masked tone (second tone within the trial) was varied while the level of the third tone was fixed. Before this background, the finding that loudness enhancement seemed to be somewhat more pronounced at 5 kHz than at 1 kHz should be viewed with great caution.

For the interval-produced accent, no data are available.

The midlevel jnd elevation caused by a forward masker increases with target frequency. Results on loudness enhancement are not incompatible with this finding, but target frequency has not been varied within the same experiment.

Note also that a midlevel hump in intensity discrimination is observed at high standard frequencies even in the absence of masking. For this reason, the increased midlevel jnd at higher frequencies could be due to two different mechanisms.

3.6 Masker frequency

The deterioration in intensity discrimination performance at midlevels is most pronounced if mask frequency equals target frequency (Zeng and Turner, 1992).

The largest amount of enhancement was observed if conditioner frequency was slightly below or equal to target frequency (Zwislocki and Sokolich, 1974).

For the accent, varying masker frequency is equivalent to presenting tone 1 and tone 2 at different frequencies. No such data are available.

The effects found for loudness enhancement and the midlevel hump in intensity discrimination are very similar. Specifically, the slope of the function relating jnd or loudness enhancement, respectively, to masker frequency appeared to be greater at frequencies above target frequency than below target frequency (Fig. 10). Such pattern is compatible with auditory filter shape (cf. Moore, 1995).

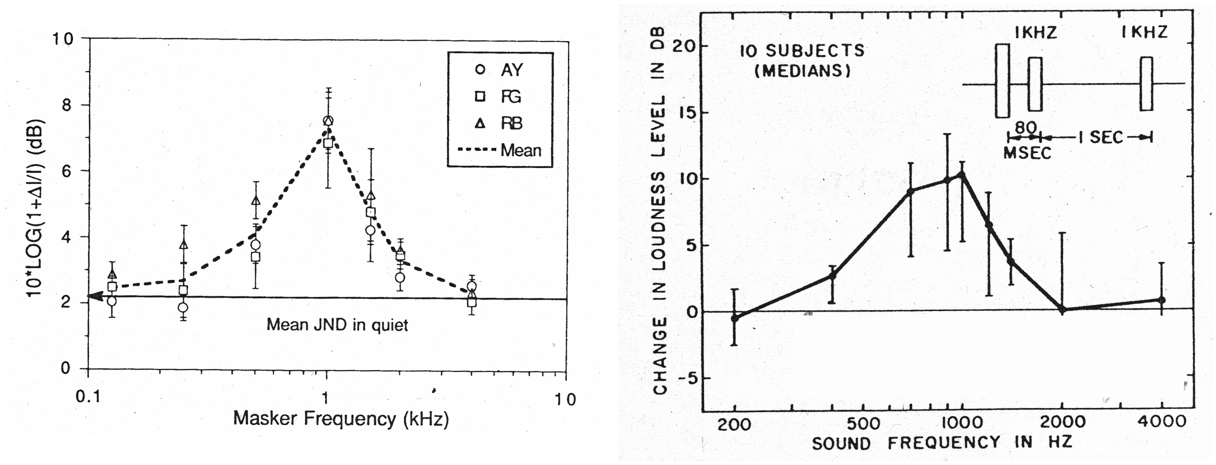

Fig. 10: Intensity jnd's *(left panel; from Zeng and Turner, 1992) and loudness enhancement (right panel; from Zwislocki and Sokolich, 1974) as a function of masker frequency. Standard/target frequency was 1 kHz in both experiments.*

3.7 Masker spectrum

Schlauch et al. (1999) used 100-ms maskers and standards in an intensity discrimination experiment. In one condition, a 1-kHz pure tone masker was presented. In the other condition, a 4.133-kHz sinusoid of the same sound-pressure level was added to the 1-kHz sinusoid, resulting in a two-tone masker. The two-tone masker reduced the size of the hump for 3 of the 4 listeners, relative to the values observed with the single-tone masker. A small midlevel jnd elevation remained for all subjects in the two-tone mask condition. The two-tone masker also reduced the jnd elevation at low levels present for two listeners and the very large jnd elevation at low levels found for one listener in an experiment by Schlauch et al. (1997).

A 10-ms broadband noise masker caused a slightly stronger increase in difference limens for a 10-ms tonal standard than a 10-ms pure tone masker in the study by Schlauch et al. (1997). Different listeners participated in the two conditions, however.

Only one experiment presented other than pure-tone conditioners. Zwislocki, Ketkar, Cannon, and Nodar (1974) used wide-band noise conditioners combined with 1-kHz pure-tone targets and comparison bursts. Target level was 45 dB SL. In one condition, the overall sound pressure level of the noise was equal that of the target. There was no effect on the perceived intensity of the target. In the second condition, the sound pressure level of the noise within a critical band centered at 1 kHz was equal to the sound pressure level of the target. For ISIs between 100 and 200 ms, there was loudness enhancement (4 dB), while loudness of the target was reduced for smaller ISIs. Unfortunately, the authors did not measure loudness enhancement produced by a pure-tone conditioner with sound pressure level equal to the total level of the louder noise. The origin of the loudness reduction at small ISIs remains unclear.

For the interval-produced accent, no data are available.

The effects of masker spectrum on midlevel jnd's seem to be best described as similarity effects. This result is important for the discussion of models of the effects of non-simultaneous masking on intensity difference limens. The data are too limited to decide whether masker spectrum has a comparable on loudness enhancement.

3.8 Effects of additional stimuli

If additional stimuli besides masker and target are present in the observation intervals, a rather complex pattern of results has been reported. As the findings are of importance for models of the midlevel hump in intensity discrimination, a summary of the observations is attempted here. The effect of additional stimuli on loudness enhancement and the interval-produced accent has not been studied.

3.8.1 Noise presented simultaneously with the pedestal

Plack and Viemeister (1992 a, b) observed that a notched noise presented simultaneously with the pedestal and with a spectrum level 30 dB below pedestal level caused the midlevel hump to disappear because the jnd elevation effected by the masker was now approximately independent of standard level. This finding was unexpected, since the hypothesis was that the notched noise should prevent off-frequency listening and thus **increase** the size of the midlevel hump according to the "recovery-rate model" (Zeng, Turner, and Relkin, 1991) that will be discussed in Chapter 4.1.1.

The noise removed the hump regardless if it was gated on and off simultaneously with the pedestal, or if the 30-ms pedestal was presented temporally centered in a 125-ms notched noise. Without forward masker, the notched noise resulted in elevated jnd's at high pedestal levels. This pattern can be explained by restricted off-frequency listening, which has an effect only at higher levels as spread of excitation is small at low levels. The jnd elevation at high levels was also present under forward masking. In a control condition, notched-noise spectrum level was fixed at 30 dB and resulted in reduced jnd's at intermediate levels, while resolution at higher levels did not change. Similarly, under backward masking, no midlevel hump was present in the notched noise condition (Plack and Viemeister, 1992 b). The increase in difference limens caused by the backward masker was stronger than for the forward masker.

Zeng (1998) emphasized that the notched noise in Plack and Viemeister's (1992 a) experiments caused not only off-frequency but 10-14 dB of on-frequency masking. Using the same stimuli as Plack and Viemeister except for a pure-tone instead of a narrow-band noise forward masker, he found that the notched noise reduced midlevel jnd's for two of the three listeners. For the third listener, the hump seemed to be shifted towards higher standard levels.

Zeng also measured forward masked intensity jnd's with notched-noise spectrum level fixed at −10, 10, 30, and 50 dB. The −10 dB noise caused only 5 dB of threshold shift at the signal frequency, while the 10, 30, and 50 dB noise caused approximately 10, 25, and 50 dB of threshold shift, respectively.

Without the noise, forward-masked intensity jnd's were elevated considerably at standard levels between 40 and 70 dB SPL. With the −10 dB noise, jnd's were reduced at 40 dB SPL and increased at 50 dB SPL standard level; all remaining jnd's were unaltered. The 10 dB noise caused a jnd decrease at 40 dB and 50 SPL, an increase at 60 dB SPL, and nearly no effect at the other standard levels. The 30 dB noise reduced jnd's at 50 and 60 dB SPL and increased them at higher levels. Thus, the midlevel hump appeared to be shifted towards higher levels with increasing notched-noise level.

In a second experiment, Zeng used narrow-band noise instead of notched noise. Noise level was selected for each listener to produce the same on-frequency threshold shift as the noise in Plack and Viemeister's (1992a) study. Again, noise level was increased with signal level. Jnd's were reduced at standard levels of 50 to 80 dB SPL and unaltered at 30, 40, and 90 dB SPL.

In the third experiment, a 50-dB spectrum level high-pass noise (2000-5000 Hz) was presented that produced only 10 dB of threshold shift at signal frequency. Intensity jnd's were unaltered at low levels but elevated at pedestal levels above 50 dB SPL.

3.8.2 Notched noise presented immediately before the pedestal

Plack and Viemeister (1992 a) also included a condition in which the offset of the notched noise (50 ms steady-state duration) coincided with the onset of the pedestal. Relative to jnd's in quiet, jnd's obtained without forward masker were elevated at higher levels (compared to the jnd's in quiet). With forward maskers, these jnd-level functions seemed to be shifted vertically (upwards) for 3 of the 4 participants. For pedestal levels between 40 and 60 dB SPL, forward-masked jnd's were strongly reduced relative to the condition without noise except for one listener.

3.8.3 Proximal burst presented before or after the pedestal

Plack, Carlyon, and Viemeister (1995) added a third tone to each observation interval in a 2I, 2AFC task. This "proximal burst" was a 30-ms 1-kHz pure tone (just as masker and standard) and was presented ipsilaterally before or after the pedestal, with a silent interval of 12.5 ms. In nearly all conditions (forward/backward masking, burst presented before/after the target, masker ipsilateral/diotic/contralateral), the burst improved performance. Interestingly, the beneficial effect of a proximal burst presented before the pedestal was greater if burst level **exceeded** pedestal level. The latter effect was less clear if the burst occurred after the pedestal. Plack (1996 b), using a comparable setting and a backward masker, found that the optimum proximal burst level depended on the ISI between the burst and the following pedestal and appeared to be 10 to 30 dB above pedestal level. Again, results were inconsistent for the burst following the pedestal, but the optimum level seemed to be *below* standard level in this situation. There was evidence for frequency selectivity: the maximum improvement occurred if proximal burst frequency equaled pedestal frequency.

3.8.4 Discussion

Comparable conditions have not been studied for loudness enhancement. Therefore, the discussion of the complex effects of additional stimuli is postponed until the review of models of the midlevel hump in intensity discrimination in Chapter 4.

3.9 Various

1. In experiments conducted by Zeng and Shannon (1995), the midlevel hump in intensity discrimination was present for cochlear implant listeners, but absent for auditory brainstem implant listeners.
2. Turner, Horwitz, and Souza (1994) found a significant midlevel elevation of intensity jnd's using the standard 2I, 2AFC, adaptive procedure, but no midlevel hump in an adjustment task. This difference between psychophysical methods is not present in experiments measuring loudness enhancement, where the effect was found both in adjustment and in 2I, 2AFC, adaptive procedures (e.g., Elmasian and Galambos, 1975; Zeng, 1994).

3.10 Summary: Empirical similarities

Many parameters show comparable effects on loudness enhancement and intensity discrimination under non-simultaneous masking. Especially important are the midlevel humps observed under both forward and backward masking and also with contralateral maskers in both phenomena. Loudness enhancement and intensity resolution also show a very similar dependence on the variation of masker frequency relative to target frequency.

Differences exist in the size of the effects under backward masking as compared to forward masking. Above that, while a strong effect of masker duration on intensity discrimination was reported, this parameter had only a very small effect on loudness enhancement. The duration difference presented in the single loudness matching experiment (Zwislocki and Sokolich, 1974) may have been too small, though.

Several parameters have not been studied systematically for loudness enhancement (e.g., target frequency, diotic backward maskers, and effects of additional stimuli).

As to the interval-produced accent, only data on the effect of the temporal structure have been reported. While loudness enhancement and the midlevel hump in intensity discrimination were observed with target-mask intervals of 400 ms, the accent seemed to disappear if the intra-pair interval exceeded 250 ms in the Experiment by Povel and Okkerman (1981). As argued above, the interpretation of Povel and Okkerman's data is problematic and Experiment 4 of this work is designed to obtain more accurate estimates of the temporal limits of accent perception.

To summarize, the review supports the idea put forward by Carlyon and Beveridge (1993) that the empirical similarities suggest a common mechanism or a causal link between loudness enhancement and the midlevel hump in intensity discrimination, while no conclusion seems possible for the interval-produced accent due to the very limited data.

4 Models proposed for the phenomena

For each of the three phenomena, several explanations have been proposed, which are reviewed now. Additionally, alternative explanations for some of the phenomena will be discussed.

4.1 Explanations for the midlevel hump in intensity discrimination

4.1.1 Different recovery-rates of high-SR and low-SR auditory nerve fibers

Zeng et al. (1991) based their explanation of the midlevel hump in intensity discrimination on the physiological findings reported in the companion paper by Relkin and Doucet (1991). As observed by Kiang et al. (1965) first, there exist at least two different populations of auditory nerve fibers. One exhibits high rates of spontaneous activity and low thresholds, the other has low spontaneous rates (SR) and high thresholds (Liberman, 1978; Sachs and Abbas, 1974; Evans and Palmer, 1980; Palmer and Evans, 1980). The operating ranges of the two populations suggest that high-SR neurons code low intensities while low-SR neurons operate at higher intensities.

Now, Relkin and Doucet (1991) found that after stimulation by an intense sound, low-SR auditory nerve neurons in chinchilla recovered **slower** than high-SR neurons. They used an adaptive procedure. In two observation intervals, “[...] spikes were counted during identical time windows in each of the two intervals, only one of which, chosen randomly, contains the probe tone” (p. 217). The procedure converged at a probe tone level that was ‘detected’ (higher spike count in the time window containing the standard-plus-increment) in 75% of the trials. Tone duration was 25 ms. Maskers of 102 ms duration were presented in both intervals in the forward masking condition. The silent inter-stimulus-interval between mask and probe tones was varied between 0 and 240 ms. The IOI between the maskers was 430 ms, the inter-trial interval was not reported. Masker and probe tone were presented at the characteristic frequency of each fiber. Masker level was 40 dB above each fiber’s threshold in quiet.

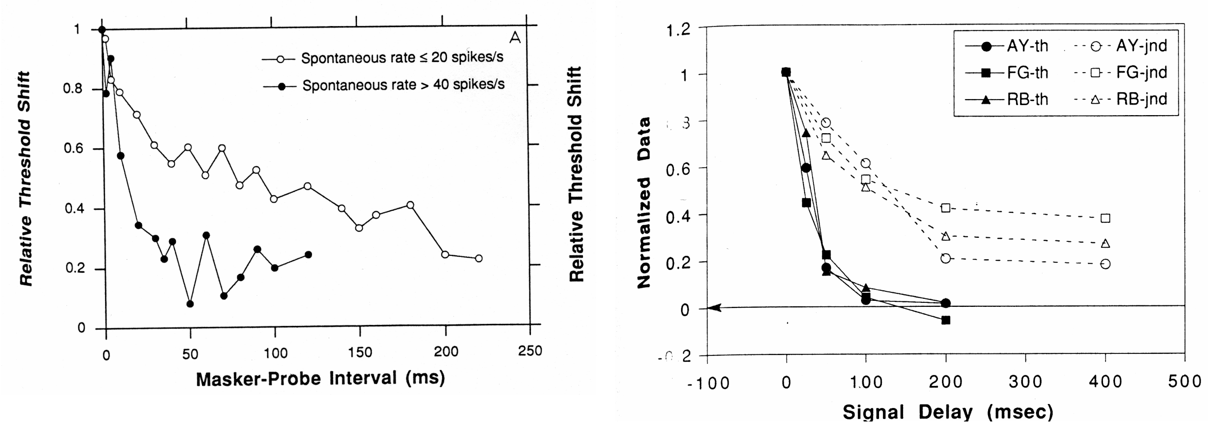

Fig. 11: Left panel: Recovery of auditory nerve neurons (from Relkin and Doucet, 1991). Right panel: Normalized shifts of detection threshold (filled symbols) and intensity jnd’s *under forward masking at an intermediate pedestal intensity (open symbols). From Zeng et al. (1991).*

The main finding was that thresholds of high-SR fibers recovered much faster than those of low-SR fibers (Fig. 11, left panel). The estimated times to complete recovery were 2.064 s for the low-SR and 215 ms for the high-SR fibers, respectively. This observation is qualitatively compatible with the finding of Salvi et al. (1986) that in chinchillas, the time needed for the neural probe rate to recover was longer for low-SR than for high-SR neurons. For a masker 50 dB above threshold of the fiber, the estimated time constants were 100 ms for the high-SR and 200 ms for the low-SR neurons, respectively. Masker duration was 100 ms, the probe was a 10-ms tone presented 10 dB above threshold, and both tones were presented at the characteristic frequency of the fiber.

The time course of threshold recovery of the high-SR neurons corresponds closely to the psychophysically measured recovery of absolute threshold in forward masking (Jesteadt et al., 1982), suggesting that high-SR neurons are responsible for detection. On the other hand, the time course of recovery of the low-SR neurons resembles the ISI dependence of intensity jnd elevations (Zeng et al., 1991; Fig. 11, right panel).

Zeng et al. (1991) argued that these physiological findings can be used to explain the midlevel hump in intensity discrimination occurring in the presence of an intense masker presented 100 ms before the probe tones. They hypothesized that in quiet, there is a smooth transition between the operating ranges of the two populations (Fig. 12, upper panel).

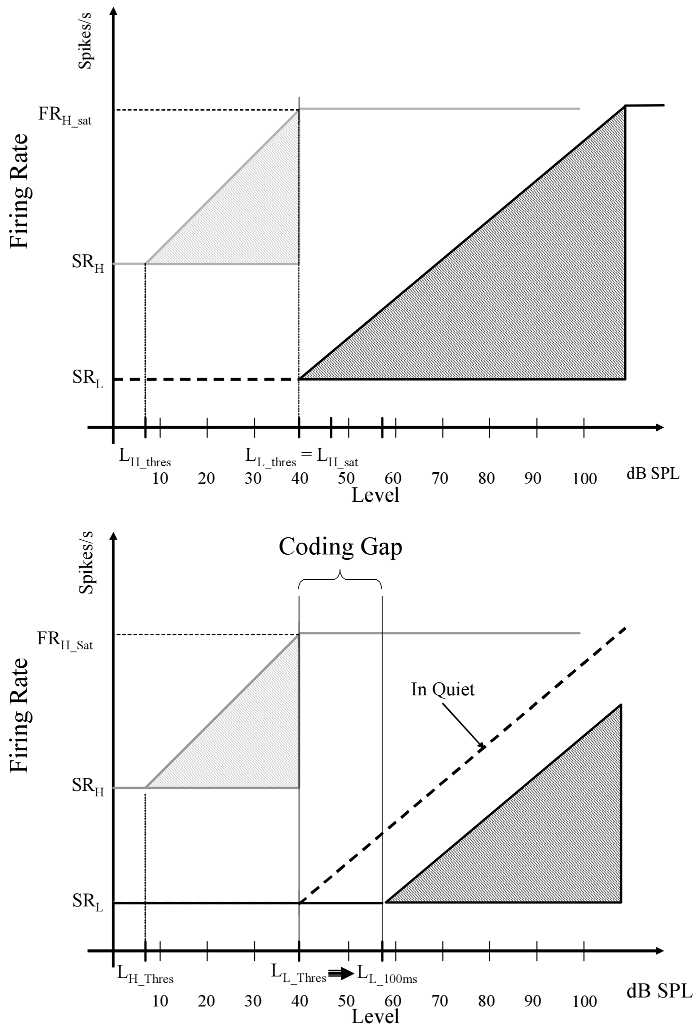

Fig. 12: Hypothetical rate-level functions of high-SR and low-SR neurons. Upper panel: in quiet. There is a smooth transition between the operating ranges. Lower panel: in forward masking (masker-tone ISI = 100 ms). The mask creates a 'coding gap' at midlevels because high-SR neurons recover within several milliseconds while the threshold of low-SR neurons is still elevated 100 ms after the masker.

The high-SR fibers code intensities in the range between L_{H_Thres} (threshold) to L_{H_Sat} (saturation). The low-SR fibers code intensities in the range between L_{L_Thres} (threshold) to L_{L_Sat} (saturation). Now, an intense forward masker elevates the thresholds of the low-SR fibers for several hundred milliseconds. As a result, 100 ms after the masker, the high-SR fibers are already recovered, while the thresholds of the low-SR fibers are still shifted from L_{L_Thres} to L_{L_100ms}. This results in a midlevel 'coding-gap' between the operating ranges of the two populations (Fig. 12, lower panel). The high-SR neurons can code intensity at low pedestal intensities in this situation. At high pedestal intensities, low-SR neurons contribute to intensity discrimination. In contrast, at pedestal levels above the operating range of high-SR fibers, but below the still elevated threshold of low-SR fibers, intensity discrimination is impaired.

Zeng and Shannon (1995) measured intensity discrimination in normal, cochlear-implant, and auditory-brainstem-implant (ABI) listeners. In normal and in cochlear-implant listeners (where the hair cell and auditory-nerve synapse is bypassed) they found a midlevel hump. The hump was absent in ABI listeners, where the auditory nerve is bypassed and intrinsic processing in the auditory nucleus does not occur. Zeng and Shannon interpreted the data as to suggest that a mechanism either peripheral to or in the cochlear nucleus is the most likely origin of the hump. Incompatible with this view, however, is the finding that contralateral masking produces a small but significant hump in normal listeners (Plack, Carlyon, and Viemeister, 1995; Schlauch et al., 1999).

4.1.1.1 Recovery-rate model versus data

For which effects of stimulus parameters on the midlevel hump in intensity discrimination can the neuronal-recovery-rate model account for?

4.1.1.1.1 Standard level

The model can account for the midlevel hump observed with an intense forward masker.

4.1.1.1.2 Masker level

The increase of the difference limens for a fixed intermediate-level standard with masker level is correctly predicted by the model, as the midlevel 'coding gap' caused by elevated thresholds of the low-SR fibers becomes more pronounced with increasing masker level.

The effect of the masker-standard level difference at different standard levels will be studied in Experiment 1 of this thesis.

4.1.1.1.3 Masker-standard ISI

The decrease of the jnd elevation as the ISI between mask and tone increased above 100 ms (Zeng and Turner, 1992) is of course predicted by the model, as the 'coding gap' disappears if the low-SR neurons are fully recovered at presentation of the target tone. The jnd elevation at low levels observed at an ISI of 50 ms (Zeng and Turner, 1992) can be understood as reflecting recovery of the high-SR neurons. Plack, Carlyon, and Viemeister's (1995) finding of a jnd reduction at ISIs smaller than 50 ms is incompatible with the model, however.

4.1.1.1.4 Masker duration, standard duration

Several studies measuring auditory-nerve responses demonstrated that adaptation increases with masker duration (e.g., Smith, 1977; Harris and Dallos, 1979). Thus, the recovery-rate model can not explain why the midlevel jnd elevation for a 10-ms pedestal was stronger for 10-ms than for 250-ms maskers in the experiments by Schlauch, Lanthier, and Neve (1997).

4.1.1.1.5 Backward masking

The midlevel hump observed under backward masking (e.g., Plack and Viemeister, 1992) is evidence against the recovery-rate model. In the backward-masking situation and at an ISI of 100 ms, it seems not possible that the noise altered the representation of the pedestal in the auditory nerve, because the time constant of the auditory filters is smaller than 5 ms (Kiang et al., 1965; Irino and Patterson, 2001). As this is a very serious limitation of the model, alternative explanations were suggested and will be discussed in the next chapters.

4.1.1.1.6 Masker laterality

According to the model, the mechanisms causing the jnd elevation under non-simultaneous masking are located at the level of the auditory nerve. The observation of small but significant midlevel humps for contralateral maskers (Zeng and Shannon, 1995; Plack, Carlyon, and Viemeister, 1995; Schlauch et al., 1999) is not compatible with the model and suggests a more central location as origin of the midlevel hump.

4.1.1.1.7 Standard frequency, masker frequency

Adaptation in the auditory nerve is most pronounced at adaptor (masker) frequency (e.g., Harris and Dallos, 1979). Therefore, thresholds of fibers with characteristic frequencies (CFs) away from masker frequency will be less elevated by the forward masker. These characteristics of the auditory periphery can account for the fact that the largest jnd elevation was observed if masker frequency equaled pedestal frequency (with a slightly slower decrease towards low masker frequencies, Zeng and Turner, 1992).

If masker and standard frequency are equal, a listener might compensate for the elevated thresholds of low-SR fibers with CFs equal to standard frequency by listening off-frequency (Plack and Viemeister, 1992 a). At high standard frequencies, however, off-frequency listening will not be possible because of the limited frequency range of the cochlea. Accordingly, the more pronounced midlevel jnd elevations at higher pedestal frequencies can also be explained by mechanisms located in the auditory periphery.

4.1.1.1.8 Masker spectrum

The observed effects of a two-tone masker present a problem for the recovery-rate model. Certainly, the 4.133 kHz tone added to the 1-kHz masker should cause nearly no adaptation (due to frequency selectivity of forward masking) and consequently no further jnd increase. The rather pronounced **improvement in performance** obtained with the two-tone masker, however, is incompatible with the model.

4.1.1.1.9 Noise presented simultaneously with the standard

Zeng (1998) proposed that the recovery-rate model can account for the effects of additional noise and emphasized the finding of Plack and Viemeister (1992 a) that the notched noise caused not only off-frequency but also on-frequency masking.

His explanation rests on two assumptions:

1. Limiting off-frequency listening has the maximal effect at higher standard levels, as spread of excitation is small at low levels.
2. On-frequency masking by simultaneously presented noise shifts the operating ranges of the auditory-nerve neurons towards higher levels due to suppression (Costaloupes, Young, and Gibson, 1984; Gibson, Young, and Costaloupes, 1985), which is believed to

be caused by a reduction in the vibration of the basilar membrane (Robles, Ruggero, and Rich, 1986).

The recovery-rate model plus the additional assumptions can now be used to explain the results in the different masking conditions presented in Zeng's (1998) study.

In the 50-dB spectrum level high-pass noise condition, on-frequency masking was 10 dB only, while there was strong masking at higher frequencies. Relative to the condition with forward maskers but without noise, resolution was degraded at higher levels only. This pattern is compatible with the model, according to which restricted off-frequency listening caused jnd's to be elevated at higher levels where spread of excitation plays a role, while on-frequency masking was too small to cause a significant shift in the operating ranges of the low-SR fibers, so that jnd's were unaltered at low and intermediate levels.

Similarly, in the experiment using a fixed notched-noise spectrum level, the lowest level noise (−10 dB) should have caused nearly no shift in the operating ranges, as on-frequency masking was smaller than 5 dB. Therefore, jnd's were unaltered at low levels. The masker was too also weak to mask spread-of-excitation at high standard levels, so that jnd's were unaffected in this region.

For the higher notched-noise levels, the model predicts that on-frequency masking shifts the operating ranges of all fibers (and thus the coding gap) towards higher levels and that resolution for standard intensities falling into the coding gap (40-60 dB SPL) in the condition without noise is therefore improved. At the same time, the noise restricts off-frequency listening, causing jnd's at higher levels to be elevated. This means that the model can explain the shifted position of the hump found for the 10-dB spectrum-level noise. It remains unclear, however, why the size of the hump was reduced at the two highest spectrum levels (10 and 30 dB).

In the condition where notched-noise level was always 30 dB below standard level, the operating range of the low-SR fibers can be assumed to have been shifted upwards, so that low-SR fibers always provided sufficient information. Therefore, relative to the condition without noise, jnd's remained approximately the same at low and high levels, but were reduced at intermediate levels.

The narrowband-noise caused nearly no off-frequency, but significant on-frequency masking. As Zeng varied the level of the NBN-noise so that it was a fixed amount below pedestal intensity, the same arguments as in the corresponding notched-noise condition apply.

To summarize, given the additional assumptions concerning on- and off-frequency masking, the recovery-rate model can -- at least qualitatively -- account for the data reported by Zeng (1998). It is obvious, however, that the assumed effects of the noise are quite complex, and that some of the explanations rely on a very specific trade-off between restricted off-frequency listening and shifted operating ranges. A quantitative model would be necessary to decide whether all of the effects observed by Zeng are indeed compatible with the model.

4.1.1.1.10 Notched noise presented immediately before the standard

Zeng's (1998) explanation of the notched-noise effects relies on the assumptions that the noise shifts the rate-intensity functions of all fibers to higher levels. This pattern can indeed be expected as the noise causes suppression, i.e., a reduction in the mechanical amplitude of the basilar membrane, which has effects on all groups of fibers.

Plack and Viemeister (1992 a) presented a condition in which the offset of the notched noise coincided with the onset of the pedestal in order to investigate whether suppression is a necessary condition for the noise-induced reduction of the jnd's at intermediate levels.

The notched noise presented with a spectrum level 30 dB below pedestal level can be expected to have produced adaptation at the signal frequency as well as at other frequencies. Physiological data by Smith (1977, 1979) indicate that adaptation shifts the rate-intensity function vertically, i.e. the neural response is reduced by the same amount at all signal levels. Put differently, the operating ranges were not shifted to higher levels as for simultaneous masking by noise. Abbas (1979), on the other hand, reported an increased operating range for some fibers (both low-SR and high-SR).

The reduced resolution at higher levels observed with the noise immediately preceding the standard can be explained by restricted off-frequency listening due to adaptation. To account for the reduction in the jnd's at intermediate levels, it would be necessary to assume that the noise caused a shift or an expansion of the operating range of the high-SR neurons so that the pedestal always fell into a region of the rate-intensity function above threshold but below saturation level. It seems that adaptation does not result in such a pattern, but the physiological data are inconsistent.

4.1.1.1.11 Proximal burst

The elevated detection thresholds observed with the "proximal burst" occurring 12.5-50 ms before the standard (Plack, Carlyon, and Viemeister, 1995; Plack, 1996 b) indicate that the burst caused adaptation at the level of the auditory nerve even for the high-SR neurons. For this reason, the same arguments as for the noise presented immediately before the standard apply. The recovery-rate model could account for the observed of reduction of midlevels jnd's data only by assuming a shift of the high-SR fiber's operating ranges towards higher levels. The fact that the maximum jnd reduction was observed for high proximal burst levels, at a small burst-target tone interval, and with burst frequency equaling standard frequency can be viewed as evidence for this idea. Situations producing most masking were not always those resulting in the lowest Weber fractions at midlevels, however (Plack, Carlyon, and Moore, 1995).

In any case, the finding that a burst presented after the pedestal also reduced midlevel jnd's is incompatible with a peripheral mechanism. Therefore, it can be concluded that the recovery-rate model can not account for the effects of the proximal burst.

4.1.2 Referential encoding

In the preceding chapter it became clear that there are several effects in non-simultaneously masked intensity discrimination the 'recovery-rate' model by Zeng et al. (1991) can not account for. Consequently, alternative explanations have been proposed for the midlevel hump in intensity discrimination. One of them is the referential encoding hypothesis.

Plack and Viemeister (1992 b) reported jnd-elevations under backward masking that were even larger than under forward masking. In the backward-masking situation with an ISI of 100 ms, it seems impossible that the noise altered the representation of the pedestal in the auditory nerve because the time constant of the auditory filters is much shorter (Kiang et al, 1965; Irino and Patterson, 2001). Following Mori and Ward (1992), Plack and Viemeister suggested that the backward masker in interval 1 caused an increase in "memory noise" (Durlach and Braida, 1969) by interfering with the representation of the first tone. The idea is that the trace of the first tone is degraded or overwritten by the masker presented in interval 1. Plack and Viemeister (1992 b) note that the "memory noise" explanation on itself can not account for the effect of standard level (larger jnd elevations at intermediate standard levels). However, Carlyon and Beveridge (1993) pointed out that it is possible to understand this effect if one recurs to two other mechanisms suggested by Durlach, Braida, and co-workers: context-coding and the role of internal references (cf. Braida and Durlach, 1988).

4.1.2.1 Excursus: Context-coding and the role of internal references

In their seminal series of papers on intensity resolution (starting with Durlach and Braida, 1969), Braida, Durlach, and co-workers developed, tested and refined a model of intensity discrimination which assumes that intensity resolution is determined by peripheral as well as by central factors. The model focuses on the central limitations. Several experiments (see Braida and Durlach, 1988, for a review) supported their hypothesis that in intensity discrimination, two different types of memory operation are involved.

In the "**sensory-trace mode**", the listener maintains a vivid trace of the sensation which is precise but fading quickly (i.e., a non-verbal, 'sensory' representation). According to the model, if the trace mode is used in a 2I, 2AFC discrimination experiment, performance depends on the stimuli, the listener, and the ISI between the two tones, but is independent of the range of levels used in the experiment. This view is compatible with the "perceptual diffusion model" formulated by Kinchla and Smyzer (1967).

In the "**context-coding mode**", the listener remembers a categorical/verbal representation of the sensation which is based on comparison with the context of sounds presented in the experiment or with other internal or external references. This representation is held to be less precise than the trace representation, but temporally very stable. The precision of the referential representation critically depends on the "[...] width of the context of sounds that must be compared, larger widths resulting in less precise comparisons." (Braida and Durlach, 1988, p. 560). Resolution in the context-coding mode is assumed to be the same in one-interval and two-interval experiments.

Evidence from one-interval (absolute identification, AI) experiments, where use of the trace mode is precluded, supported the view that in the context coding mode, sensitivity decreases with increasing stimulus range (e.g., Braida and Durlach, 1972). This finding was termed the "range effect". In contrast, in a "roving-level" two-interval paradigm, where different standard levels were presented in each block, resolution was independent of the range of standard levels if the inter-stimulus interval was smaller than 1 s (Berliner and Durlach, 1973).

The fact that in the one-interval paradigm, resolution was superior for stimuli at the extremes of the intensity range (the "edge effect") led the authors to conclude that the edges of the range are used as references in the coding process. This mechanism was modeled in detail in the "perceptual anchor model" (Braida et al., 1984). According to this model, "*[...] the observer codes the sensation arising from a given stimulus presentation relative to the context by determining its position relative to one or more [noisy, DO] internal references or perceptual anchors. 'Distance' measurements on the sensation axis are made in noisy discrete steps of finite mean size. The variance of a distance measurement is proportional to the distance to be measured and inversely proportional to mean step size. When more than one anchor is available it is assumed that the distance measurements to the various anchors are combined in coding the sensation*" (p. 722). In other words, resolution in context-coding mode depends on sensation noise, anchor noise, and measurement noise.

Braida et al. proposed that besides the edges of the stimulus range, detection threshold and discomfort level can be used as internal references and that additional stimuli can also serve as anchors. Berliner et al. (1977) found that in an identification experiment, presenting an explicit standard in each trial improved resolution for pedestals in the middle of the intensity range but not for pedestals at the edges. Chase et al. (1983) reported that increasing the presentation probability of tones of intermediate intensity in an AI experiment did not significantly improve resolution at midlevels and concluded that no additional anchor at the most frequently encountered stimulus seemed to have been established.

The presence of a range effect and an edge effect in roving-level intensity discrimination experiments represents strong evidence for the use of context-coding in 2I, 2AFC tasks where the ISI exceeds several hundred milliseconds. Results from Berliner et al. (1977) indicate, however, that for ISIs “ [...] less than roughly 10-15 s, the edge effect in roving-level discrimination becomes significant only when *R* [range in Bels, DO] exceeds roughly 3, and that for *R* in the vicinity of 5, it becomes significant only when *T* is at least roughly 2 s.” (p. 1581).

4.1.2.2 Context coding and the midlevel hump

To turn to the midlevel hump in intensity discrimination again, the referential-encoding hypothesis uses the concepts formulated by Durlach and Braida to explain why jnd’s are elevated at medium intensities under non-simultaneous masking. If a masker presented between the two target tones overwrites or degrades the trace of the first tone, a listener is forced to switch to context-coding mode even if the interval between the to-be-compared tones is only a few hundred milliseconds. In such case, the perceptual anchor model predicts low intensity tones to be efficiently coded relative to detection threshold and high intensity tones relative to discomfort level or masker level. For midlevel tones, however, the perceptual distance to the anchors is large, resulting in increased ‘measurement noise’.

The midlevel hump observed under backward masking can now be explained in exactly the same way as for forward masking, as the masker presented in the first interval is assumed to interfere with the trace of the first tone. As to the more pronounced hump relative to forward masking reported by Plack and Viemeister (1992 b), one would have to speculate that ‘overwriting’ the trace after a shorter silent interval (100 ms in backward masking versus 600-800 ms in forward masking) has a stronger effect. The referential encoding hypothesis in its original form can not explain this difference.

The assumed influence of trace degradation was partly corroborated by an experiment in which Carlyon and Beveridge (1993) presented a forward masker in either the first or the second interval only. Performance was worse if the masker occurred in interval 2, where it would interfere with the trace of the first stimulus. The effect was rather small, however, given the fact that in trials where the forward masker occurred in interval 1 only, there should have been no trace degradation at all. Moreover, in a condition where the mask preceded either the pedestal or the pedestal-plus-increment only, performance was better if only the pedestal was masked. Such effect is not easily understood in terms of the referential encoding hypothesis, but can be accounted for by an alternative model discussed below (Chapter 4.1.3).

The referential encoding hypothesis can account for the (small) effects of contralaterally presented maskers simply by making the reasonable assumption that the auditory store is located at a level beyond the first intra-aural interaction.

The finding of Plack and Viemeister (1992 a, b) that a notched noise presented simultaneously with the pedestal and 30 dB below pedestal intensity caused the midlevel hump to disappear can also be explained by the referential encoding hypothesis. Plack and Viemeister (1992 b) suggested that listeners might have used the notched noise as a “[...] within-interval basis for detecting an intensity increment, perhaps by a cross-frequency comparison similar to 'profile analysis'" (p. 3100), so that this additional within-interval coding reference caused the disadvantage at medium sound levels to disappear. A related observation was made by Carlyon and Moore (1986 a), who found that the midlevel hump in intensity discrimination of short pedestals at 6 kHz disappeared when the pedestal was not gated with the pedestal, but continuous. With a continuous pedestal, listeners should be able to code pedestal-plus-increment intensity relative to pedestal intensity shortly before or after the increment.

Surprisingly, the simplest situation, namely intensity discrimination in quiet, presents a problem to the referential encoding hypothesis. Equally surprising, this issue has not been discussed previously.

First, regardless of the presence or absence of masking sounds, listeners should profit from context- or referential encoding in using a combination between trace mode and context coding mode (Braida and Durlach, 1988). At first sight, one might thus conclude that the model predicts performance at low and high sound levels to be better than at midlevels **in quiet** also. This stands in sharp contrast to the near-miss to Weber's law observed in this condition. The problem can be resolved, however, by following the preliminary theory of intensity resolution in assuming that the trace mode is so effective that additionally using context coding results in no observable improvement in performance (e.g., Berliner and Durlach, 1973).

A more serious problem is related to the assumption that use of the trace mode is impossible if the interval between the to-be-compared tones is long. In other words, an inter-tone interval of, e.g., 10 s should force the listener to use context-coding, even if no masker is present. This assumption underlying the preliminary theory was corroborated by the observation of an edge effect and a range effect in the roving-level two-interval paradigm at long ISIs (Berliner and Durlach, 1973). Now, if the perceptual distance between the internal references and the target was indeed the cause of the elevated jnd's at midlevels under forward masking, a midlevel hump should be observed in quiet if long ISIs are used. Unfortunately, there seem to be no studies reporting level-dependent jnd's in a fixed-level two interval paradigm at long ISIs. For 800-ms, 1-kHz tones presented with an ISI of 2.5 s, Berliner and Durlach (1973) found no indication of a midlevel hump, but this ISI is too small to present a critical test of the hypothesis. We can use data from one-interval experiments to tackle the question, however, as use of the trace mode is impossible in this situation. In their experiment 5, Braida and Durlach (1972) measured intensity resolution for 500-ms, 1-kHz tone bursts in a small-range condition and at different tone levels. Performance could be characterized by the near-miss to Weber's law, no midlevel hump was observed. Similar results were reported by McGill and Goldberg (1968 b). Braida, Durlach, and colleagues (e.g., Braida and Durlach, 1988) attributed these findings to the formation of a precise long-term representation of pedestal intensity. In terms of their model, this formation is possible if the range of intensities presented in the experiment is small (< 10 dB).

The absence of a midlevel hump in quiet in the small-range one-interval paradigm raises the question of why listeners seem not to be able to form an equally effective long-term representation of pedestal intensity in the forward-masked intensity discrimination paradigm. In all relevant experiments, pedestal intensity was fixed within each block. The Durlach and Braida (1969) model would thus predict performance in context coding mode to be *independent* of pedestal intensity (if the near-miss is ignored for a moment), as no edge effect is observed in experiments with small intensity ranges. This stands in contrast to the referential encoding hypothesis assuming that a listener has to rely on detection threshold and discomfort level as internal coding references. Now it could be argued that the maskers create a roving-level situation even if pedestal level is fixed. Within a block, a listener encounters, e.g., a 60-dB SPL pedestal, rather similar pedestal-plus-increment levels, but also a 90-dB SPL masker, corresponding to a level range of 30 dB. It remains unclear, however, why listeners should not be able to use the edges of the intensity range (pedestal and masker intensity, respectively) as references in this situation (Berliner and Durlach, 1973), but have to rely on detection threshold and discomfort level.

A difference between the forward-masker intensity discrimination and the one-interval experiments in quiet is tone duration, which was about 25 ms in the former and 150 to 800 ms in the latter case. Unfortunately, no small-range one-interval studies using tone durations of about 30 ms seem to be available. On the other hand, it is not apparent why tone duration should have a strong effect on the pattern of results.

4.1.2.3 Referential encoding hypothesis versus data

The capability of the referential encoding hypothesis to account for the data will now be discussed.

4.1.2.3.1 Standard level

The model can account for the midlevel hump.

4.1.2.3.2 Masker level

To explain why for a fixed intermediate-level standard, difference limens increase with masker level, the reasonable assumption has to be introduced that trace degradation caused by the masker increases with masker level.

4.1.2.3.3 Masker-standard ISI

If a masker interpolated between the first tone and the second tone degrades the trace of the first tone, this can be expected to happen regardless of the tone-mask ISI. Consequently, the referential encoding hypothesis can not account for the effect of the ISI without additional assumptions. Even if the amount of trace degradation was a decreasing function of the masker-tone ISI, plausible predictions would result only for the case of a backward masker. For a forward masker, the standard presented in interval 1 and the masker presented in interval 2 are always separated by several hundred milliseconds, even if the masker-standard ISI in each interval is short.

4.1.2.3.4 Masker duration, standard duration

If masker duration has any effect at all on the amount of trace degradation, it seems likely that degradation should **increase** with masker duration. At the same time, use of the mask as a within-interval reference should become less effective due to the greater perceptual distance between the long masker and the short standard. Out of these reasons, the model can not account for the improvement in performance observed for longer masker durations (Schlauch et al., 1997, 1999).

4.1.2.3.5 Backward masking

As already discussed, the ability to explain the hump found under backward masking is one of the main advantages of the referential encoding hypothesis.

4.1.2.3.6 Masker laterality

As discussed above, the midlevel hump found with contralaterally presented maskers is compatible with the referential encoding hypothesis.

4.1.2.3.7 Standard frequency

In its original form, the referential encoding hypothesis is incompatible with the observation of a stronger midlevel hump at high standard frequencies. Plack and Carlyon (1995) proposed a solution to this problem. Against the background of neural models simulating information available in the auditory nerve, they suggested that “Perhaps we should not be asking why human performance is so good at high intensities, but rather, why it is not better at low intensities.” (p. 142). For example, the physiologically oriented model by Viemeister (1988) predicts much lower jnd’s at levels slightly above detection threshold than found in psychophysical experiments. Plack and Carlyon hypothesized that in most settings, intensity discrimination performance is limited by mechanisms central to the auditory nerve. I.e., they assumed that information available in the auditory nerve is not optimally used by the observer due to the central limitation, degrading performance at low and medium levels and resulting in the near-miss to Weber’s law. They noted that a midlevel hump is observed in circumstances where peripheral stimulus information is restricted due to the presentation of very brief sounds, high signal frequencies, or masking. Plack and Carlyon pointed out that a possible mechanism which could explain why restricted peripheral information results in a performance deficiency at medium intensities only is a level dependence of the central limitation. The inferior performance at intermediate levels associated with context coding represents such level dependence. Plack and Carlyon suggested that while “[...] in normal circumstances the memory trace is rich enough to be relatively immune to the effects of degradation” (p. 148), context coding has to be used by the listeners in the case of restricted peripheral information. Following this argumentation, presenting high frequency pedestals under non-simultaneous masking represents a two fold-degradation of the memory trace. The increased humps at higher pedestal frequencies can thus be explained.

A problem with this explanation is related to the question of why the precision of context-coding at low and high standard levels should not be impaired in situations where the memory trace is degraded. After all, some ‘early’ representation of the stimulus must necessarily be used for context-coding also. So why should context-coding of intensities at the edges of the dynamic range still work well in conditions where the trace is degraded due to, e.g., a high standard frequency? It seems that for the explanation proposed by Plack and Carlyon to work, it has to be assumed that using for example a high standard frequency or a short tone duration has a stronger detrimental effect on the memory trace than on the ‘initial representation’ used for context-coding.

4.1.2.3.8 Masker frequency

To explain why the effect of the masker was most pronounced for masker frequency equal to standard frequency, it has to be assumed that trace degradation is frequency selective. Such an assumption is certainly not implausible.

4.1.2.3.9 Masker spectrum

A similar argumentation applies to the two-tone mask that effectively reduced the midlevel hump in the experiments by Schlauch et al. (1999). Introduction of the additional frequency component should result in trace degradation either remaining the same or increasing. Use of the mask as a within-interval reference should also be either less effective due to the greater perceptual distance between the two-tone masker and the standard, or unaltered if the listener can simply ignore the additional frequency component.

To explain the beneficial effects of the two-tone masker, it would be necessary to assume that a listener greatly profits from across-frequency comparisons between masker and target.

4.1.2.3.10 Noise presented simultaneously with the standard

As discussed above, the referential encoding hypothesis can explain why the notched noise 30 dB below target level reduced midlevel jnd's in the experiments by Plack and Viemeister (1992a) by assuming that the noise was used as a within-interval coding reference.

The same argument applies to the condition of narrow-band noise presented at a fixed level difference to target level (Zeng, 1998). In terms of perceptual similarity on a spectral dimension, the narrowband noise should have been even more efficient. Zeng's observation that the reduction of midlevel jnd's was smaller for the narrowband than for the notched noise might be related to the smaller overall level of the noise, as Zeng chose the level of the narrowband noise to produce the same on-frequency masking as the notched noise.

For the fixed notched-noise condition studied by Zeng (1998), it can be argued that the low-level noise (−10 dB) provided a within-interval reference at low levels, where the coding reference detection threshold was available even without noise, so that the noise had no effect. The referential encoding hypothesis can not explain why there was a slight increase in the jnd for a 50-dB SPL standard.

The intermediate-level notched-noises (0-dB and 10-dB spectrum level) were used as references at low and intermediate levels according to the model, so that the jnd reduction at the 40 and 50 dB SPL standard levels can be accounted for. The noise-induced deterioration of performance observed with the 30-dB noise at higher standard levels is incompatible with the model. It should be noted, however, that Plack and Viemeister (1992a) reported no increase at higher levels in the presence of a 30-dB spectrum level notched noise.

The intense high-pass noise should not have provided an efficient coding reference at intermediate levels. But this does not explain why jnd's were elevated at higher standard levels, so that a peripheral explanation (restriction of off-frequency listening) seems necessary in this condition.

4.1.2.3.11 Notched noise presented immediately before the standard

In the condition in which the offset of the notched noise coincided with the onset of the pedestal (Plack and Viemeister, 1992 a), the reduction in the jnd's at intermediate standard levels can be explained using the same argumentation as for the noise presented simultaneously with the target tones. It is only necessary to assume that the noise did not cause significant additional trace degradation, which seems reasonable if trace degradation is again assumed to be frequency selective.The referential encoding hypothesis can not account for the reduced resolution at higher standard levels, however. Again, it seems necessary to assume a peripheral mechanism (restricted off-frequency listening due to adaptation) here.

4.1.2.3.12 Proximal bursts

In terms of the referential encoding hypothesis, there are two potential effects of the proximal burst, as discussed by Plack, Carlyon, and Viemeister (1995) and Plack (1996 b):

First, due to the temporal proximity, the burst could be used as an efficient coding reference. This hypothesis is corroborated by the observation of Plack, Carlyon, and Viemeister (1995) that in the forward-masked intensity discrimination setting without proximal bursts, midlevel jnd's were smaller at very short masker-standard intervals (< 50 ms) than at longer ISIs. The authors argued that the finding of Zeng and Turner (1992), who reported midlevel jnd's to maximum at the shortest ISI could be due to the use of a narrowband-noise masker in the Zeng and Turner study. A narrowband noise signal exhibits random level fluctuations which should render its use as a coding reference less effective. As context coding is a central mechanism, the jnd reduction observed with proximal bursts presented after the standard presents no problem to the model.

As the perceptual distance between reference and target tone determines coding efficiency according to the perceptual anchor model (Braida et al., 1984), the observation of maximum jnd reduction for burst frequency equaling target frequency is compatible with the model.

Out of the same reason, however, the fact that the maximum jnd improvement occurred at proximal burst levels exceeding standard level is incompatible with the anchor model, according to which a same-level reference should be most effective (Plack, 1996 b).

The second potential explanation of the proximal burst effects is that partial masking by the proximal burst reduces the loudness of the target tone. Therefore, the loudness of a midlevel standard presented together with the burst would be similar to the loudness of a low-level standard presented without burst. Put differently, partial masking reduces the perceptual distance to detection threshold and thus increases the efficiency of referential encoding.

As already discussed in Chapter 4.1.1.1.11, the masking hypothesis can account for the observations that maximum jnd reduction occurred with intense proximal burst levels, at a small burst-target tone interval, and with burst frequency equaling standard frequency. Situations producing most masking were not always those resulting in the lowest Weber fractions at midlevels, however (Plack, Carlyon, and Viemeister, 1995). Additionally, the intensity jnd **increases** at low levels in quiet. Therefore, the explanatory value of the masking hypothesis seems questionable.

4.1.2.3.13 Midlevel jnd elevation in a loudness matching task

Zeng (1994) and Plack (1996 a) obtained loudness matches between a tone preceded by a forward masker and an unmasked comparison tone (Fig. 5). They used a 2I, 2AFC adaptive procedure with interleaved upper and lower tracks converging on the 71% and 29% “louder” points on the psychometric function, respectively (Jesteadt, 1980). As Schlauch and Wier (1987) pointed out, these points can be used as “[...] essentially symmetrical estimates of the just-detectable increment and decrement, respectively” (p. 14). Half the difference between the two values can be taken as a measure of loudness variability (Zeng, 1994). Plack (1996 a) used the same measure and termed it “jnd”. The so-defined jnd’s were elevated at intermediate standard levels with a forward masker level of 90 dB SPL. These findings are difficult to understand in terms of the referential encoding hypothesis, as no sound is presented in the interval between the first and the second target tone (Fig. 5). Therefore, as trace degradation can not present a problem in this situation, no midlevel hump should be observed according to the referential encoding hypothesis.

4.1.3 Loudness enhancement hypothesis

As demonstrated in Chapter 3, the effects of stimulus parameters such as masker-target ISI, masker frequency, and masker laterality on the size of the midlevel hump in intensity discrimination and on loudness enhancement are quite similar. This observation led Carlyon and Beveridge (1993) to propose that the midlevel elevation of intensity jnd’s under non-simultaneous masking might be caused by loudness enhancement.

The stimulus configuration used in forward-masked intensity discrimination studies closely resembles the stimuli in loudness enhancement experiments, as a brief target tone is preceded by a more intense sound (cf. Fig. 4 and Fig. 5). Carlyon and Beveridge presented evidence for the forward mask introducing a systematic bias: listeners appeared to overestimate the loudness of sounds preceded by the mask in an experiment where only one observation interval contained a masker, as performance was superior if only the pedestal-plus-increment was preceded by the mask than if only the pedestal was masked.

To explain why loudness enhancement should impair intensity discrimination in a situation where **both** the pedestal and the pedestal-plus-increment are preceded by a masker, Carlyon and Beveridge suggested that loudness enhancement introduces **variability** in the perception of the pedestal. In the following, this idea is termed the loudness enhancement hypothesis.

Data by Plack (1996 a) support their hypothesis. He used a 2I, 2AFC adaptive procedure with two interleaved tracks converging on the 70.7 and 29.3 points of the psychometric function, respectively (Jesteadt, 1980; Schlauch and Wier, 1987). With this method it is possible to simultaneously measure loudness enhancement and loudness variability. Plack found the two measures to be significantly correlated (Spearman rank correlation coefficient r_s = 0.5-0.76) for 3 of the 4 listeners. Zeng (1994) reported loudness enhancement, loudness variability, and intensity jnd's for the same listeners. All of the three measures showed a very similar dependence on target level, being maximum at intermediate stimulus levels (midlevel hump).

To explain the midlevel hump by loudness enhancement has the advantage to explicitly account for the apparent similarity between the two phenomena. On the other hand, Plack (1996 a) has argued that explaining increased jnd's by loudness variability is in some respects a "truism", as the intensity jnd is a measure of variability in perceived intensity. In fact, the only exception seems to be the case where a response bias rather than variability causes the level difference necessary for the listener to reach a certain level of performance to be large. Plack (1996 a) proposed to rather "[…] *ask why loudness is variable in conditions that produce loudness enhancement*" (p. 1024). In this study, the problem will be treated together with an associated problem: even if the proposed cause-effect relation was correct, it would still be unclear which mechanism causes loudness enhancement. In the new model introduced in Chapter 5, an explanation for loudness enhancement will be proposed. The effects of loudness enhancement on intensity resolution will be discussed in terms of a signal-detection model in Chapter 7.2.

Are there any existing models that can explain the effects of loudness enhancement on intensity resolution? Certainly, many attempts have been made in relating intensity perception and intensity discrimination, an effort dating back to Fechner (1860). Fechner's law predicts that relative jnd's (Weber fractions) should follow the derivative of loudness with respect to sound intensity (i.e., the slope of the loudness function). Equivalently, increasing sound intensity by one jnd should produce a constant increment in loudness. Such a model is not compatible with the data. For example, Hellman, Scharf, Theghtsoonian, and Theghtsoonian (1987), presenting tones in quiet and in noise, found that intensity jnd's were unrelated to the slope of the loudness function. Their results supported the "equal-loudness, equal-jnd" hypothesis first formulated by Zwislocki and Jordan (1986) that predicts jnd's to be equal at equal loudness levels (Schlauch and Wier, 1987; Johnson et al., 1993). Subsequent studies showed, however, that the model can not account for loudness and intensity resolution of pure tones in the presence of a continuous high-pass masking noise (Schlauch, 1994), gated masking noise (Rankovic et al., 1988) and for listeners with certain types of sensorineural hearing losses (Florentine et al., 1993). Applied to loudness of pure tones under non-simultaneous masking, Zeng (1994) pointed out that the model must be rejected simply because there are no stimulus levels in quiet for which the large jnd's observed under non-simultaneous masking are found.

A theory that can account for the relation between loudness and intensity resolution in background noise is the "proportional-jnd theory" by Lim et al. (1977), which is based on a hypothesis formulated by Riesz (1933). According to this model, two sounds are matched in loudness if their intensities divide the two dynamic ranges proportionally in terms of jnd's. To use an example from Houtsma et al. (1980), if the dynamic range for sound 1 contains 100 jnd's, the dynamic range for sound 2 contains 200 jnd's, and the level of sound 1 is 24 jnd's above threshold for sound 1, then the level of sound 2 producing a loudness match is 48 jnd's above threshold for sound 2 according to the proportional-jnd theory. In other words, the proportional-jnd theory postulates a close relation between loudness and the jnd just as Fechner (1860), but the Fechnerian assumption predicts loudness to be a function of the jnd count *without* normalization. The proportional-jnd theory was successfully used to account for loudness comparisons between tones presented at different frequencies (Lim et al., 1977), and for the relation between loudness matches and intensity discrimination for tones in quiet, tones in background noise (Houtsma et al., 1980; Rankovic et al., 1988; Schlauch, 1994; Schlauch, Harvey, and Lanthier, 1995), and noise stimuli (Houtsma et al., 1980).

Zeng (1994) used the model to predict the loudness function for the forward masked target tone from intensity discrimination data obtained in his experiment. These results will be discussed in Chapter 4.2.1. Just as the Fechnerian assumption, the proportional-jnd theory can also be used to predict intensity jnd's from loudness matching data. According to Eq. 8 in Lim et al. (1977),

$$\frac{\beta_q\, \delta'(\mathrm{I}_q)}{\beta_m\, \delta'(\mathrm{I}_m)} = \frac{\mathrm{I}_q\, \mathrm{M}'_{q,m}(\mathrm{I}_q)}{M_{q,m}(\mathrm{I}_q)}, \qquad (4\text{-}1)$$

where $M_{q,m}(\mathrm{I}_q)$ denotes intensity of the masked tone (target) that is matched in loudness to a tone in quiet (comparison) with intensity I_q, $M'_{q,m}(\mathrm{I}_q)$ is the first derivative of the matching function with respect to I_q, δ' denotes sensitivity-per-Bel (Durlach and Braida, 1969), and β_q and β_m are the sensation variances, which are assumed to be independent of intensity.

Up to a multiplicative constant, Eq. 4-1 predicts resolution in forward masking to be resolution in quiet divided by the slope of the loudness matching function with comparison tone intensity expressed in decibels plotted on the *x*-axis and target tone intensity expressed in decibels plotted on the *y*-axis (Houtsma et al., 1980, p. 808). Now, the intensity jnd (ΔL_{DL}) corresponding to $d' = 1$ is just 10/δ' (Braida and Durlach, 1988). Therefore, ΔL_{DL} in forward masking is ΔL_{DL} in quiet multiplied by the slope of the matching function according to the model.

If one considers, e.g., the loudness matching data provided by Zeng (1994; see his Fig. 2), it is evident that the observation of a midlevel hump in loudness enhancement in the presence of a 90-dB SPL masker is equivalent to the slope of the matching function increasing with target level. As the increase in slope overcompensates for the slight decrease in ΔL_{DL} in quiet (the near-miss), the proportional-jnd theory predicts jnd's to increase montonically with stimulus level. This pattern is incompatible with the data.

Another model proposed for the relation between loudness and intensity resolution is a generalization of the McGill and Goldberg (1968 a, b) model. This generalization formulated by Zwislocki and Jordan (1986) and Hellman and Hellman (1990) assumes that sensitivity expressed as d' is proportional to the loudness difference ΔN between two tones divided by the standard deviation of the decision variable,

$$d' = k\, \Delta N/\sigma_N. \qquad (4\text{-}2)$$

Zwislocki and Jordan (1986) and Hellman and Hellman (1990) introduced an approximation based on the Taylor-series expansion of the loudness function, writing

$$d' = k\frac{N'(I)\,\Delta I}{\sigma_N}, \qquad (4\text{-}3)$$

where N'(I) is the first derivative of the loudness function with respect to signal intensity I, and ΔI is the intensity difference between the two tones. According to the Poisson assumption made by McGill and Goldberg, $\sigma_N = k_2\sqrt{N}$. Following McGill and Goldberg, Hellman and Hellman (1990) actually based their model on neural counts rather than on loudness. As they assumed loudness to be a linear function of the neural count, however, this model is equivalent to the loudness-based formulation used in Eq. 4-3.

The model can be used to predict intensity jnd's in forward masking from loudness matches and intensity jnd's in quiet. For tones in quiet, the loudness function can be computed according to the model by Zwislocki (1965) that will be discussed in greater detail in Chapter 7.1.1. The model predicts loudness of a tone in quiet to depend on sound pressure P of the tone and on sound pressure P_{Th} at detection threshold as

$$N = K\,[(P^2 + 2.5\,P_{Th}^2)^\theta - (2.5\,P_{Th}^2)^\theta], \qquad (4\text{-}4)$$

where K is a scale constant. From this function and the loudness matches between the forward masked target tone and the unmasked comparison tone obtained in loudness matching experiments, the loudness function for the masked tone can be derived (Zeng, 1994). As intensity jnd's represent intensity differences ΔI corresponding to a constant value of d', the jnd function in quiet can be used to predict jnd's in forward masking (i.e., intensity differences corresponding to the same d' as in quiet). From Eq. 4-3 and the assumed Poisson-relation between loudness and its standard deviation, it follows that

$$\Delta I_{DL\text{-}m}(I) = \frac{N_q'(I)\sqrt{N_m(I)}}{N_m'(I)\sqrt{N_q(I)}}\,\Delta I_{DL\text{-}q}(I), \qquad (4\text{-}5)$$

where $\Delta L_{DL\text{-}m}$ denotes the just-noticeable intensity increment of the masked tone at standard intensity I, $\Delta L_{DL\text{-}q}$ is the just-noticeable intensity increment of the tone in quiet at the same intensity, $N_q(I)$ and $N_m(I)$ are loudness in quiet and in forward masking, respectively, $N_q'(I)$ is the first derivative of the loudness function in quiet with respect to signal intensity I, and $N_m'(I)$ is the first derivative of the loudness function in forward masking with respect to I (Hellman and Hellman, 1990; Schlauch, Harvey, and Lanthier, 1995).

The loudness matching data by Zeng (1994) were again used to evaluate the model. The loudness function in quiet (comparison tone) was computed using Eq. 4-4. According to Zeng and Shannon (1995), detection threshold was 17 dB SPL for the listeners in Zeng (1994), although this value seems rather high for the 30-ms, 1-kHz pure tones presented. The parameter θ of 0.27 estimated by Zwislocki (1965) for the magnitude estimation data by Hellman and Zwislocki (1963) was used. The scale parameter K cancels out in Eq. 4-5. To construct the loudness function of the masked tone (target), mean loudness matches between the masked target tone and the unmasked comparison tone reported by Zeng (1994) were interpolated by a third-order polynomial. For each level of the masked tone, the level of the tone in quiet producing the loudness match was computed. The latter level was then used as input to the loudness function in quiet (Eq. 4-4). The derived loudness function for the masked tone is shown by the dashed line in Fig. 13, left panel. Jnd's in quiet reported by Zeng and Shannon (1995; cf. Fig. 5 in Zeng, 1994) were interpolated by a linear function. The dashed line in Fig. 13, right panel, shows predicted difference limens in quiet according to Eq. 4-3. The constant k was selected to minimize the squared deviation between observed and predicted jnd's. Jnd's in forward masking (predicted by Eq. 4-5) are shown by the solid line in Fig. 13, right panel. Unlike the models discussed up to now, the McGill and Goldberg model indeed predicts the jnd to be a non-monotonic function of standard level. If one compares the predicted pattern to the midlevel hump found by Zeng and Shannon (1995; cf. Fig. 5 in Zeng, 1994), however, it is obvious that the predicted hump is smaller (4 dB) and located at a higher standard level (near 80 dB SPL).

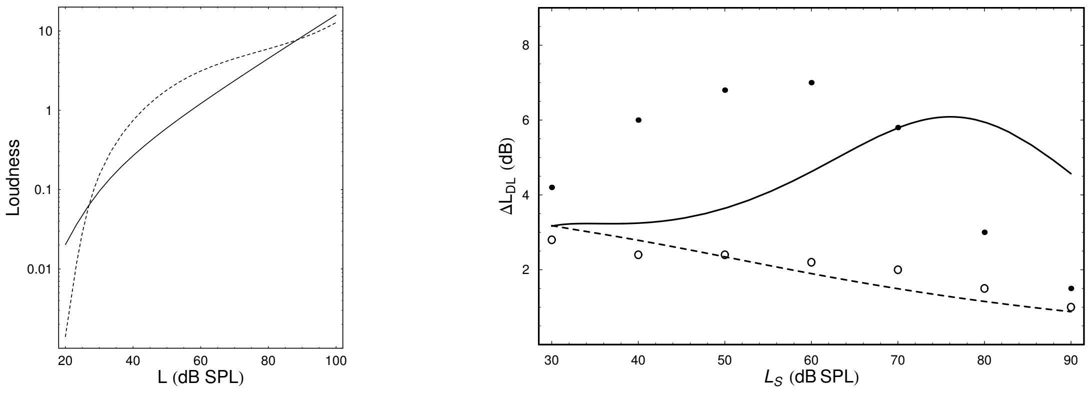

Fig. 13: McGill and Goldberg model applied to the data by Zeng (1994). Left panel: Loudness function in quiet according to Zwislocki (1965; solid line), loudness function of masked tone derived from matching function (dashed line). Right panel: Observed jnd's *(open circles: in quiet, closed circles: in forward masking), predicted ΔL_{DL} in quiet (dashed line), and predicted ΔL_{DL} in forward masking (solid line).*

It can be concluded that the models proposed for the relation between loudness and intensity resolution, especially the proportional-jnd theory, can explain many peripheral effects as masking by background noise or using tones of differing frequencies. The models fail to account for the presumable more central effects of non-simultaneous masking, however.

4.1.3.1 Loudness enhancement hypothesis versus data

The loudness enhancement hypothesis simply predicts that jnd's are elevated where loudness is enhanced. As discussed in Chapter 3, there is a close correspondence between the amounts of loudness enhancement and jnd elevation for variations of target and masker level, the ISI between masker and target, masker frequency relative to target frequency, and the laterality of masker and target. Both phenomena are observed under both forward and backward masking, but size of the effects differs. It is also unclear if the dependence of intensity jnd's on masker duration applies to loudness enhancement also.

Above that, there are several parameters for which no loudness enhancement data have been reported, as for the effects of additional stimuli.

Explicit correlations between loudness enhancement and loudness variability were reported in only one study (Plack, 1996 a).

An additional point follows from the model. If loudness enhancement was indeed the cause of the midlevel hump, what could be expected for conditions in which loudness **decrement** is observed

(Zwislocki and Sokolich, 1974; Elmasian and Galambos, 1975)? If a change in perceived intensity induced by the masker increases variability in the representation of intensity, a jnd increase should result in these conditions, too. Difference limens for masker levels smaller than standard level were measured in an experiment by Zeng and Turner (1992), where no jnd elevation was observed in this situation. In Experiments 1, 2, and 3 of this thesis, the masker-target level difference will be varied in order to induce different amounts of loudness enhancement/loudness decrement and to measure the correlation between loudness enhancement and the jnd.

4.1.4 Summary

The *recovery-rate model* certainly represents an attractive model of the effects of forward masking on intensity discrimination as it attempts to explain the findings by properties of the auditory periphery. The model can indeed account for a range of findings, as for example the effects of masker-standard ISI and masker frequency on the midlevel hump or, given some additional assumptions, the effects of noise presented simultaneously with the standard. The fact that the model can not explain the midlevel hump observed under backward masking and with contralaterally presented maskers is a very serious limitation, however, that is reason to reject the model. Before this background, the failure to account for the reduced midlevel jnd's observed with longer masker durations and two-tone maskers is only additional evidence against the model.

The *referential encoding hypothesis* is an important alternative to the recovery-rate model as it can account for effects that must be caused at more central locations (e.g., midlevel humps observed with a backward masker and with a contralaterally presented masker). Some basic effects (masker level, masker frequency) can be explained if additional plausible assumptions are introduced. The most important effects of presenting additional stimuli are compatible with the model. As for the recovery-rate model, it must be concluded, however, that it seems necessary to consider both peripheral and central mechanisms in order to account for the whole range of findings (Zeng, 1998). The referential encoding hypothesis can not account for the effects of masker duration and masker waveform. It also has difficulty explaining the smaller midlevel humps found for long masker-standard ISIs, at least for the case of a forward masker. Above that, the increase in loudness variability caused by a forward masker in a loudness matching task is difficult to explain in terms of the model, as no masker is interpolated between the two target tones in such an experiment. Summarizing the discussion presented in Chapter 4.1.2.2, it has to be concluded that the most serious objection against the model seems to be the discrepancy between intensity resolution in quiet observed in a small-range one-interval paradigm and difference limens found in non-simultaneous masking.

The *loudness enhancement hypothesis* represents the only model proposed for the midlevel hump in intensity discrimination that explicitly accounts for the similarity between loudness enhancement and intensity resolution in non-simultaneous masking. It is therefore important for the discussion of a common mechanism underlying the two phenomena and will be included in the new model introduced in Chapter 5. The few differences existing in the effects of parameters on the two phenomena are evidence against the model, but problems as serious as for the previous two models do not seem to exist. It has to be noted, though, that the data basis that can be used for evaluation of the model is smaller, as several parameters have not been studied for loudness enhancement. Experiments 1 to 3 of this thesis are designed to measure intensity jnd's and loudness enhancement for the same stimuli and listeners, so that the prediction of a correlation between the two measures can be tested. The data will also provide further evidence as to the effect of masker duration and allow conclusions concerning the relation between loudness decrement and difference limens.

4.2 Explanations for loudness enhancement

4.2.1 Models of the relation between intensity resolution and loudness

As discussed in Chapter 4.1.3, several models have been proposed for predicting loudness from intensity discrimination data.

Zeng (1994) used the proportional-jnd theory (Lim et al., 1977) to compute the loudness function for the forward masked target tone on the basis of intensity jnd's obtained in his experiment. For comparison, he first constructed the empirical loudness function for the tone in forward masking by combining the loudness balance function (masked target tone level versus unmasked comparison tone level) and the loudness function for a 1000-Hz pure tone in quiet (Hellman and Zwislocki, 1963). This function shows a steeper growth of loudness at low levels, but is flatter than in quiet at intermediate levels, so that the two functions intersect at a target level of approximately 90 dB SPL (Fig. 13, left panel, dashed line). The proportional-jnd theory predicts the slope of the loudness function to be inversely proportional to the size of the jnd. Due to the midlevel hump in intensity discrimination, the predicted loudness function is steeper at low than at intermediate levels, which is compatible with the empirical loudness function for the masked tone described above. The slope of the function predicted by the proportional-jnd theory is also greater at high than at intermediate levels, however (Fig. 4 in Zeng, 1994), which is incompatible with the loudness function derived from the loudness matches.

Essentially the same argument applies to the Fechnerian assumption of loudness increasing by a constant amount if stimulus level is increased by one jnd.

Schlauch (1994) pointed out that an "equal-jnd, equal-loudness" model formulated as the 'inverse function' of the "equal-loudness, equal-jnd" hypothesis by Zwislocki and Jordan (1986) makes meaningless predictions, as for example the jnd for broadband noise is independent of level, while loudness of course increases with level.

For the generalized McGill and Goldberg model, Hellman and Hellman (1990, see their Eq. 14) have shown that loudness N(I) of a pure tone presented with intensity I is predicted to be

$$\sqrt{N(I)} = k \int_{I_{\mathrm{Lower}}}^{I} \frac{1}{\Delta I(x)}\,\mathrm{d}x + \sqrt{N(I_{\mathrm{Lower}})}, \qquad (4\text{-}6)$$

where I_{Lower} is the lower limit of integration (i.e., the lowest intensity for which a jnd is available), k is a constant, and $\Delta I(I)$ is the intensity increment corresponding to a given value of d' at standard intensity I. From Eq. 4-6 it is obvious that the square root of loudness is predicted to be proportional to the integral of the inverse of the intensity-jnd function $\Delta I(I)$. Therefore, the slope of the square-root-of-loudness function at a given signal intensity is inversely proportional to the jnd at that intensity. In other words, the generalized McGill and Goldberg model predicts the slope of the loudness function to be monotonically related to the inverse of the jnd, just as the proportional-jnd theory (cf. Eq. 4 in Lim et al., 1977). For this reason, the same arguments as for the latter theory apply, and it can be concluded that the McGill and Goldberg model can not account for the relation between intensity difference limens and loudness in forward masking.

4.2.2 Physiological findings

In physiologically oriented models of intensity coding, loudness is generally assumed to be a function of the (weighted) sum of the spike counts of auditory nerve neurons (e.g., Fletcher and Munson, 1933; Lachs and Teich, 1981).

The specific relation between auditory nerve activity and loudness remains subject to discussion (e.g., Relkin and Doucet, 1997; Chatterjee and Zwislocki, 1998). It seems safe to assume, however, that loudness increases monotonically with total spike count.

At early stages of auditory processing, forward maskers have almost exclusively been found to cause adaptation, i.e., to reduce the neural response to the test tone (e.g., Bauer and Galambos, 1975; Smith, 1977; Harris and Dallos, 1979; Kramer and Teas, 1982; Relkin, Doucet, and Sterns, 1995; Murnane, Prieve, and Relkin, 1998). Bauer, Elmasian, and Galambos (1975) measured both brainstem evoked responses and loudness enhancement. They found that wave V amplitudes were not enhanced in situations where loudness was enhanced. On the contrary, they observed significant adaptation in some cases. The single contradicting result was presented by Ananthanarayan and Gerken (1983), who also recorded auditory brainstem responses in humans and reported wave *V* amplitude to be enhanced at masker-target ISIs between 5 and 45 ms.

At later stages of auditory processing, response enhancement by preceding sounds has been reported in several studies. Not quite surprisingly, the first observations were made in echolocating bats at an ISI of 10 ms (Suga, O'Neill, and Manabe, 1978). Subsequently, similar patterns were reported for other mammals and for humans (e.g., McKenna, Weinberger, and Diamond, 1989; Nishimura et al., 2003). To give an example, Loveless, Hari, Hämäläinen, and Tiihonen (1989) used MEG to record evoked fields to pairs of sounds. They presented pairs of 50-ms duration, 80-dB SPL noise bursts . The first sound of each pair can be viewed as the conditioner and the second sound as the target in a loudness enhancement experiment. The inter-onset interval (IOI) between the first and the second sound was varied between 70 and 500 ms, the inter-pair interval was 1.2-1.4 s. For IOIs smaller than 300 ms, the amplitudes of *N*100*m* to the *second* stimulus of a pair were larger than the amplitudes of *N*100*m* evoked by the first stimulus. Subsequent studies confirmed these findings (Loveless and Hari, 1993; Loveless et al., 1996; McEvoy, Levänen, and Loveless, 1997).

Budd and Michie (1994) observed a similar pattern in the auditory ERP (event-related potential). Stimuli were 1000-Hz tones at 80 dB SPL and separated by random ISIs between 100 and 1000 ms. The *N*100 peak was stronger if the silent interval preceding a tone was 100-300 ms rather than 500-1000 ms.

Brosch, Schulz, and Scheich (1999) recorded tone-evoked responses from single units und multiunit groups in macaque auditory cortex. Responses to 100-ms tones were enhanced by a preceding tone of the same intensity in a substantial proportion of units and groups, although a reduction in response was also found frequently. Maximum enhancement was observed for an ISI of 125 ms and a frequency separation of one octave. The latter finding is incompatible with the psychophysical data showing maximum loudness enhancement and jnd-elevation for same-frequency masker and target (Zwislocki and Sokolich, 1974).

Note that in all above studies, response enhancement was observed in pairs of identical tones. In psychophysical measurements, loudness enhancement for identical conditioners and targets was either found to be absent or to be very small (e.g., Zwislocki and Sokolich, 1974; Elmasian and Galambos, 1975). This discrepancy might be related to the fact that with a maximum inter-pair interval of 1.5 s, the sequences presented in the neurophysiological studies are within the tempo range for which tone sequences are perceived as rhythmic (Fraisse, 1982). Therefore, the stimuli used in the experiments more closely resemble the short-long sequences described in Chapter 2.4 than the three-tone sequences used in loudness enhancement experiments.

McEvoy, Levänen, and Loveless (1997) presented tone pairs separated by a fixed ISI of 210 ms and varied the inter-pair interval in the range between 0.6 and 8.1 s. Evoked fields to the first tones of the tone pairs were smaller than responses to single tones presented with the same ISI. The response to the second tones was enhanced relative to the response to the first tones only at an inter-pair interval of 0.6 s, while it was substantially smaller at longer intervals. It could be argued the data are compatible with the psychophysical results insofar as in the three-tone paradigm, where a pair of sounds is followed by a moderately long ISI (500-1500 ms), the comparison tone and then a relatively long inter-trial interval corresponds to inter-pair intervals greater than 2 s in the experiment of McEvoy et al., where no enhancement of the fields evoked by the second tones was observed. The short-long sequences presented by Vos (1977) and Povel and Okkerman (1981) on the other hand correspond to the shortest inter-pair intervals, where response enhancement was present.

Generally, it remains questionable if the neuromagnetic responses and loudness perception are correlated, as in none of the physiological experiments, loudness enhancement was measured. Another important issue is that if loudness indeed closely followed the neural response, the results by McEvoy, Levänen, and Loveless (1997) indicate that loudness of the first tone should significantly exceed loudness of the second tone in pairs of identical tones presented with a long inter-pair interval. Such pattern has not been observed in studies measuring the time order error, however (see Hellström, 1985).

To summarize, it seems precluded that loudness enhancement is caused by early mechanisms, while some neural responses observed at more central locations are not completely incompatible with the psychophysical data.

To my knowledge, no study that reported neural response enhancement also varied conditioner level, so that nothing can be concluded about the effect of the level difference between conditioner and target on the neural responses. In this context, it is important to note that the idea of an integration of 'neural energy' seems misleading. If remaining activation resulting from the first stimulus were simply added to the activation associated with the second stimulus, loudness enhancement should be observed regardless of conditioner level. This stands in sharp contrast to the results of Elmasian, Galambos, and Bernheim (1980), and Zwislocki and Sokolich (1974) who observed loudness **reduction** in trials where conditioner level was below target level. As no physiological data seem to have been collected using masker levels below target tone level, the conclusions that can be drawn from the reports of response enhancement are limited. It would be desirable to record evoked fields of potentials in a setting similar to the three stimulus configuration common in loudness enhancement studies (Chapter 2.3) and to compare the amplitudes of responses to target tone and comparison burst at conditioner levels both above and below target tone level. Such data would present a critical test of the hypothesis that loudness enhancement is related to response enhancement in central locations.

4.2.3 Recovery-rate model

On the basis of the assumption that loudness is a monotonically increasing function of the (weighted) sum of the spike counts from high-SR and low-SR neurons, could the neuronal-recovery-rate model by Zeng, Turner, and Relkin (1991, cf. Chapter 4.1.1) explain loudness enhancement? As discussed above, the physiological data show that the rate-level functions of auditory nerve fibers are either shifted to the right or simply downwards by a forward masker (e.g., Abbas, 1979). Both effects result in spike count under masking being smaller than in quiet. It can thus be concluded that the recovery-rate model generally predicts loudness reduction rather than loudness enhancement (Zeng, 1994).

4.2.4 Referential encoding hypothesis

As formulated by Carlyon and Beveridge (1993) and Plack, Carlyon, and Viemeister (1995), the referential encoding hypothesis can account for intensity resolution only.

Within the framework of the model, a potential explanation for loudness enhancement would be that the memory representation of target loudness is biased towards the coding reference. If the conditioner was used as a within-interval coding reference, an intense conditioner would bias the memory representation for the target towards higher intensities, i.e., loudness enhancement would be observed.

Results from experiments by Berliner, Durlach, and Braida (1977) speak against the hypothesis that the memory representation of the first tone ‘drifts’ towards reference loudness, however. They measured intensity discrimination in a roving-level two-interval paradigm (in quiet). No feedback was provided. Systematic response biases were observed at ISIs of 3.5 s and 14 s. At low standard levels, listeners tended to respond “First tone louder”, while at high standard intensities, there was a bias towards responding “Second tone louder”. As detection threshold or the lower edge of the intensity range is assumed to be used as a reference at low levels and discomfort level or the upper edge of the intensity range at high levels, it appears that memory was in fact biased **away** from reference loudness towards the middle of the intensity range.

Another critical issue is that the proposed effect should increase with the level difference between conditioner and target. A 90 dB SPL conditioner should cause more ‘drift’ for a 30 dB SPL target than for a 60 dB SPL target. In other words, the observation of a midlevel hump is incompatible with the idea.

4.2.5 Mergence hypothesis

Elmasian, Galambos, and Bernheim (1980) attributed loudness enhancement to “mergence”. They argued that “*[...] overlapping in the brain causes some of the processes engendered by the conditioner to be interpreted as belonging to the target [...]. As a result, the final percept of the target is approximated by a weighted average of the separate sensations each interactor would produce if presented alone.*” (p. 606). They also introduced a memory-related aspect in stating that mergence might be effective for memory processes.

The concept of mergence was based mainly on their finding that loudness is reduced if the conditioner is less intense than the target, while a more intense conditioner results to enhancement. In other words, target perception always seems to be shifted towards conditioner perception.

The mergence hypothesis can explain why loudness is enhanced at masker levels above target level and reduced at masker levels below target level.

If the weight masker loudness receives decreases with the temporal separation between masker and target, the model can also explain the decrease in loudness enhancement with increasing ISI. The observation of a non-monotonic time course, where loudness enhancement at very small ISIs is greater than at intermediate ISIs could only be accounted for by assuming that the masker reduces loudness of the target by causing adaptation in the auditory nerve.

The influence of the difference between masker and target frequency is compatible with the model if one assumes that masker loudness receives a smaller weight if masker and target are separated spectrally. The mergence hypothesis can *not* account for the midlevel hump in loudness enhancement, however. Given a constant masker level of, e.g., 90 dB SPL, loudness enhancement should increase monotonically with decreasing target level. In Chapter 5, an extended version of the mergence model will be formulated that is not subject to this limitation.

4.2.6 Concept related to mergence: Trace drift

If two sounds presented sequentially are compared for intensity, systematic biases have been reported since Fechner (1860). The "Zeitfehler"/time error/time order error (TOE) as defined by Fechner is positive if loudness of the first stimulus is overestimated and negative if it is underestimated. In the method of constant stimuli, the stimuli are frequently presented in the temporal order standard-comparison. In such case, the TOE corresponds to constant error

$$CE = PSE - \Phi_S, \quad (4\text{-}7)$$

where PSE is the physical magnitude of the comparison that is judged to just as loud as the standard and Φ_S is the physical magnitude of the standard. Most often, **negative** TOEs have been reported (i.e., the first stimulus is underestimated; cf. Woodworth and Schlosberg, 1954). More specifically, at high stimulus intensities, TOEs were more negative than at low stimulus intensities, where positive TOEs were found in some experiments (e.g., Needham, 1935).

Hellström (1985) discussed time order error in an extensive review. He concluded that response bias (e.g., a preference for the response louder) can not account for the TOE. This conclusion was based on data from experiments in which listeners judged either the first or the second stimulus, presentation order was either standard-comparison or comparison -standard, and different response formats were required (Hellström, 1977, 1978). Instead, Hellström attributed TOE to "assimilation": he proposed that "[...] introduction of a (relatively long) ISI causes a regression of the memorial magnitude of the first-presented stimulus towards its mean" (p. 41).

This explanation is similar to the quantitative theory of TOE formulated earlier by Michels and Helson (1954). According to their model, the veridical impression of the second stimulus is compared to a memory representation of the first stimulus, the so-called "comparative adaptation level" (CAL). CAL was assumed to be a weighted average between the perceived magnitude of the first stimulus and the "series adaptation level" (SAL), which represents the geometric mean of all preceding stimulation. Such "regression towards the mean" view is compatible with the idea that the trace of the first stimulus drifts towards the middle of the stimulus range in a roving level intensity discrimination experiment (Berliner, Durlach, and Braida, 1977). It can also explain why in roving level experiments, TOE was more negative at higher stimulus levels than at lower stimulus levels, where in some cases even positive TOEs were observed. One should note, however, that in Michels and Helson's model, depending on the specific weights, also a contrast effect could result (i.e., the trace drifts **away** from the reference value).

As mentioned, the geometric mean of preceding stimulation and the middle of the stimulus range were proposed as adaptation levels. It also seems reasonable to assume a proximity effect: stimuli closely preceding or following the signal should have greater impact on the adaptation level than temporally distant sounds. For example, the fighter jet that just flew over your head should dominate the adaptation level effective for the next bird you hear singing, while the relative silence of the Negev desert you experienced just before that jet now lost its influence. Hellström (1985) noted that **loudness enhancement and decrement** can be explained in this context if one assumes that the conditioner influences SAL_T (the adaptation level effective for the target). If so, the remembered loudness of the target would drift towards conditioner loudness during the target-comparison burst interval.

This concept is very similar to the mergence hypothesis. If in the forward masking situation, SAL_T is influenced by the conditioner, the drift of the remembered loudness for the target towards its adaptation level results in CAL_T (the memory representation of target loudness used in the comparison) being a *weighted average* between conditioner loudness and target loudness. Essentially the same argument applies to settings where the conditioner follows the target ('backward masking'): the tone interpolated between target and comparison influences SAL_T and therefore causes the trace of the target to drift towards conditioner loudness. For both conditions, the explanation rests on the assumption that the SAL for the target is influenced more strongly by the conditioner than the adaptation level effective for the comparison tone.

To summarize, the basic idea behind the assimilation/trace drift concept put forward in the context of research on the time order error is essentially equivalent to the mergence hypothesis, as the loudness representation of the target effective in the loudness matching task is assumed to be a weighted average between masker loudness and target loudness, with additional effects of average loudness encountered during the experiment.

4.2.7 Summary

Physiological findings preclude peripheral mechanisms as the cause of loudness enhancement, while response enhancement observed at more central locations is not incompatible with the behavioral effects. On the other hand, there are several inconsistencies between the neurophysiological and the behavioral findings, so that the explanatory value of the physiological findings remains questionable.

The only model capable of accounting for both loudness enhancement and decrement is the mergence hypothesis by Elmasian, Galambos, and Bernheim (1980). In its original form, this model has to be rejected, however, as it is incompatible with the midlevel hump in loudness enhancement. A modified version of the mergence hypothesis that is not subject to this shortcoming will be formulated in Chapter 5.

4.3 Explanations for the interval-produced accent

The phenomena found in short-long sequences are interesting because they are not easily understood against the background of existing models of intensity perception. In the following, the relevance of empirical and theoretical aspects from the context of loudness enhancement, the midlevel hump in intensity discrimination, and a new aspect based on a model of processing in auditory sensory memory will be discussed.

4.3.1 Loudness enhancement

In the short-long sequences, listeners report the second sounds of the pairs (i.e., sounds initiating long IOIs) to be more intense than the first tones. This could be interpreted as loudness enhancement of the second tone of a pair by the first tone.

Consequently, Povel and Okkerman (1981) discussed the relation between loudness enhancement and the interval-induced accent. They falsely stated, however, that no loudness enhancement had been found for conditioner and target tone of the same frequency. Their reference (Zwislocki, Ketkar, Cannon, and Nodar, 1974) applied only to the situation where conditioner and target were of the same frequency **and** of the same intensity. In a situation where the conditioner was 20 dB more intense than the target, Zwislocki and Sokolich (1974) demonstrated that loudness enhancement was strongest if conditioner frequency was *equal to* or slightly below target frequency. The more important point is the second argument brought forward by Povel and Okkerman: either no or only a small amount of loudness enhancement was found if conditioner and target were of the same intensity, frequency, and duration (cf. Chapter 3.2). Povel and Okkerman (1981) reported loudness matches between the first and second tones of the tone pairs for one listener. For short intra-pair intervals, the level difference required to make the first tones sound equally loud as the second tones was as large as 4 dB. This value is larger than the loudness matches of 1-3 dB between the second tone of the pair (target) and the comparison tone found in the three-tone matching task for identical conditioner and target (e.g., Zwislocki and Sokolich, 1974; Elmasian and Galambos, 1975).

At this point, however, it is important to note a fundamental difference between a three-tone matching task and an accent-perception experiment. In the former case, the listener compares loudness of the masked target tone to loudness of a tone “in quiet”, as the comparison tone is separated from the target by a silent interval of at least 500 ms. In the short-long sequences, loudness of a “backward masked” tone (the first tone of a pair) is judged relative to loudness of a “forward masked” tone (the second tone of the pair). For this reason, loudness matches obtained in the two experiments are not equivalent.

First, from an accent reported on the second sounds of the pairs, it is principally impossible to infer whether the first tone *enhanced* loudness of the second tone or whether the second tone *reduced* loudness of the first tone. This argument is important for the discussion of an alternative model, the interrupted-processing hypothesis, presented below.

Second, one could argue that the task of comparing loudness of the two sounds comprising a pair might be more sensitive to small loudness differences than the three-tone task, in which loudness of the second tone of the pair is compared to the comparison tone separated by a longer interval. This would especially be the case if loudness of the first tone was indeed reduced by the second tone. In the three-tone matching task using a forward masker, this effect can not be detected, as the listener compares only second tone and comparison tone loudness. Results from a three-tone experiment by Elmasian and Galambos (1975) can be taken as evidence for a reduction of first tone loudness. Recall, that in their loudness matching task, the authors found enhancement to be only 1-2 dB if forward conditioner level was equal to target tone level (70 dB SPL). The target was the second tone of the tone pair conditioner-target. In their experiment 4, the same listeners were tested again with the same stimuli, but using a different procedure. Instead of producing a loudness match between target and comparison tone, listeners rated the loudness of all of the three tones (conditioner, target, and comparison). The loudest tone was to be scored a “3”, the softest a “1”, and ties were allowed. In the condition where the level of all three tones was identical (70 dB SPL), mean scores were 1.3, 2.2, and 2.5 for conditioner, target, and comparison, respectively. These results indicate that, compatible with the loudness matches, the loudness difference between target and comparison tone was small. Instead, loudness of the first tone (the conditioner) was significantly *reduced* in this condition. On the other hand, loudness enhancement has also been studied in a “backward masking” setting. In these experiments, listeners matched the loudness of the first tone of a trial (the target) to loudness of the last tone (the comparison), while the level of the second tone (the conditioner) was varied. Now, for a target equal in level to the conditioner presented 100 ms after the target, Elmasian, Galambos, and Bernheim (1980) reported up to 4 dB of loudness enhancement for some listeners, while loudness of the target was reduced for other listeners. Clearly, these loudness matches do not indicate a general reduction of tone 1 loudness by the second tone.

Loudness of two sounds comprising a pair is compared directly in experiments measuring loudness matches in quiet, i.e., the time order error. As discussed in Chapter 4.2.6, the level of the second tone matching loudness of the first tone was frequently found to be *below* first-tone level (Hellström, 1985). These results indicate either a reduction of first tone loudness or an enhancement of second tone loudness, just as for the interval-produced accent. The TOE data are *not* compatible with the accent, however, because the overestimation of second-tone loudness *increased* with the silent interval between first and second tone (Fig. 1 in Hellström, 1979). In contrast, the interval-produced accent on the second tones of the pair is strongest at small ISIs (Povel and Okkerman, 1981).

To summarize, the small discrepancy between the ‘enhancement’ of the second tone of a pair found in the three-tone matching task and the short-long sequences, respectively, can not be taken as evidence against loudness enhancement as the mechanism underlying the accent, as the two tasks are not equivalent. Above that, results for the short-long sequences are available only from one experiment using exactly one listener. A more important argument against the idea that loudness of the second tone is enhanced by the first tone are the results of the rating experiment conducted by Elmasian and Galambos (1975), which indicate loudness of the first tone being reduced relative to a tone in quiet. On the other hand, loudness matches in “backward enhancement” between the first tone and the third tone of a trial did not show a reduction of tone 1 loudness if conditioner and target were identical.

As pointed out above, neither the proportion of accents reported on the second sounds, nor the loudness match between first and second sounds can be used directly to decide whether the accent is due to reduced loudness of the first sounds or to enhanced loudness of the second sounds. Conclusions can be drawn from the time course or the level dependence of the accent, however.

Loudness enhancement decreases with the silent interval between conditioner and target (at least for ISIs larger than 50 ms; cf. Chapter 3.3.1). This pattern is compatible with the dependence of the interval-produced accent on the intra-pair interval (e.g., Fig. 8, left panel). The effect of the *inter*-pair interval on the accent (Fig. 8, right panel) can also be explained by loudness enhancement. For identical intra- and inter-pair intervals, the first tone of a pair enhances loudness of the second tone. The second tone, however, causes exactly the same amount of enhancement for the first tone of the following pair if intra- and inter-pair interval are identical. A loudness difference between first and second tones can result only if the interval separating the tone pairs is larger than the intra-pair interval.

If loudness enhancement was the mechanism underlying the accent, the maximum ISI at which the two phenomena are observed should also be similar. The maximum silent interval between conditioner and target at which loudness enhancement was found is at least 400 ms (Zwislocki and Sokolich, 1974). Povel and Okkerman (1981) concluded that accents were reported on the second sounds of a pair of sounds only if the within-pair offset-onset-interval (OOI) did not exceed 250 ms. As pointed out above, some of the data contributing to this estimate were collected in a condition where the difference between the long and the short ISI *decreased* with short interval duration in the experiment. As the temporal asymmetry between short and long IOI is essential for the interval-produced accent, the negative correlation between intra-pair interval duration and the interval difference can be assumed to have biased the estimate towards a smaller value. A more reliable estimate of the temporal limit of accent perception will be obtained in Experiment 4.

Loudness enhancement also depends on presentation level. Thus, the observation of, e.g., a midlevel hump in the interval-produced accent would be strong evidence for a relation between the accent and loudness enhancement. Experiment 5 will study the effect of presentation level.

We now turn to two *theoretical* arguments against loudness enhancement as an explanation for the interval-produced accent.

First, substantial loudness enhancement was found in three-tone sequences where the conditioner *followed* the target (‘backward masking’; Elmasian and Galambos, 1975; Elmasian, Galambos, and Bernheim, 1980; Plack, 1996 a). Now, if forward and backward enhancement were exactly *symmetric* (i.e., of the same magnitude and following the same time course), loudness enhancement would cause no loudness difference between the first and the second sounds. Enhancement of the second sound by the first sound would be compensated by enhancement of the first sound by the second sound. To decide whether the temporal asymmetry in loudness enhancement is sufficiently large to account for the interval-produced accent, measurements of the time course of loudness enhancement under forward and backward masking precise to at least 1 dB would be necessary. On the basis of the data available to date, no definitive conclusion seems possible.

Second, if loudness enhancement indeed caused the accent, a model of loudness enhancement itself would be desirable. The most successful model, according to the discussion in Chapter 4.2, is the mergence hypothesis. Yet, this model predicts the *absence* of loudness enhancement when conditioner and target are *identical* because a weighted average of conditioner and target loudness is simply equal to target loudness in this situation.

4.3.2 Referential encoding hypothesis

As already discussed for loudness enhancement, referential encoding can not account for systematic changes in perceived intensity.

4.3.3 Trace drift

The idea proposed by Hellström (1979, 1985) and Berliner, Durlach, and Braida (1977) that the auditory trace of a stimulus drifts toward some reference level (e.g., the middle of the intensity range) seems at first sight capable of accounting for the interval-produced accent. One necessary assumption is that in the short-long sequences, the listener compares only the temporally most proximal sounds, i.e., loudness of the first and the second tone of each pair, but makes no comparison between the second tone of a pair and the first tone of the following pair. According to this assumption, the task performed by the listener is simply a multiple-looks loudness comparison. We know from results on time order error that in the case of loudness comparisons between two tones in quiet, the trace of the first tone seems to drift towards a reference level during the intra-pair interval. If it is now additionally assumed that the adaptation level is lower than the loudness of the tones (e.g., the weighted mean of sound intensity and the quiet background; Hellström, 2000), loudness of the first tone should be **underestimated** relative to loudness of the second tone, as its trace drifts towards the lower reference level during the silent intra-pair interval. This mechanism would result in the perception of an accent on the second tone. According to this hypothesis, however, the effect should **increase** with the ISI between the first and the second tone of a tone pair, as the amount of trace drift increases over time (Hellström, 1979). This prediction stands in clear contrast to the effects found in Experiment 4 and in the experiments conducted by Povel and Okkerman (1981), where the second tones were perceived as accented more frequently if the interval between the two tones was **short**.

Moreover, any comparison between sounds separated by the long IOI (e.g., second sound of pair one and first sound of pair two) would result in the second tone of a pair being underestimated relative to the first tone according to the trace drift mechanism.

The trace-drift hypothesis can therefore be rejected as a model of the interval-produced accent.

4.3.4 Interrupted processing in auditory sensory memory

Povel and Okkerman (1981) suggested that in the short-long sequences, perception of the second sounds as louder than the first sounds is the result of interrupted processing in auditory sensory memory. The authors assumed that an auditory sensation needs time to build. After transformation by the auditory receptor system, a ‘trace’ of the sound is stored and processed in sensory memory. Magnitude of sensation (as well as the amount of information extracted from the stimulus) grows during processing and reaches an asymptote after about 300 ms. In a short-long sequence, processing of the first sound of a pair is terminated by input of the second sound. Processing of the second sound is in turn interrupted by input of the first sound of the next pair. As a result, all second sounds can be processed for a longer time than the first sounds because the inter-pair interval (IOI_L) is longer than the intra-pair interval (IOI_S, Fig. 7). As loudness is assumed to be an increasing function of processing time, loudness of the second sounds therefore exceeds loudness of the first tones (Fig. 14).

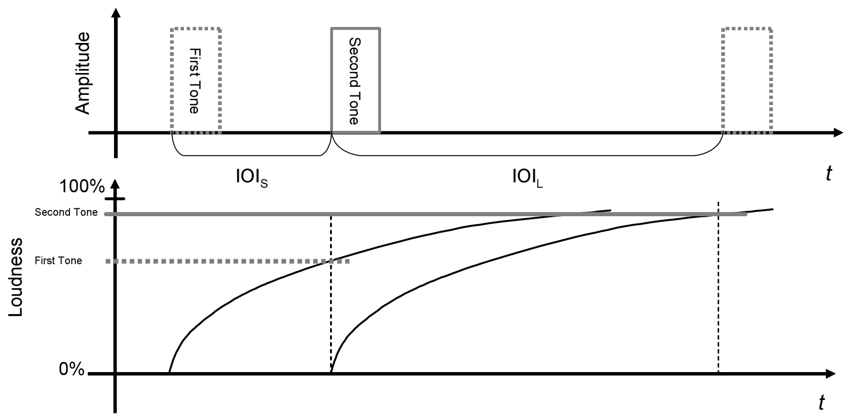

Fig. 14: Loudness in a short-long sequence according to the interrupted-processing model. Black curves: growth of sensation. Presentation of the next stimulus interrupts processing. Therefore, processing of the first tone is interrupted earlier than processing of the second of each pair, as $IOI_S < IOI_L$. Consequently, loudness of the second tone (solid gray line) is greater than loudness of the first tone (dashed gray line).

Note that from the interrupted-processing hypothesis, exactly the same pattern as for loudness enhancement as the cause of the accent follows: loudness of the second tones is expected to exceed loudness of the first tones. The interrupted-processing hypothesis attributes this effect to a completely different mechanism, however, because it assumes loudness of the first tone to be reduced rather than loudness of the second tone to be enhanced. As discussed above, results from a rating experiment by Elmasian and Galambos (1975) are evidence for this hypothesis, while loudness matches obtained in 'backward masking' experiments, where the conditioner followed the target, no significant reduction of tone 1 loudness was observed in the condition of identical target and conditioner. Again, neither the proportion of accents reported on the second sounds, nor the loudness match between first and second sounds can be used directly to decide between the two models, however. It is rather necessary to interpret for instance the time course or the level dependence of the two measures.

Povel and Okkerman's theorizing refers to the "stage model of auditory information processing" formulated by Massaro (1970, 1975). In this model, sensory or "preperceptual" storage plays an important role. According to the model, the stimulus is transformed by the auditory receptor system. A detection process then stores a preperceptual representation in the "preperceptual auditory store" (PAS, Fig. 15). The time needed for detection decreases with stimulus intensity. Processing during storage in PAS results in the recognition of features (e.g., pitch). Massaro calls this process "primary recognition". The result is a "synthesized percept" which is stored in "synthesized auditory memory" (SAM). Categorical labeling or naming of synthesized percepts is done by a process termed "secondary recognition" and results in a representation in "generated abstract memory" (GAM), which is more or less identical to short-term or working memory.

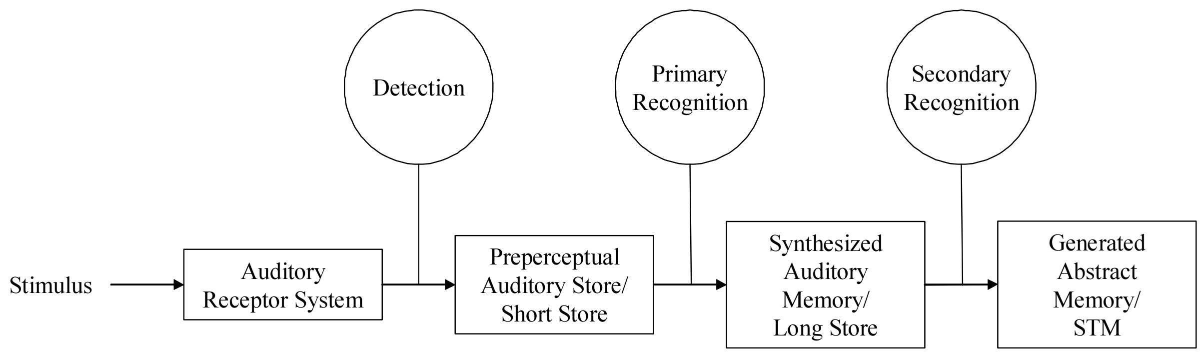

Fig. 15: Stage model of auditory information processing (Massaro, 1975).

The important assumption is that processing of a sound in PAS is interrupted by input of any following sound because the sensory trace is overwritten. In several auditory backward recognition masking (ABRM) experiments, Massaro and his coworkers demonstrated that recognition of sound attributes such as loudness, pitch, and duration were impaired if another sound was presented within approximately 250 ms after presentation of the relevant stimulus (see Massaro, 1975, for a review). The paradigm was an absolute identification task (AI) where participants identified a stimulus followed by a backward masker in each trial (e.g., identification of a pure tone as 'high' or 'low' in pitch). Performance improved monotonically with the interval between tone and masker and reached an asymptotic value at ISIs greater than 250 ms. These findings can be explained by assuming that the input of the next stimulus overwrites the stored information and interrupts processing and that complete analysis (feature recognition) of a stimulus in PAS takes approximately 250 ms.

According to Massaro, the next incoming stimulus does not completely overwrite the information associated with the first stimulus at the third stage of processing, SAM. Rather, new information can result in what Kallman and Massaro (1979) termed "memory confusion": information associated with the second stimulus is falsely attributed to the first stimulus. Evidence from a pitch discrimination experiment conducted by Kallman and Massaro partly supported this concept.

The notion of the existence of two distinct sensory stores in the auditory modality was supported by Cowan (1984). He concluded from widely varying estimates of the duration of auditory storage in the literature (100 ms to 20 s) that there must be two types of auditory sensory memory. According to his review, the "short auditory store" (p. 341) decays within about 300 ms, is indexed by experiments on masking, auditory persistence, and temporal integration, and can be interpreted as an unanalyzed, vivid representation of the stimulus which is experienced as and extension of sensation. The "long auditory store" (p. 341) contains partly analyzed information, retains them for several seconds, and is experienced as a vivid recollection (Cowan, 1988). This type of storage is indexed in auditory partial report (Darwin, Turvey, and Crowder, 1972; cf. Massaro, 1975, for a critical discussion) and suffix-effect experiments (Crowder and Morton, 1969). In Cowan's (1988) concept of memory storage, only the short store is a separate system, whereas the long sensory store is part of STM. According to Massaro (1975), Cowan (1984), and Massaro and Loftus (1996), the *short* auditory store is closely linked to sensation (e.g., auditory persistence), whereas phenomena related to the *long* auditory store are more likely experienced as memory. Against this background, the **short** store rather than the long store should be relevant for the effect of temporal sequential structure on intensity perception.

Povel and Okkerman (1981) applied the interrupted processing concept to perceived intensity. They proposed that in the short-long paradigm perception of the second sounds as louder is the result of interrupted processing in sensory memory because loudness is an increasing function of processing time and the second sound can be processed for a longer time than the first sound, as the inter-pair interval (IOI_L) is longer than the intra-pair interval (IOI_S).

A study by Cowan (1987) supported this notion. A standard tone was presented 750 ms before a pair of sounds. Listeners judged the loudness of the standard tone as well as the loudness of the first sound of the pair (the target) using a rating scale. All sounds were 40-ms 700-Hz pure tones presented at 83, 85, or 87 dBA. Standard and target intensity were always equal. The level of the third tone (the masker) was varied independently of target level. Loudness judgments of the first tone of the pair increased as the IOI between target and masker increased from 50 to 200 ms. Cowan interpreted these results as to support the hypothesis that neural activity resulting from the sound is integrated temporally and that a mask presented shortly after the target interrupts this process. He proposed that single stimuli are analyzed within a time span corresponding to the time constant of the short auditory store. If another stimulus arrives before the end of that time span, processing of the first stimulus is interrupted. Loudness is proportional to the degree of processing.

A similar idea was promoted by Zwicker and Fastl (1999) as an explanation for temporally masked loudness, i.e., the reduction in loudness of a test tone by an intense noise masker presented after offset of the test tone. The authors concluded that "[…] the effect indicates that our hearing system needs some time to develop the sensation of loudness" (p.219). In their data, loudness of the tone reached an asymptote at a tone-masker ISI of 150 ms.

It can be concluded that the interrupted-processing model is principally capable of explaining the perception of accents on the second tones of the sound pairs in short-long sequences. The question remains whether the assumed characteristics of preperceptual auditory storage are compatible with the effects of the temporal sequence structure on the interval-produced accent. Povel and Okkerman (1981) interpreted their data as to demonstrate the maximum intra-pair-interval at which an accent on the second tones is observed to be 250 ms. This value would be compatible with the time constant of 250-300 ms of the short auditory store (Cowan, 1984). As discussed above, however, the value of 250 ms most likely underestimates the temporal limit of accent perception due to the negative correlation between intra-pair interval duration and the difference between short and long ISI in the relevant experiment. In Experiment 4 of this thesis, the accent is measured in a design where this correlation is not present. Generally, the estimation of time constants from psychophysical data is not without pitfalls. Many problems can be avoided by using a mathematical model for the estimation, however. In Chapter 8, such a quantitative model will be used to obtain reliable estimates of the temporal characteristics of the interval-produced accent.

As an additional test between loudness enhancement and interrupted processing as two potential mechanisms underlying the interval-produced accent, the effect of presentation level is studied in Experiment 5. This parameter should have no effect according to the interrupted-processing hypothesis.

4.3.5 Physiological Findings

If there is a temporal asymmetry in a rhythmic sequence because the interval between the sounds of a pair is smaller than the interval between pairs, adaptation in the auditory nerve recovery could potentially cause perception of the second sounds to be different from perception of the first sounds. For same-level sounds and ISIs longer than 100 ms, however, there should be nearly no interaction between the first and the second tone of a pair at early stages of auditory processing.

As discussed in Chapter 4.2.1, several studies reported an enhancement of the neural response to the second tone of a pair of identical sounds at locations central to the auditory nerve. The stimuli used in these experiments closely resemble the rhythmic sequences presented by Vos (1977) and Povel and Okkerman (1981). Intra-pair intervals varied between 70 ms and 1000 ms, inter-pair intervals were 1.2-1.5 s in most studies. The fact that the response to the second tones was stronger than the response to the first tones at intra-pair intervals smaller than 300 ms (Loveless et al., 1989) is compatible with the behavioral data reported by Povel and Okkerman (1981).

MEG data by Loveless et al. (1996) could be interpreted as evidence against the interrupted processing hypothesis, though. They presented short-long sequences of 50-ms 1-kHz tones. The intra-pair interval was varied between 70 and 500 ms while the inter-pair interval varied randomly between 1.2-1.4 s. Neuromagnetic responses to the second tones decreased with increases in the intra-pair interval, while amplitude and latency of responses to the first tones remained approximately constant. This is exactly the opposite of what the interrupted processing hypothesis predicts, as a variation of the intra-pair interval affects processing time of the first tone but not of the second tone of a pair.

4.3.6 Summary

Identifying loudness enhancement as the cause of the interval-produced accent would be fantastic in terms of parsimony, as it would allow interpreting loudness perception in the rhythmic short-long sequences simply as a ‘multiple looks’ variant of loudness perception in a single pair of tones, without requiring to recur to completely different mechanisms. It was pointed out that the small discrepancy between the amounts of loudness enhancement observed for identical masker and target and the magnitude of the interval-produced accent can not be taken as an argument against loudness enhancement, as the two tasks are not equivalent. A more important point are indications that the time course of loudness enhancement differs from the time course of the accent, but the estimate of the temporal limit of accent perception proposed by Povel and Okkerman is likely to be biased.

The interrupted processing model can account for the effects of temporal sequence structure. Still, it remains unclear whether the temporal limits of accent perception are compatible with the assumed time constant of the short auditory store.

In Experiment 4 of this thesis, data will be obtained that allow arriving at a more reliable estimate of the temporal characteristics of the interval-produced accent. For deriving the estimate, a quantitative model based on the concept of interrupted-processing is used. This approach many some of the pitfalls associated with the estimation of a time constant from simple psychophysical 'markers', like for example the ISI at which the percentage of accents on the second sounds drops to 50 %. Therefore, the model can also be used to derive a more reliable estimate of the time constant from the data by Povel and Okkerman (1981).

Presentation level produces a characteristic pattern of results for loudness enhancement and intensity resolution. The interrupted processing hypothesis, on the other hand, predicts no effects of this parameter. The influence of presentation level on the interval-produced accent is studied in Experiment 5.

5 The “similarity model” - Effects of a masker on loudness and intensity resolution

Summarizing the foregoing review of empirical evidence and models, it can be concluded that intensity jnd’s under non-simultaneous masking and loudness enhancement share several important features. For example, a midlevel hump was observed for both phenomena. Plack (1996 a) also reported a significant correlation between loudness variability and the amount of loudness enhancement.

Models proposed for the two phenomena differ in their ability to account for the empirical evidence.

The recovery-rate model, while being attractive because it attempts to explain the midlevel hump by basic properties of the auditory periphery, can not account for the humps found under backward masking and under contralateral stimulation. Moreover, it can not explain the effects of masker duration and the improvement in performance by a two-tone masker.

Accounting for the effects of backward and contralateral stimulation presents no problem to the referential-encoding hypothesis that is based on more central mechanisms. The model can also explain beneficial effects of within-interval references like notched noise. It fails to account for the effects of masker duration, however and it remains unclear whether the superior performance observed with a two-tone masker can be understood in terms of across-frequency comparisons. The model has serious difficulty accounting for the effects of the masker-standard ISI. Most important, however, it seems to be incompatible with the fact that no midlevel hump is observed in one-interval paradigms.

Loudness enhancement as a cause of the midlevel jnd elevation is an attractive explanation because it explicitly accounts for the similar effects many parameters have on the two phenomena by assuming a cause-effect relation. The loudness enhancement hypothesis does not provide an explanation for loudness enhancement itself, however. Above that, it has to be noted that some parameters have different effects on the jnd elevation and loudness enhancement (e.g., the effect of forward maskers compared to backward maskers).

Neither the recovery-rate model not the referential encoding hypothesis can account for loudness enhancement. In the auditory periphery, an intense forward masker should cause --if anything-- a reduction in loudness due to adaptation. The referential encoding hypothesis could explain loudness enhancement only by assuming that in context-coding, the intense masker is used as a reference and memory for the target tone is biased towards anchor loudness. Results by Berliner et al. (1977) indicate a bias in the opposite direction, however.

The mergence hypothesis proposed by Elmasian, Galambos, and Bernheim (1980) as an explanation for loudness enhancement can elegantly account for the effects of weak and intense maskers on perceived intensity. As such, it appears to be the only model capable of explaining at least the very basic phenomenon of loudness enhancement. Regarding the question of a common mechanism underlying the two phenomena, many effects could be explained by combining the mergence hypothesis (which accounts for loudness enhancement) and the loudness enhancement hypothesis that assumes that the jnd will be elevated in situations were loudness is enhanced. Unfortunately, though, the mergence hypothesis predicts loudness enhancement to be a monotonically increasing function of mask-target level difference. Therefore, the important observation of a midlevel hump is incompatible with the model. In the following, a new model will be presented that is based on a modified version of the mergence hypothesis and that predicts a midlevel hump.

5.1 Effects of masker-target similarity on the weight assigned to masker loudness

I propose that by extending the mergence hypothesis by an important aspect, it is possible to arrive at a combined model (mergence hypothesis plus loudness enhancement hypothesis) that can account for most effects demonstrated empirically and that is also compatible with the idea of a common mechanism underlying the two phenomena.

The additional parameter included in the new model is **masker-target similarity.** If one recalls the discussion in Chapter 4, it becomes clear that several observations that are problematic to explain in terms of the existing models can be described as *similarity effects*. The most obvious case is the effect of the two-tone masker used by Schlauch and colleagues (1999) that reduced the size of the midlevel hump in intensity discrimination. Closely related are the effect of masker duration (250-ms masker causes a smaller hump for 10-ms pedestal than a 10-ms masker does) and the smaller effect of a diotic forward masker on both loudness enhancement and difference limens (compared with an ipsilateral masker).

In the "similarity model" proposed here, the original mergence hypothesis is extended by assuming an effect of the *perceptual similarity* of masker and target. The effect of this modification is that masker loudness receives a smaller weight in the computation of the weighted average if masker and target are perceptually different (e.g., in spectral content, duration, or loudness).

As an example, consider perceptual similarity on the loudness dimension for the case of masker and target differing in presentation level only. For a 90-dB SPL masker and a 90-dB SPL target, the weighted average of the two perceived intensities is simply identical to the loudness of each interactor presented alone (*no enhancement*), although perceptual similarity and thus the weight assigned to masker loudness can be assumed to be maximum in this condition. The similarity model predicts a 30-dB SPL target and a 90-dB SPL masker to be perceptually too different for pronounced mergence to occur (only small weight assigned to masker loudness, *nearly no enhancement*). A 60-dB SPL target and a 90-dB SPL masker, however, are assumed to be sufficiently similar for masker loudness to receive significant weight (*pronounced enhancement*).

Obviously, extending the mergence hypothesis by assuming a similarity effect results in the prediction of a midlevel hump in loudness enhancement. According to the loudness enhancement hypothesis, the midlevel hump in intensity discrimination follows from the increased variability in perceived intensity caused by loudness enhancement.

5.1.1 Masker-target similarity: Previous considerations

The assumed effects of masker-target similarity introduced here can be linked to several previous observations and considerations. For example, Schlauch, Lanthier, and Neve (1997) hypothesized: "A possible explanation for the additional masking noted with short duration maskers than for longer duration ones is confusion caused by a perceptually similar masker" (p. 465*f*).

One of the best-established findings in cognitive psychology is the effect of target-distractor similarity, for example in experiments studying recognition memory (see Baddeley, 1997, for a review), where the detrimental effect of distractors on performance is smaller if target and distractor items differ phonologically or semantically.

Similarity also plays a role in less "symbolic" tasks involving memory, like pitch discrimination. Associated with the third stage of Massaro's (1975) model discussed in Chapter 4.3.4, synthesized auditory memory (SAM), is the concept of memory confusion or memory interference (Kallman and Massaro, 1979). According to this concept, if a subject responds on the basis of information stored in SAM, he or she in some trials confuses information associated with another stimulus (e.g., a mask or distractor) with the relevant target information. Massaro (1975) explicitly assumes that perceptual similarity plays an important role in memory confusion so that the probability of memory confusion decreases with perceptual conditioner-target differences. A frequency discrimination experiment conducted by Kallman and Massaro (1979) supported this hypothesis. The standard was always presented in the first interval, followed by the test tone (higher or lower in frequency), followed by a backward masker. Results indicated a similarity effect at short test tone - masker intervals. If for example test tone frequency was lower than standard frequency and the test tone was followed by a mask slightly higher in frequency, listeners were biased to respond 'test tone frequency higher', resulting in a decrease in percent correct values (Fig. 16, solid dots). Such bias was not observed if standard and masker frequency differed by more than 30%, however.

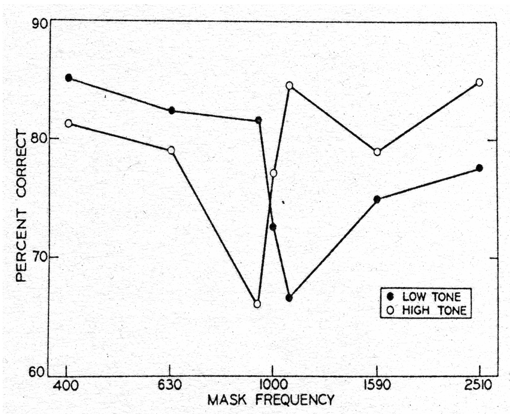

Fig. 16: Percentage of correct identifications of low and high test tones. Standard frequency: 1000 Hz. Low test tone frequency (filled circles) was slightly below 1000 Hz. High test tone frequency (open circles) was slightly above 1000 Hz. The masker introduced a bias to respond, e.g., 'test tone higher' if its frequency was slightly above test tone frequency ('error imbalance', Hawkins and Presson, 1977). At large frequency differences, the effect of the masker was small, indicating a similarity effect. From Kallman and Massaro (1979).

According to the referential encoding hypothesis, the effect of the forward masker is to force the listener to switch to context coding mode. The memory representation used for context-coding is 'non-sensory' or 'verbal' (Durlach and Braida, 1969) and thus corresponds to SAM in the Massaro model. Therefore, if the probability of a memory confusion increased with the level difference between masker and target, the small jnd's found at low standard levels could be explained by a similarity effect.

Another auditory phenomenon in which similarity plays a prominent role is sequential streaming. If alternating tone sequences (ABAB…) are presented in which tones A and *B* are made increasingly different in one ore more dimensions, listeners perceive these stimuli no longer as one single but as two separated sequences or "streams" (cf. Bregman, 1999). Van Noorden (1977) demonstrated that level differences as small as 5 dB can result in streaming if the rate of presentation is fast (ISI < 200 ms), indicating that the effect of perceptual similarity on the loudness dimension assumed in the similarity model is plausible.

5.2 Similarity-model versus data

The similarity model predicts any manipulation causing masker and target to be perceptually less similar to result in decreased loudness enhancement and therefore also a smaller jnd elevation. For this reason, the model can account for the smaller effects of using a two-tone masker as well as presenting masker and target at different frequencies or contralaterally. The fact that a diotic masker resulted in a slightly smaller amount of loudness enhancement than an ipsilateral masker in an experiment by Elmasian and Galambos (1975) is also evidence for an effect of perceptual similarity, as the stimulation in the ear receiving the target was identical in the two conditions. For a comparable stimulus configuration, Plack, Carlyon, and Viemeister's (1995) reported a smaller jnd elevation effected by a diotic masker than by an ipsilateral masker. In other settings (e.g., backward masking at short standard-masker intervals), the diotic masker had a stronger effect than the ipsilateral masker, however.

Maskers differing in duration from the standard also resulted in significantly smaller midlevel humps in intensity discrimination, while the effect of masker duration on loudness enhancement was small but followed the same pattern.

Obviously, separating masker and target temporally should also reduce the amount of interaction assumed by the mergence hypothesis, so that the decrease of both loudness enhancement and the jnd elevation with increases in the masker-target ISI is compatible with the model. The findings that loudness enhancement was smaller at a masker-target interval of 10 - 30 ms than at an ISI of 70 - 150 ms (Zwislocki and Sokolich, 1974) and that the jnd elevation caused by a forward masker was more pronounced at ISIs of 12.5-25 ms than at 100 ms (Plack, Carlyon, and Viemeister, 1995) is at odds with the assumed effect of 'perceptual proximity' on the amount of mergence. Yet, it could be argued that at the smallest ISIs, the masker caused adaptation. In fact, Plack et al. (1995) found a 10 to 15 dB increase in detection threshold at an ISI of 12.5 ms. As loudness can be assumed to be proportional to total spike count (e.g., Fletcher and Munson, 1933; Lachs and Teich, 1981), adaptation should result in loudness of the target being reduced. For the intense masker combined with a midlevel target presented by Plack et al. (1995), reduction of target loudness results in a decrease in the loudness similarity between masker and target, so that the similarity model predicts a smaller amount of loudness enhancement in this situation. Certainly, this explanation is speculative and rests on the assumption that the increase in masker weight due to the reduced temporal separation between masker and target is overcompensated by the decrease in loudness similarity due to adaptation. Additionally, the reduction in loudness of a 50-dB SPL tone caused by a 15-dB increase in detection threshold can be expected to be quite small due to loudness recruitment (Zwislocki, 1965). For intensity resolution, the jnd reduction at small ISIs might also be related to reduced variability in the neural responses. Smith (1977) stated that "[...] short-term adaptation reduces both the presumed noise, i.e., spontaneous activity, and the response to the test signal. As a consequence, the neural signal-to-noise ratio may actually increase if based only on the response of a single auditory fiber." (p. 1111). This explanation is valid for the forward masking situation only, as a backward masker causes no adaptation in the auditory nerve.

Unfortunately, the effects of several parameters have not been studied for both loudness enhancement and intensity discrimination. For example, it seems less easy for the model to explain the effects of the notched noise presented simultaneously with the standard in the experiments of Plack and Viemeister (1992 a) and Zeng (1998). On the one hand, the noise should provide an efficient cue decreasing the perceptual similarity between masker and standard, as the listener hears a pure tone (masker) and a tone-in-noise (standard), respectively. This being the case, the reduction of midlevel jnd's in the presence of an additional notched noise could be accounted for. On the other hand, the similarity model predicts the same effect for the intense high-pass noise, which did not remove the midlevel jnd elevation in Zeng's study. Data on the effects of a noise presented simultaneously with the target on loudness enhancement have not been reported. In each case, it would be necessary to take into account the reduction of loudness and threshold shifts due to partial masking in this condition.

The proximal bursts that reduced jnd's in the experiments by Plack and colleagues can be viewed simply as additional maskers in terms of the similarity model. If a loudness matching experiment was conducted presenting a proximal burst, not only the representations of masker and target tone, but also of the proximal burst could be assumed to interact. Therefore, the loudness representation of the target effective in comparing target and comparison tone loudness should be a weighted average between the three initial loudness values. No data on loudness enhancement in this condition have been reported, but for a proximal burst similar in spectrum and level to the masker, no reduction in loudness enhancement can be expected according to the similarity model. It could only be argued that due to the short silent interval between proximal burst and target (12.5 – 50 ms), the burst causes adaptation, so that the arguments discussed above for the case of small ISIs would apply.

Another effect that seems to be incompatible with the model is the observation of a more pronounced midlevel jnd elevation at high standard frequencies. The effects of this parameter on loudness enhancement have not been studied, but according to the model, a variation of frequency should have no effect on loudness enhancement, as long as masker and target are presented at the same frequency.

To conclude, the similarity model correctly predicts the midlevel humps observed with an intense forward masker. In the following, it will become evident that unlike previous models, the similarity model also predicts a ***mid-difference hump***, i.e., maximum loudness enhancement and jnd elevation at intermediate masker-target level differences. The effects of similarity in duration and spectrum previous models could not account for directly follow from the similarity model. The complex effects of the proximal bursts used by Plack and colleagues and of noise presented simultaneously with the target, which also present some problems for previous models, are difficult to explain by the similarity model. It was pointed out above that in these conditions, both peripheral effects like adaptation and higher level effects like masker-target similarity are likely to play a role (Zeng, 1998). The difficulties might be related to the interaction between the different types of mechanisms.

6 Testing the similarity model: Effects of the masker-target level difference

In most experiments studying the midlevel-hump in intensity discrimination or loudness enhancement, the forward masker was presented at a fixed level (most frequently 90 dB SPL - 100 dB SPL). According to the similarity model, the small amount of enhancement at low target levels is the consequence of the large perceptual distance between the low-intensity tones and the high-intensity maskers. I.e., the assumed effect of perceptual similarity between masker and target results in a midlevel hump in loudness enhancement.

Now, if a 30-dB SPL target is combined with maskers presented at different levels, the similarity effect should result in a non-monotonic relation between enhancement and masker level. At a masker level of 30-dB SPL, no enhancement is predicted as the weighted average between masker and target loudness is identical to target loudness. With increasing masker level, enhancement should increase up to a level difference sufficiently large for the amount of mergence to decrease again. Insofar, it seems more appropriate to speak of a *mid-difference* rather than of a *mid-level* hump. According to the similarity model, the large perceptual distance between masker and target loudness present at low target levels causes the non-monotonic relation between standard intensity and the jnd if an intense masker is used.

From the preceding argument, it follows that a critical test of the similarity model is to vary the masker-target level difference ($\Delta L_{M\text{-}T} = L_M - L_T$) independently of target level L_T, for instance, by combining a 30-dB SPL target with different masker levels corresponding to level differences of −15, 0, +30, and +60 dB.

6.1 Model predictions

The recovery-rate model, the referential encoding hypothesis and the similarity model predict distinctly different patterns for the effects of the masker-target level difference.

6.1.1 Recovery-rate model

According to the model proposed by Zeng, Turner, and Relkin (1991), the intensity of low level standards (< 40 dB SPL) is coded by the high-SR fibers. As long as the threshold of these neurons is not elevated to values near or above standard intensity at standard onset (100 ms after presentation of the forward masker), masker level will have no effect. For even higher masker levels, the jnd is expected to increase monotonically with masker level (Fig. 17). As recovery of the high-SR fibers is fast (Relkin and Doucet, 1991), the threshold shift can be expected to be very small at a masker-standard interval of 100 ms. Compatible with this view, a 90 dB SPL masker caused no jnd elevation at standard levels of 20-30 dB SPL in most --but not all-- previous studies.

In the forward masking situation, standard intensities above the operating range of the high-SR fibers can be coded as precisely as in quiet only if they fall into the operating range of the low-SR fibers. At low masker levels, the threshold shift effective for the latter population of fibers is negligible 100 ms after masker offset, so that no ‘coding gap’ is present. At higher masker levels, however, the threshold shift becomes large enough for a coding gap to appear. This results in elevated jnd’s for standard levels falling into this gap. Further increases in masker level cause the threshold shift at standard onset to be more pronounced and therefore the coding gap to be larger. Consequently, the jnd elevation at intermediate standard levels is predicted to be a monotonically increasing function of the masker-standard level difference (Fig. 17). This prediction is compatible with results reported by Zeng and Turner (1992), who presented 40-60 dB SPL pedestals and varied masker. They found no jnd elevation for masker levels below pedestal level and an approximately linear jnd increase at higher masker levels.

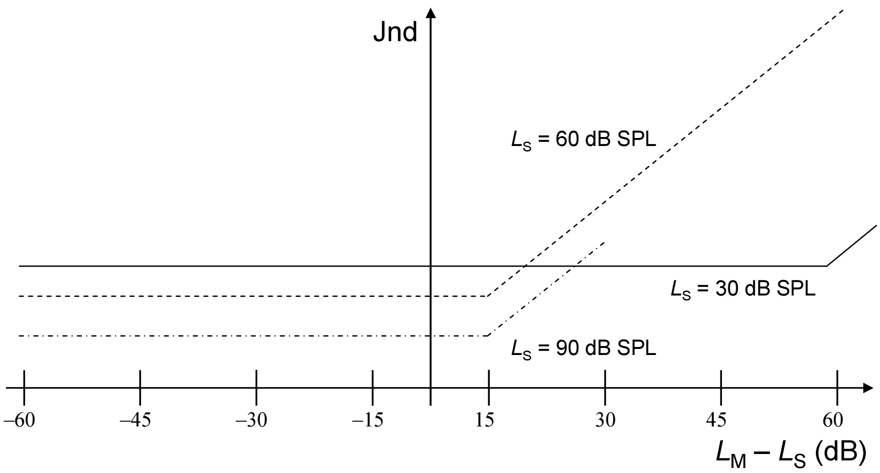

Fig. 17: Recovery-rate model: Schematic representation of the predicted effect of the masker-standard level difference $L_M - L_S$. The jnd is assumed to be elevated only if the masker causes the thresholds of the high-SR fibers (low standard levels; solid line) or of the low-SR fibers (intermediate and high standard levels; dashed line and dot-dashed line, respectively) to be elevated above standard level at presentation of the standard. Due to the faster recovery of the high-SR fibers, the masker-standard level difference at which the jnd starts to increase is considerably greater at low than at higher standard levels. Above the critical masker level, the jnd is predicted to increase monotonically with masker level.

As high intensities are also coded by the low-SR neurons, essentially the same arguments as for the intermediate standard levels apply. A jnd elevation is expected only if the masker is sufficiently intense to shift the threshold of the low-SR neurons to values near or above standard level. Above this point, jnd elevation should increase monotonically with masker level. No jnd elevation was observed for a 90-dB SPL standard combined with 90-100 dB SPL maskers. This finding is compatible with the model, as the threshold of the low-SR fibers should have been way below 90 dB SPL in this condition.

As noted above, the recovery-rate can predict jnd's because it can not account for loudness enhancement.

6.1.2 Referential encoding hypothesis

In terms of the referential encoding hypothesis (Plack and Viemeister, 1992*b*; Carlyon and Beveridge, 1993), the effects of masker-standard level difference are somewhat difficult to predict as the masker can be assumed to have two contrary effects.

First, trace degradation caused by the masker should be a function of masker level. For masker intensity below detection threshold, certainly no degradation of the trace of the standard should occur. In the other extreme, if the masker is so intense that the standard is below threshold, intensity discrimination is impossible. Therefore, trace degradation can be expected to be not all-or-none but a monotonically increasing function of masker level. The referential encoding hypothesis also makes no explicit assumptions about the relation between the masker-standard level difference $L_M - L_S$ and the amount of trace degradation. It seems reasonable, however, to assume that the amount of trace degradation depends on the level difference. To give an example, a 60 dB SPL masker should have a stronger detrimental effect on the trace of a 30 dB SPL standard tone than on the trace of a 90 dB SPL tone. Second, the masker could serve as a within-interval coding reference. In intensity discrimination experiments, the ISI between masker and tone presented in each observation was generally smaller than the ISI between the target tones (e.g., 100 ms vs. 650 ms). Above that, masker intensity was constant within each block, in most studies even throughout the complete experiment. Consequently, a listener should be able to establish an effective long-term representation of masker intensity and to use this as a reference in context-coding mode. As before, however, the perceptual distance between masker and standard is crucial according to the perceptual anchor model (Braida et al., 1984) that predicts perceptually similar maskers to be more effective as a coding reference. For this reason, the utility of the masker as a reference should be maximal if masker level equals standard level.

To summarize the two considerations, according to the referential encoding hypothesis it can be expected that context-coding is not necessary at masker levels well below standard level, as the trace of the first target tone is essentially unimpaired in this case. Consequently, jnd's in this condition should not differ from jnd's in quiet (Fig. 18). This prediction applies to all standard levels.

As masker level increases, trace degradation is assumed to become more and more pronounced. Consequently, trace mode gets less effective, and the system has to rely increasingly on context-coding. The effect on intensity resolution depends on standard level in this situation: If masker level is increased to values above standard level, trace degradation increases, and at the same time the effectiveness of the masker as a coding reference becomes smaller. At intermediate pedestal levels, no internal references are effective so that the jnd elevation is predicted to increase monotonically with the level difference between masker and standard. In contrast, at low and high pedestal levels, threshold and discomfort level are still effective as internal references. Thus, no or only a small jnd increase is expected in these conditions. In fact, Berliner, Durlach, and Braida (1978) demonstrated that presentation of an explicit standard in each trial of an identification experiment resulted in improved performance at midlevels only. These results hint to the effectiveness of the internal references. In intensity discrimination experiments, a 90 - 100 dB SPL masker caused no or only a small jnd elevation at standard levels of 20-30 dB SPL and 80-90 dB SPL. A large jnd elevation resulted at pedestal levels between 40 and 70 dB SPL, however and data by Zeng and Turner (1992) indicate that at intermediate levels, the jnd increases monotonically with the masker-standard level difference. A very intense masker (e.g., 120 dB SPL) can also be expected to cause adaptation in the auditory nerve that is still substantial at standard onset, so that intensity discrimination performance will be impaired by the resulting threshold shift, at least at low standard levels.

It is difficult to predict the effect of maskers slightly below standard level. If masker level is decreased below standard level, trace degradation decreases, but the effectiveness of the masker as a coding reference also decreases. Depending on the specific effects of masker level on trace degradation on the one hand and 'perceptual distance' on the other hand, the jnd could be either a monotonically decreasing or a non-monotonic function of masker level in this case. By and large, however, jnd's must approach the jnd's in quiet as masker level is reduced to values below detection threshold.

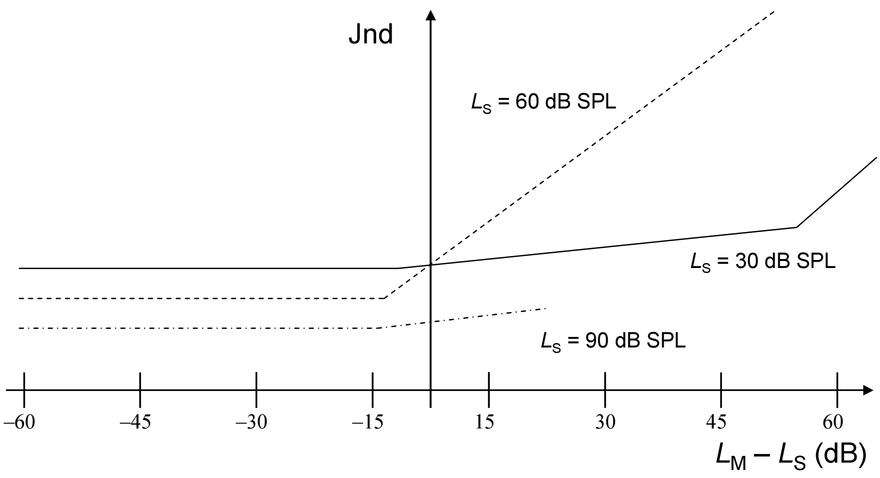

Fig. 18: Referential encoding hypothesis: Schematic representation of the predicted effect of the masker-standard level difference $L_M - L_S$. The jnd is assumed to be approximately identical to the jnd in quiet for masker levels below standard level, but to increase monotonically as L_M is increased to and above standard level. The jnd elevation is predicted to be very small or completely absent at low (solid line) and high (dot-dashed line) standard levels due to the availability of internal coding references, but to be pronounced at midlevels (dashed line).

The model can be used to predict jnd's only because it can not account for loudness enhancement.

6.1.3 Similarity model

The similarity model predicts loudness enhancement and intensity discrimination difference limens to be non-monotonic functions of the masker-target level difference. For example, a 60-dB SPL target combined with a 60-dB SPL masker causes no loudness enhancement according to the model as the weighted average between target and masker loudness is equal to target loudness. As masker level increases above 60 dB SPL, loudness enhancement is predicted to increase. At the same time, the weight assigned to masker loudness in computation of the weighted average between masker and target loudness decreases because masker and target become increasingly dissimilar. For this reason, loudness enhancement is expected to decrease again at large masker-target level differences (Fig. 19). As the difference limen is monotonically related to the amount of loudness enhancement according to the loudness enhancement hypothesis (Carlyon and Beveridge, 1993), the same pattern is predicted for intensity resolution.

Essentially the same arguments apply for masker levels **smaller** than target level. Loudness decrement is predicted to first increase as masker level is decreased below target level, and than to become smaller again as the perceptual similarity decreases. In line with the results of Elmasian, Galambos, and Bernheim (1980), these effects are expected to be smaller than for loudness enhancement. The reason for this behavior is not immediately apparent. In terms of the mergence hypothesis, it would be necessary to assume that maskers lower in level than the targets generally receive less weight than maskers higher in level. Exactly this pattern is indeed predicted by the quantitative version of the similarity model formulated in Chapter 7, where the smaller amount of loudness decrement relative to enhancement arises because of properties of the loudness function.

In each case, as masker level is reduced to values below threshold, loudness decrement must vanish completely. As discussed in Chapter 4.1.3, loudness decrement can also be expected to introduce an increased variance in the intensity representation of the standard, just as loudness enhancement. In other words, the similarity model predicts the difference limen to be a monotonically increasing function of the *absolute value* of the loudness change.

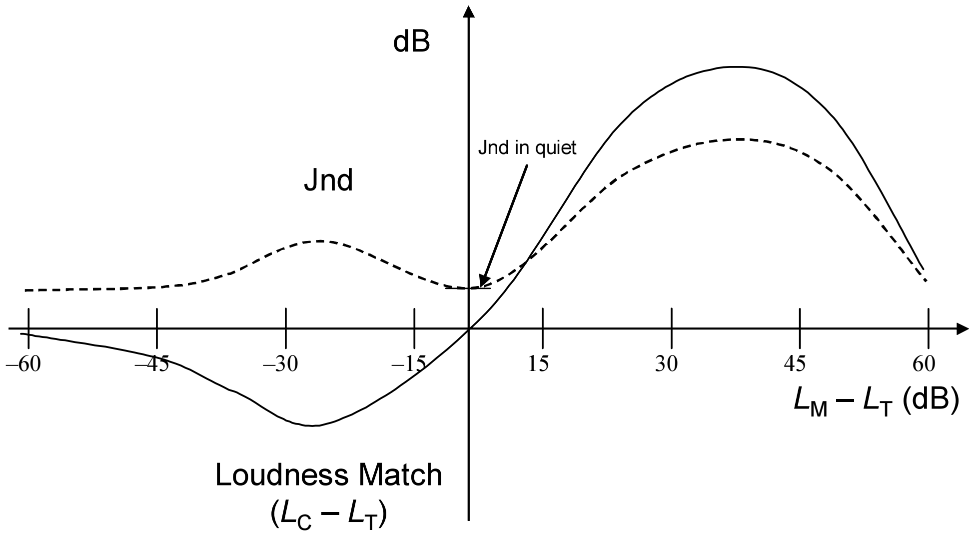

Fig. 19: Similarity model: Schematic representation of the predicted effect of the masker-target level difference $L_M - L_T$. The loudness change (solid line) caused by the forward masker is a non-monotonic function of $L_M - L_T$. The jnd (dashed line) is assumed to be monotonically related to the absolute value of the loudness change. If masker level is neither below threshold nor sufficiently high to mask detection of the target, the same pattern is predicted at each target level (see text). Consequently, for the sake of simplicity, only one pair of functions is displayed.

Except for the case of extreme masker levels (near or below detection threshold or sufficiently high to mask detection of the target), the model predicts the function relating loudness enhancement/jnd and masker-target level difference to be the same **at each target level**. This is an additional important difference to the predictions of the recovery-rate model and the referential-encoding hypothesis. It is also the reason why –for the sake of simplicity- only one pair of functions is displayed in Fig. 19.

6.2 Experiment 1: Intensity Discrimination

In Experiment 1, intensity jnd's were measured as a function of standard level and the level difference between masker and standard.

The standard was presented at low, intermediate, and high levels (L_S = 25, 55, and 85 dB SPL, respectively). The level difference between masker and standard ($\Delta L_{M\text{-}S} = L_M - L_S$) was varied between −60 and +60 dB in 15-dB steps. The smallest masker level was 10 dB SPL. For the listeners in this experiment, this level was between 1 and 5.5 dB above detection threshold (Table 1). Out of concern for the listeners, the highest masker level was 100 dB SPL and the maximum sound pressure level was restricted to 105 dB SPL.

To save experimentation time, the 10-dB SPL masker was not presented with the 55-dB SPL standard, and the 10-dB SPL, 25-dB SPL and 40-dB SPL maskers were not used with the 85-dB SPL standard, as the effect of those maskers on the difference limen was expected to be negligible. Due to a technical problem, listener AL did not receive the 10-dB SPL masker combined with the 25-dB SPL standard and the 25-dB SPL masker combined with the 55-dB SPL standard; listener BS did not receive the 70-dB SPL masker/85-dB SPL standard combination.

Jnd's in quiet were obtained for all listeners at the three standard levels.

6.2.1 Method

6.2.1.1 Stimuli

Standard and masker were 1-kHz pure tones, with a steady- state duration of 20 ms. They were gated on and off with 5-ms cosine-squared ramps (all durations and silent intervals are measured between zero-voltage points). This configuration is essentially the same as in the experiments by Plack, Carlyon, and Viemeister (1995), who also used a pure-tone masker and standard of same the duration.

In each trial, there were two observation intervals (Fig. 20). In each interval, a masker and a standard were presented. The silent interval between masker offset and standard onset was 100 ms. The interval between offset of the first standard and onset of the second standard was 650 ms. In one of the observation intervals (selected with an equal a-priori probability), an increment of same frequency, duration and temporal envelope was added in-phase to the standard. A trial started with a visual attention signal that was on for 300 ms, followed by a silent interval of 500 ms and then the first tone.

6.2.1.2 Apparatus

A MATLAB program was used for stimulus generation and control of the experiment. Stimuli were 24-bit resolution, 44.1-kHz sampling-frequency wave files. They were played back via an *M*-Audio Delta 44 PCI audio-card. One channel was used for the masker, a separate channel for standard and standard-plus-increment. The increment was produced digitally. Masker and standard/standard-plus-increment were fed into separate channels of a custom-made programmable attenuator, summed in an inverting summing amplifier, amplified in a headphone amplifier, and fed into the right channel of Sennheiser HDA 200 headphones.

For calibration, a Brüel and Kjær Artificial Ear 4153 (IEC 318 specification), a Brüel and Kjær Microphone 4134**,** a Rhode and Schwartz calibrator, and a Norwegian Electronics Sound Meter 108 were used.

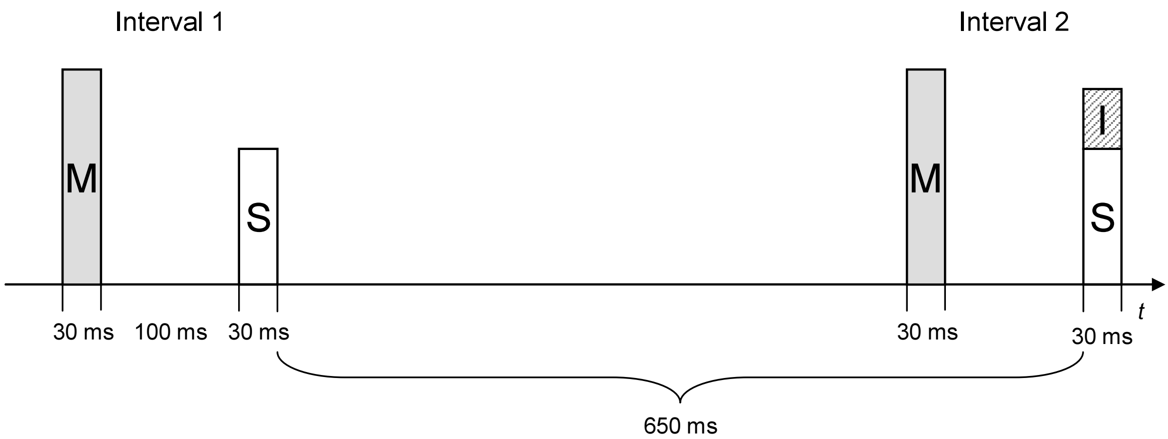

Fig. 20: Trial configuration used in Experiment 1. 2I, 2AFC, adaptive procedure. In each of the two observation intervals, a forward masker M and the standard S were presented. The level increment I was presented in interval 1 or interval 2 with equal a-priori probability. Listeners responded whether the increment had occurred in the first or in the second interval. After two consecutive correct responses, the increment was decreased. After each incorrect response, the increment was increased (2-down, 1-up).

The experiment was conducted in a sound-proof chamber ("Camera Silenta" of the Institute of Technical Acoustics at the University of Technology Berlin). Listeners were tested individually.

6.2.1.3 Listeners

Six listeners participated in the experiment. One of them (DO) was the author; the others were paid an hourly wage. For the ear used, all had hearing levels (as measured by pure-tone audiometry according to ISO 8253) better than 10 dB in the frequency range between 125 Hz and 8000 Hz, with exception of DO (HL = 10.9 dB at 1500 Hz), and AL (HL = 17.6 dB at 8000 Hz). All listeners except BS received stimulation to their right ear. For BS, the left ear was used because in the right ear, HL was 19.6 dB at 8000 Hz, while HLs in the left ear were better than 6.6 dB in the complete frequency range measured.

Listeners were fully informed about the course of the experiment. All except the author were naïve with respect to the aim of the experiment. None except DO had previous experience in psychoacoustic experiments. Listeners received at least 2 hours of practice. If necessary, it was allowed for further practice until performance stabilized.

6.2.1.4 Procedure

A 2I, 2AFC adaptive procedure with a 2-down, 1-up tracking rule (Levitt, 1971) was used to measure jnd's ($\Delta L_{DL} = 10 \cdot \log[1 + \Delta I/I]$) corresponding to 70.7% correct. In each interval, a masker and a standard were presented. In one of the observation intervals (selected with equal a-priori probability), an increment was added to the standard.

Listeners selected the interval containing the louder target tone by pressing one of two response buttons. They were instructed to ignore the maskers. To help them ignore the maskers, standard and standard-plus-increment were marked by visual signals. Two light-emitting diodes (LEDs) were used. LED 1 was switched on 50 ms before the onset of the standard in interval 1 and switched off 50 ms after its offset. LED 2 marked the standard presented in the second interval in exactly the same way.

After two consecutive correct responses, the increment was reduced. After each incorrect response, the increment was increased. Until the fourth reversal, step size was 5 dB. For the remaining eight reversals, step size was 2 dB. At the end of each run, the jnd expressed as ΔL_{DL} was taken as the mean of 10 log(1 + $\Delta I/I$) at the last eight reversals. A run was discarded if the standard deviation was greater than 5 dB within the run. At least three runs were obtained for each data point. Time permitting, additional tracks were run if the standard deviation of jnd's obtained in the first three runs exceeded 5 dB; so each data point is based on three-to-nine runs.

Visual false/correct feedback was provided after each trial. The experiment was self-paced. In each block, only one masker-standard level combination was presented. Listeners received the conditions in pseudo-random order with the exception that blocks with the 100-dB SPL masker were always presented at the end of a session to avoid that a potential temporary threshold shift due to the intense maskers have an effect on detection thresholds in subsequent blocks.

An experimental session lasted approximately one hour. Listeners took one or two short breaks in each session.

Five of the six listeners also participated in Experiment 2, where loudness matches were obtained for the same masker-target level combinations. Therefore, they received both intensity discrimination and loudness matching blocks within each session.

6.2.2 Absolute thresholds

Before the main experiment, absolute thresholds were measured for the 30-ms 1-kHz tones in quiet and under forward masking.

The same temporal configuration as for the intensity discrimination experiment was used. Thresholds were obtained in a 2I, 2AFC, adaptive procedure (2-down, 1-up rule). In one interval (selected randomly), the signal was presented, while no tone was presented in the other interval. The two observation intervals were again marked by two LEDs. In the forward masking conditions, a masker was presented in both intervals. The silent interval between masker offset and signal onset was 100 ms. Masker levels were 25, 55, 85, and 100 dB SPL, except for listener BS. Listeners DO and SD also received 100-ms duration maskers.

Listeners selected the interval containing the signal by pressing one of two response buttons. Initially, signal level was 20 dB SPL. Step size was 5 dB until the fourth reversal, and 2 dB for the remaining eight reversals. Visual false/correct feedback was provided after each trial.

Threshold intensity (corresponding to 70.7% correct detections) was computed as the arithmetic mean of the signal sound-pressure-levels at the last eight reversals. For each condition, at least three such measurements were made. If within a run, the standard deviation was larger than 5 dB, the run was repeated. If the standard deviation of thresholds obtained in the three runs exceeded 5 dB, additional runs were obtained.

Individual results are displayed in Fig. 21 and in Table 1. Thresholds in quiet lay between 5 and 10 dB SPL. Thresholds in quiet and under forward masking were virtually identical for listener BS and elevated by only about 2.5 dB for DO (except for the case of a 100-dB SPL 100-ms masker). This finding is compatible with the results of Zeng, Turner, and Relkin (1991). For the remaining listeners, the masker caused an elevation of detection thresholds that tended to increase with masker level. The maximum elevation of 9.53 dB is comparable to the effect on detection threshold Carlyon and Beveridge (1993) reported for a 106-ms, 85.5-dB SPL, narrowband-noise masker.

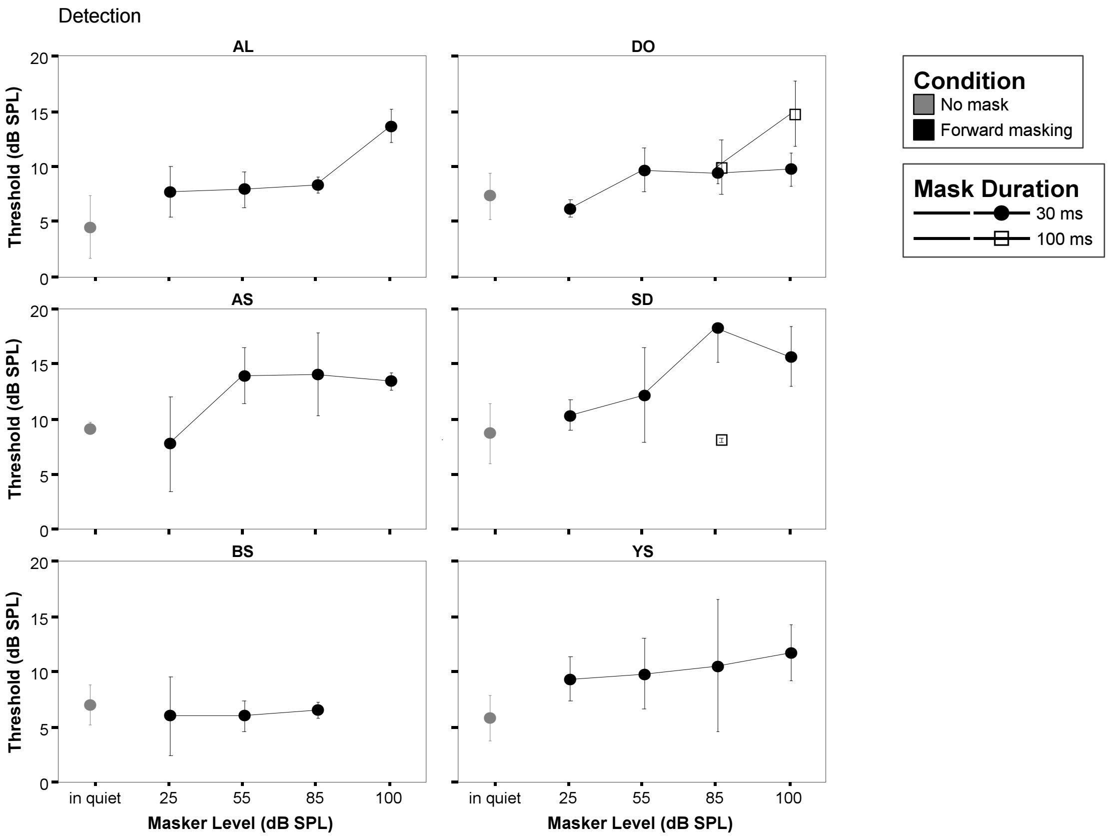

Fig. 21: Absolute thresholds in quiet (gray symbols) and under forward masking (black symbols). Each panel represents one listener. Filled circles: masker duration 30 ms. Open boxes: masker duration 100 ms. Error bars show ± 1 SD.

The fact that at the 85-dB SPL masker level, SD's threshold was much lower for the longer duration masker seems to indicate that a 'confusion effect' (cf. Neff, 1985) rather than neural adaptation caused by the masker was the reason for the elevated thresholds. In contrast, for listener DO, the 100-dB SPL, 30-ms masker produced only 2.45 dB of threshold elevation, while the masker 100-ms, 100-dB SPL masker resulted in a 7.45-dB increase in threshold. This finding can be explained by peripheral masking, as longer mask durations produce stronger adaptation in the auditory nerve (Harris and Dallos, 1979).

		Masker Level									
		In Quiet		25 dB SPL		55 dB SPL		85 dB SPL		100 dB SPL	
Listener	Mask Duration	L_{Th} (dB SPL)	SD (dB)	L_{Th} (dB SPL)	SD (dB)	L_{Th} (dB SPL)	SD (dB)	L_{Th} (dB SPL)	SD (dB)	L_{Th} (dB SPL)	SD (dB)
AL	30 ms	4.50	2.88	7.69	2.26	7.92	1.63	8.33	0.72	13.67	1.51
AS	30 ms	9.17	0.55	7.75	4.27	13.92	2.56	14.06	3.75	13.42	0.80
BS	30 ms	7.00	1.80	6.00	3.54	6.00	1.41	6.50	0.71		
DO	30 ms	7.30	2.08	6.17	0.76	9.69	2.01	9.44	0.97	9.75	1.52
	100 ms							9.92	2.50	14.75	2.95
SD	30 ms	8.72	2.73	10.33	1.38	12.19	4.34	18.25	3.08	15.67	2.74
	100 ms							8.13	0.18		
YS	30 ms	5.83	2.05	9.33	2.01	9.81	3.20	10.56	5.97	11.75	2.54

Table 1: Detection thresholds for 30-ms 1-kHz tone bursts corresponding to 70.7% correct. Arithmetic means and standard deviations for three to five measurements per condition.

6.2.3 Results

Individual results from the intensity discrimination experiment are displayed in Fig. 22, where each row represents one listener while each column represents one standard level.

In quiet (diamonds), the data can be described by the near-miss to Weber's law as ΔL_{DL} decreased with standard level.

At the 25-dB SPL standard level, ΔL_{DL} was a non-monotonic function of the masker-standard level difference for listeners AL, AS, SD, and YS. The remaining two listeners deviated from this pattern. For listener BS, difference limens increased dramatically with the masker-standard level difference. His data are in some aspects similar to the pattern Schlauch et al. (1997) reported for one subject of their study, where DLs were even larger (up to 50 dB) for low-level standards combined with a 90-dB SPL masker. Schlauch et al. hypothesized that the listener integrated the intensities of masker and standard, presumably because of the perceptual similarity between the tones. For listener DO (the author), the jnd also showed a rather small monotonic increase with masker-target level difference. For listeners AL and SD, the difference limen at $L_M - L_S = 60$ dB was slightly larger than at 45 dB.

For the 55-dB SPL standard, jnd's at the largest masker-target level difference (45-dB) were smaller than at the 30-dB level difference for two listeners (SD and YS), resulting in a mid-difference hump. Jnd's at $L_M - L_S = 45$ dB were still larger than in quiet. The remaining listeners showed no mid-difference hump at this standard level. Zeng and Turner (1992) also reported a monotonic increase of the jnd's for a midlevel standard as the masker-standard level difference was increased from 0 to 50 dB. The maximum jnd elevation was only 8 dB in their experiment. This difference to the present findings could be attributed to the use of a narrowband-noise, 100-ms masker by Zeng and Turner. According to the similarity model, the reduced perceptual similarity due to duration and waveform may have resulted in a smaller effect of the masker. This issue has previously been discussed by Plack, Carlyon, and Viemeister (1995), who also observed DLs of up to 20 dB for an ipsilateral forward masker (80 dB SPL) equal in duration to the 50-dB SPL standard.

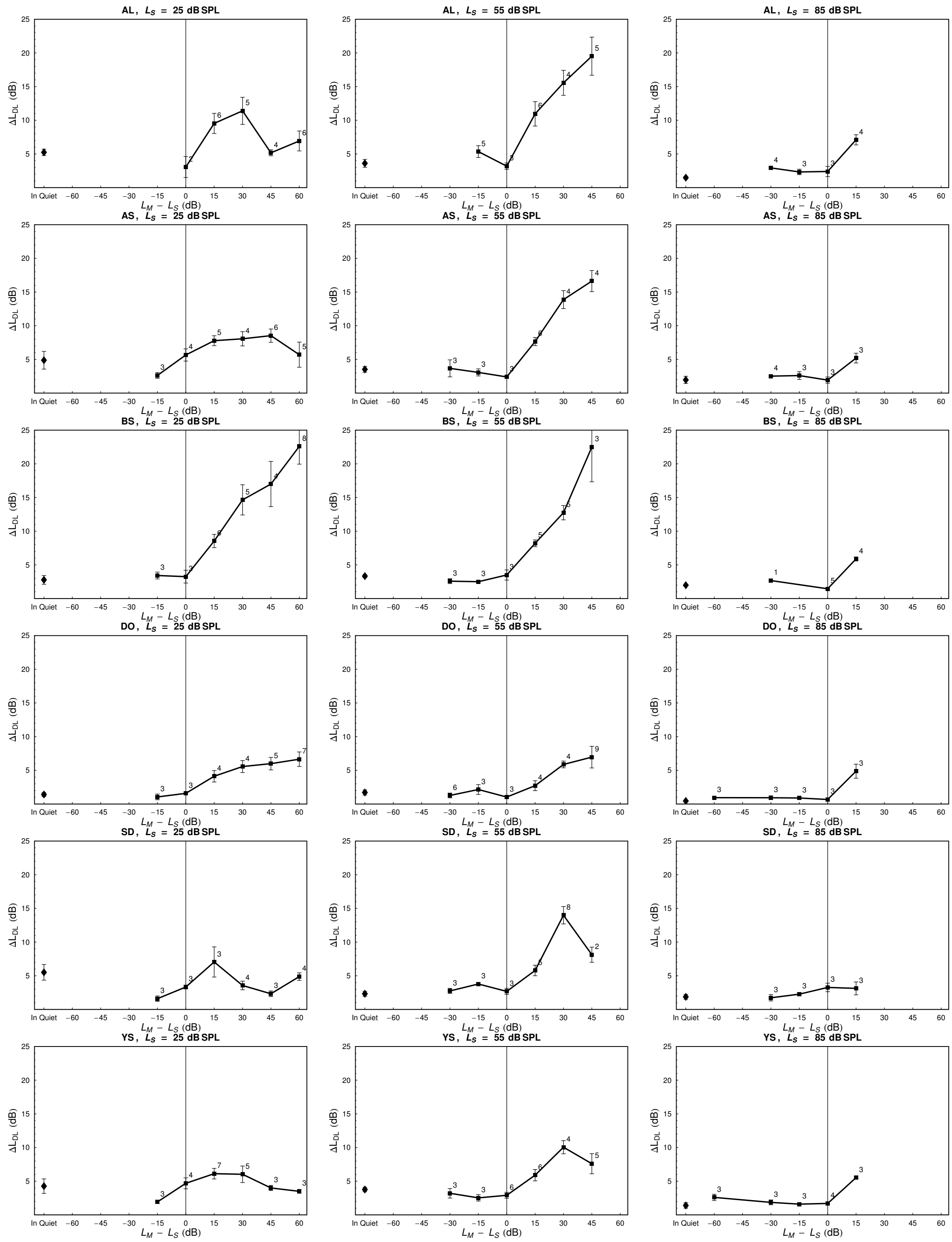

Fig. 22: Jnd's ($\Delta L_{DL} = 10 \log_{10}[1 + \Delta I_{DI}/I]$) as a function of the masker-standard level difference $L_M - L_S$ and standard level L_S. Rows represent listeners, columns standard levels. Diamonds: in quiet. Squares: in forward masking. Numbers denote the number of runs a data point is based on. Error bars show ± 1 standard errors of the mean (SEM).

Except for listener SD, difference limens for the 85-dB SPL standard increased by several dB as masker level was increased from 85-dB SPL to 100-dB SPL (0 dB and 15 dB masker-standard level differences, respectively).

For masker levels smaller than or equal to standard level, jnd's were close to those in quiet. For three listeners, the 10-dB SPL masker even caused a reduction in the difference limen for the 25-dB SPL standard. A potential explanation for this finding is a cueing effect (Moore and Glasberg, 1982) in the sense that the masker provided a reliable cue to signal onset. In several other cases, jnd's at negative values of $\Delta L_{M\text{-}S}$ were larger than in quiet and/or for equal masker and standard level.

In the mean data, jnd's for the 25-dB SPL standard plateaued at a level difference of 30 dB, while there was a monotonic increase at the 55-dB SPL standard level (Fig. 23).

As the data for listener BS deviated strongly from the data for the other listeners, his data were excluded from the following analyses.

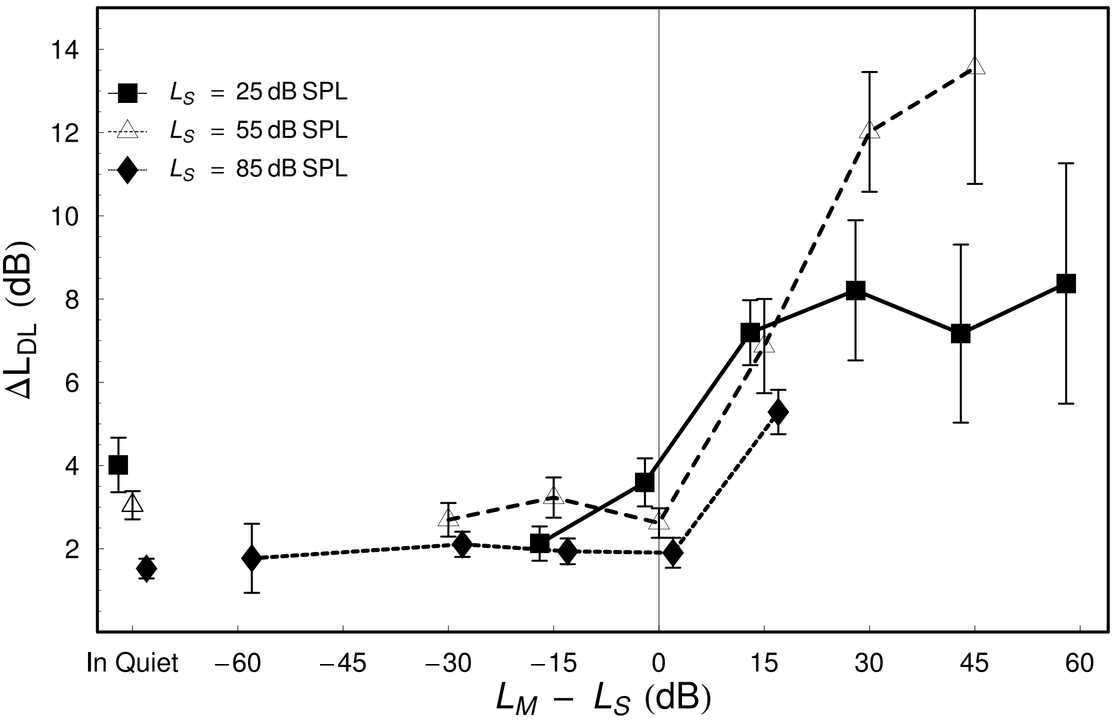

Fig. 23: Mean difference limens (ΔL_{DL}) as a function of $L_M - L_S$ for all six listeners. Boxes: 25-dB SPL standard. Triangles: 55-dB SPL standard. Diamonds: 85-dB SPL standard. To avoid cluttering, lines are shifted by 2 dB on the x-axis. Error bars show ± 1 SEM.

Repeated-measures analyses of variance (ANOVAs) were conducted using a maximum-likelihood approach (SAS PROC MIXED; Littell et al., 1996) because three data points in the set are missing due to a technical problem (see above). The restricted maximum-likelihood method was used to estimate the covariance parameters and the Satterthwaite method for computing denominator degrees of freedom for approximate F-tests of fixed effects. Note that the latter method results in fractional degrees of freedom in the F-values reported below. The "heterogeneous compound symmetry" (CSH) was selected to model the covariance structure. This structure is similar to the covariance structure assumed in repeated-measures ANOVAs based on a univariate approach with Huynh-Feldt df-correction (cf. Maxwell and Delaney, 2004) in that it has the same number of parameters and heterogeneity along the main diagonal of the variance-covariance matrix. However, it constructs the off-diagonal elements by taking geometric rather than arithmetic means.

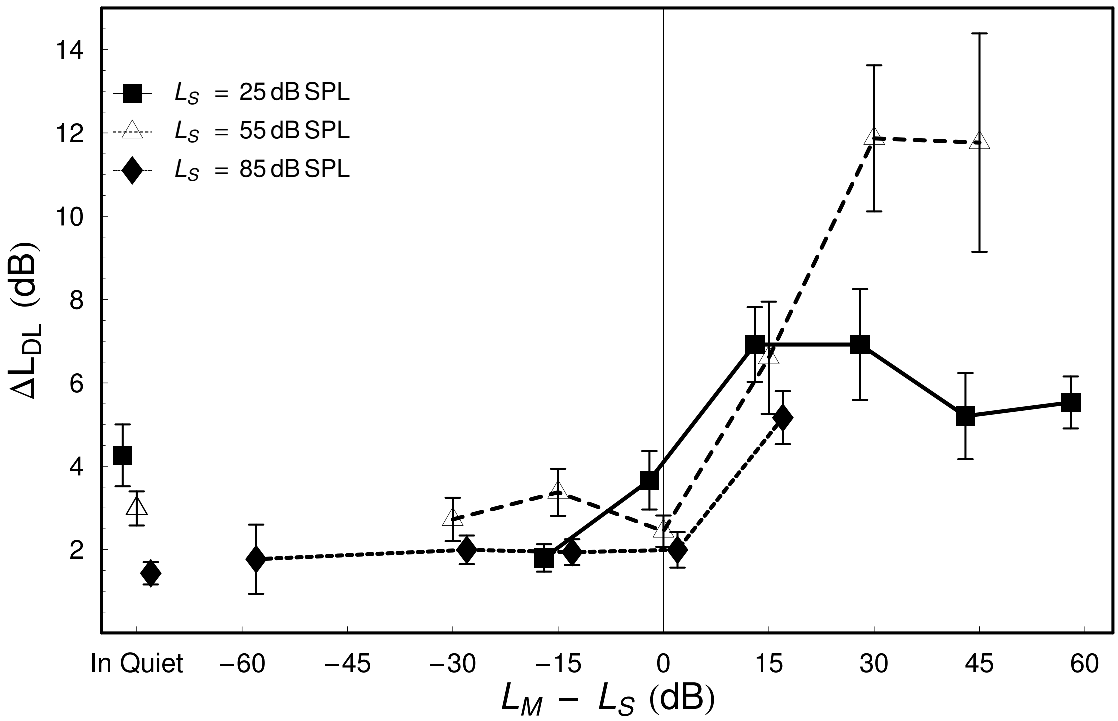

Fig. 24: Mean difference limens (ΔL_{DL}) as a function of $L_M - L_S$ for five listeners, data from listener BS excluded. Same format as in Fig. 23.

If the data from listener BS are excluded, mean difference limens at the 25-dB SPL standard level are maximum for intermediate masker-standard level differences ($\Delta L_{M\text{-}S}$ = 15 and 30 dB, Fig. 24), corresponding to the predicted mid-difference hump. A one-factorial ANOVA conducted for the data at the 25-dB SPL standard level confirmed this observation. Only forward-masked data were included in the analysis. There was a significant effect of $\Delta L_{M\text{-}S}$, $F(5, 8.78) = 11.51$, $p < 0.002$. Trend analyses were conducted using the method of orthogonal polynomials. The significant quadratic trend, $F(1, 9.82) = 16.38$, $p < 0.003$, is evidence for the difference limen being a non-monotonic function of $\Delta L_{M\text{-}S}$ at this standard level. The linear trend was also significant, $F(1, 9.71) = 21.69$, $p < 0.002$, while the cubic trend did not reach significance, $F(1, 13.4) = 0.65$.

For the 55-dB SPL standard, the jnd showed no further increase as $L_M - L_S$ was increased from 30 to 45 dB. The difference limen was slightly larger at a masker-standard level difference of −15 dB than for the case of equal masker and standard level. Surprisingly, jnd's obtained with low masker levels were even slightly smaller than in quiet. A one-factorial ANOVA conducted for the forward-masked data at this standard level showed a significant effect of $\Delta L_{M\text{-}S}$, $F(5, 6.34) = 9.05$, $p < 0.008$. The cubic trend was significant, $F(1, 14.1) = 13.38$, $p < 0.003$. This result is compatible with the prediction of the similarity model (Fig. 19), according to which jnd's should be larger at both positive and negative intermediate values of $\Delta L_{M\text{-}S}$ (e.g., −15 dB and +15 dB) than at a 0-dB level difference. Note, however, that the quadratic trend was only marginally significant, $F(1, 4.2) = 5.75$, $p < 0.08$. There was also a significant linear trend, $F(1, 4.07) = 22.77$, $p < 0.009$. Note that the jnd showed no further increase as $\Delta L_{M\text{-}S}$ was increased from 30 dB to 45 dB. The deviation from a linear trend for values of $\Delta L_{M\text{-}S}$ between 0 and 45 dB was confirmed by an ANOVA for these data, which indicated a significant linear, $F(1, 3.99) = 18.68$, $p < 0.013$, but also a significant quadratic trend, $F(1, 8.99) = 6.49$, $p < 0.032$.

For the 85-dB SPL standard, only two listeners received the −60-dB masker-standard level difference. Data from this condition were excluded from the analysis. In the ANOVA conducted for the remaining data points, there was a significant effect of $\Delta L_{M\text{-}S}$, $F(3, 4.5) = 9.14$, $p < 0.023$.

To compare the jnd elevation caused by an increase of $\Delta L_{M\text{-}S}$ from 0 dB to 15 dB at the three standard levels, a two-factorial L_S (25, 55, 85 dB SPL) × $\Delta L_{M\text{-}S}$ (0 dB, 15 dB) ANOVA was conducted. The main effects of standard level ($F[2, 14.3] = 5.62$, $p < 0.016$) and masker-standard level difference ($F[1, 8.48] = 43.84$, $p < 0.001$) were significant, while the $L_T \times \Delta L_{M\text{-}S}$ interaction was not, $F(2, 12.6) = 0.37$. In other words, the jnd increase caused by the masker in the +15 dB condition was not stronger at the 55-dB SPL standard level than at the low and high standard level.

An L_S (25, 55 dB SPL) × $\Delta L_{M\text{-}S}$ (6) ANOVA was used to analyze the data collected with masker-standard level differences between −15 and 60 dB at the 25-dB SPL and 55-dB SPL standard levels. Both main effects were significant (L_S: $F(1, 10.5) = 16.23$, $p < 0.003$; $\Delta L_{M\text{-}S}$: $F(4, 6.91) = 13.05$, $p < 0.003$). The $L_S \times \Delta L_{M\text{-}S}$ interaction was also significant, $F(4, 9.55) = 6.29$, $p < 0.01$. This analysis thus confirms the observation that the effects of the forward masker were more pronounced at the intermediate than at the low standard level.

Finally, to examine whether the difference limens obtained in quiet and in forward-masking with masker levels smaller than or equal to standard level were statistically different, the mean value of ΔL_{DL} for $\Delta L_{M\text{-}S} \leq 0$ dB was computed at each of the three standard levels. These values were compared to ΔL_{DL} obtained in quiet by means of an L_S (25, 55, 85 dB SPL) × Condition (In Quiet, Masked) repeated-measures ANOVA. A multivariate approach was used (SPSS GLM). As it can be expected from Fig. 24, there was a significant effect of standard level, $F(2, 3) = 17.48$, $p < 0.023$, that reflects the near-miss to Weber's law. The effect of Condition was only marginally significant, $F(1, 4) = 7.045$, $p < 0.058$. On the one hand, this is compatible with the prediction of the similarity model that the jnd increase should be absent or small at a 0-dB level difference as well as at masker levels significantly below standard levels. The similarity model also predicts a jnd elevation at intermediate negative values of $\Delta L_{M\text{-}S}$, however, which is not present in the data. The L_S × Condition interaction was significant, $F(2, 3) = 19.73$, $p < 0.021$. This effect is most likely due to the mean jnd observed with a low level masker ($\Delta L_{M\text{-}S}$ −15 dB or 0 dB) being smaller than in quiet at a standard level of 25 dB SPL.

For comparison with previous studies where a fixed-level intense masker was used and only the level of the standard was varied, the mean data obtained with the 85-dB SPL masker are plotted in Fig. 25. The size of the "midlevel hump", i.e., the jnd elevation at the intermediate standard level, was approximately 9 dB, which is comparable to the value of approximately 13 dB reported by Plack, Carlyon and Viemeister (1995), who also used equal-duration pure-tone maskers and standards (80-dB SPL and 50-dB SPL, respectively).

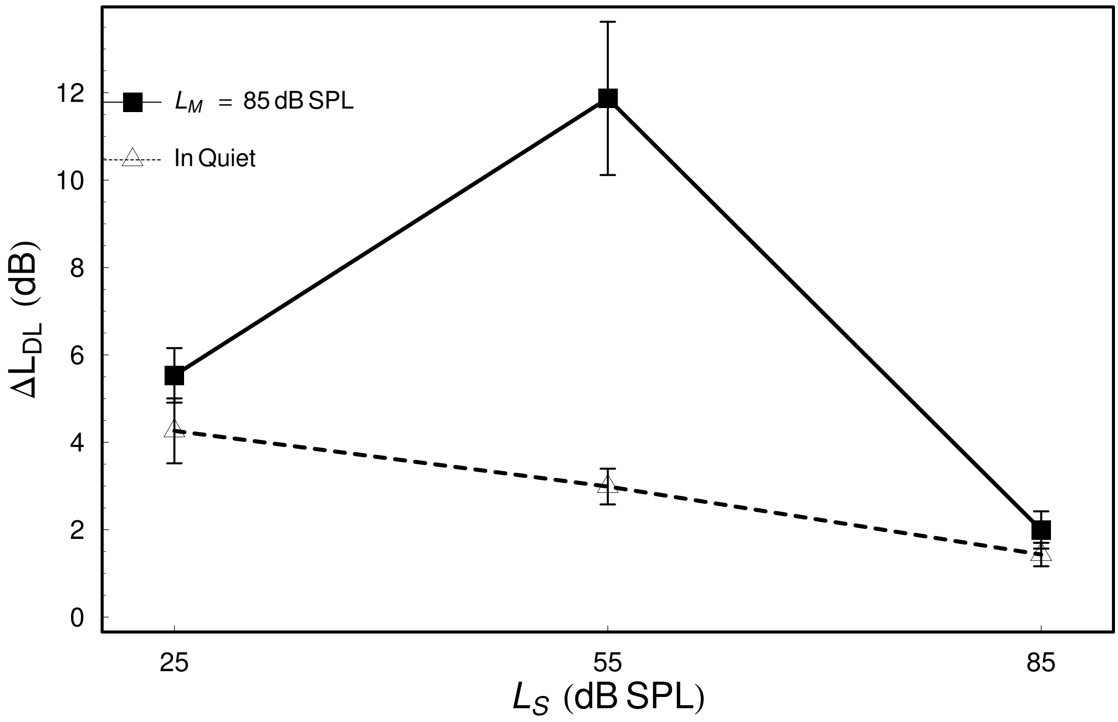

Fig. 25: The 'midlevel hump' in intensity discrimination: mean difference limens (ΔL_{DL}) for 5 listeners (data from BS excluded) plotted as a function of standard level L_S. Boxes: 85-dB SPL forward masker. Triangles: in quiet. Error bars show ± 1 SEM.

6.2.4 Discussion

At the 25-dB SPL standard level, both the individual results and the statistical analyses indicated an effect of the forward masker for all masker levels above standard level. This is evidence against the recovery-rate model, which predicts the forward masker to have no effect at low standard levels except for very intense maskers (Fig. 19). The non-monotonic functions present for four of the six listeners and the significant quadratic trend (with the data from listener BS excluded) are evidence against the referential encoding hypothesis. In its original form, this model also predicts the masker to have no effect at low standard levels, as low intensities can be efficiently coded relative to detection threshold. Even if the additional assumption is made that trace degradation increases with masker level, this is not compatible with the observation that the jnd first increased with the masker-standard level difference but then decreased again at large values of $\Delta L_{M\text{-}S}$.

Except for the reduced jnd at the −15 dB-SPL masker-standard level difference, results for the 25-dB SPL standard are evidence for the similarity model. Listener BS clearly deviated from the pattern predicted by this model. The hypothesis that he integrated masker and standard intensity in a similar manner as one listener in the experiment by Schlauch, Lanthier, and Neve (1997) will be tested below in Experiment 3 and an auxiliary experiment. Listener DO also produced no mid-difference hump; the jnd elevation at the largest masker-standard level difference (60 dB) was much smaller than for BS, however.

A mid-difference hump was observed at the 55 dB SPL standard level for two listeners only. It remains unclear whether for the remaining listeners, the maximum masker-standard level difference of 45 dB was not large enough for the jnd to decrease again. In the mean data, the quadratic trend at the 0 dB to 45 dB masker-standard level differences indicates a departure from linearity, which is compatible with the predictions of the similarity model but evidence against the recovery-rate model and the referential encoding hypothesis. The fact that jnd's observed with large masker-standard level differences were much larger at a standard level of 55 dB SPL than at 25 dB SPL is difficult to explain in terms of the similarity model, according to which the jnd elevation should be a function of $\Delta L_{M\text{-}S}$ only but independent of standard level. In Chapter 7, a quantitative version of the similarity model will be introduced, which predicts the effect of the masker-target level difference on loudness enhancement to be less pronounced at low target levels due to the shape of the loudness function, so that the pattern observed in the present experiment is predicted.

For the 85-dB SPL standard, the 100-dB SPL masker ($\Delta L_{M\text{-}S}$ = 15 dB) caused a significant jnd elevation. This effect is not compatible with the recovery-rate model, as a 100-dB SPL masker can not be expected to shift the threshold of the low-SR fibers to values near 85 dB SPL. Again, the referential encoding hypothesis can explain the data only if the additional assumption is made that trace degradation increases with masker level. Even in this case, the jnd increase with the step from 0 dB to 15 dB level difference would be expected to be smaller for the 85-dB SPL standard than for the 55-dB SPL standard, as the perceptual distance to the internal reference discomfort level is smaller at high intensities. The fact that there was no significant effect of standard level on the jnd increase supports the similarity model rather than the two previous models.

The fact that in the mean data, there was either no or only a small jnd elevation at masker levels smaller than target level is at odds with the similarity model that predicts a jnd elevation at intermediate negative values of $\Delta L_{M\text{-}S}$. The pattern of results in some respects resembles data by Stellmack and Viemeister (2000). They measured difference limens for 10-ms, 1-kHz tone pulses presented at 75 dB SPL and combined with forward or backward masker at masker-target level differences of −10, 0, or +10 dB. The jnd elevation (relative to the jnd in quiet) increased with masker level and was only 1-3 dB (in units of 10 log[1 + $\Delta I/I$]) at the −10 and the 0-dB level differences, while at the 10-dB level difference, nearly 8 dB of jnd elevation were observed. Stellmack and Viemeister did not report jnd's in masking that were smaller than the jnd in quiet, as it was the case at the 25-dB SPL standard level in the present experiment.

To summarize, the data provide clear evidence for a mid-difference hump because difference limens were maximum at intermediate masker-standard level differences. It remains unclear whether it would be appropriate to also speak of a mid-level hump (albeit in a new definition) because the effect of the masker was stronger for the 55-dB SPL than for the 25-dB SPL standard. As maximum masker level was 100 dB SPL and the use of significantly more intense maskers seems to be precluded, it remains unclear whether a non-monotonic jnd function would also be observed at the 85-dB SPL standard level if extremely high masker intensities were used.

6.3 Experiment 2: Loudness matches and loudness variability

In Experiment 2, a three-tone loudness matching task in combination with the 2I, 2AFC interleaved-staircase procedure proposed by Jesteadt (1980) was used to obtain loudness matches and a measure of loudness variability (Schlauch and Wier, 1987). To be able to compare the effects of the masker-target level difference and target level on intensity jnd's and loudness enhancement, the same stimuli as in Experiment 1 were used and the same listeners participated in the experiment.

6.3.1 Method

6.3.1.1 Stimuli, apparatus

The same stimuli, temporal configuration, and apparatus as in Experiment 1 were used, except that no masker was presented in interval 2. The target was always presented in the first interval, followed by a comparison tone after a silent interval of 650 ms (Fig. 26). In forward-masked trials, a masker was presented in interval 1 only (ISI = 100 ms).

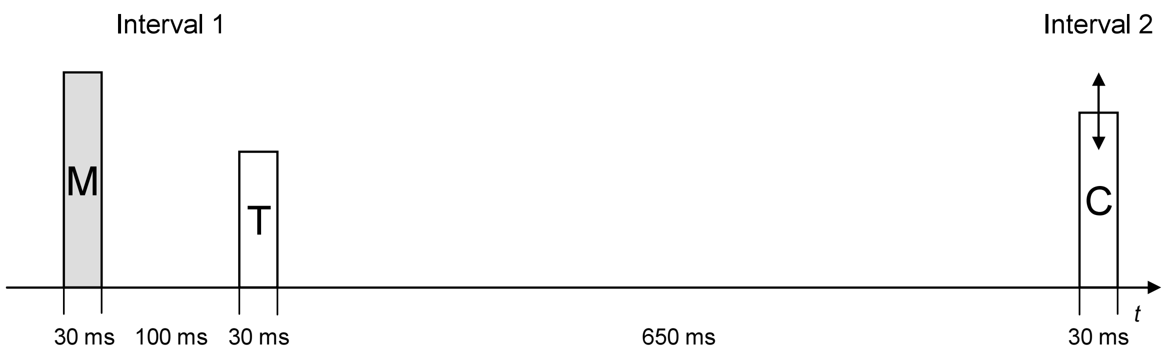

Fig. 26: Trial configuration used in Experiment 2. In interval 1, masker and target were presented. A comparison tone was presented in interval 2. Listeners responded whether the target or the comparison tone had been louder. Comparison tone level was adjusted according to an adaptive procedure using two interleaved tracks (Jesteadt, 1980) that converged on the 70.7 and 29.3% "Comparison Louder" points on the psychometric function, respectively.

Target and masker levels were identical to Experiment 1. Due to a technical problem, the 25-dB SPL target/10-dB SPL masker combination was not presented to listener SD and the 85-dB SPL target/25-dB SPL masker level was not presented to listener YS.

6.3.1.2 Listeners

With the exception of listener AS, who could not participate in Experiment 2 due to lack of time, all listeners from Experiment 1 also participated in Experiment 2.

6.3.1.3 Procedure

Listeners compared the loudness of the target tone presented in the first interval to loudness of the comparison tone presented in interval 2. In each trial, a listener responded whether tone 1 (the target) or tone 2 (the comparison tone) had been louder by pressing one of two response buttons. They were instructed to ignore the mask. Again, target and comparison tone were marked visually by two LEDs, just as in Experiment 1. No feedback was provided. The experiment was self paced.

Each run consisted of two randomly interleaved tracks (double-staircase procedure, Jesteadt, 1980). The upper track converged on the 70.7% “Comparison louder”-point on the psychometric function. After two consecutive “Tone 2 louder” responses, the intensity of the comparison tone was reduced. After each “Tone 1 louder” response, the intensity of the comparison tone was increased. In the lower track, a 2-up, 1-down rule was used to track the 29.3% “Comparison Louder” point on the psychometric function.

The upper track started with a comparison tone sound-pressure-level 11 dB above target level, while in the lower track, comparison tone level was 11 dB below target level initially. Step size was 5 dB until the fourth reversal, and 2 dB for the remaining eight reversals. If one of the tracks had already ended (because 12 reversals had occurred) before termination of the other track, it was still presented with an a priori probability of 0.15.

In each run, the arithmetic mean of the level differences between comparison tone and target at the last 8 reversals were computed for the upper and for the lower track. The arithmetic mean of these two values was taken as the loudness match $L_C - L_T$. A run was discarded if the standard deviation was greater than 5 dB in either the upper or the lower track. At least three runs were obtained for each data point. Time permitting, additional runs were presented if the standard deviation of $L_C - L_T$ obtained in the first three runs exceeded 5 dB; so each data point is based on 3 to 11 runs.

As Schlauch and Wier (1987) pointed out, the 70.7% and 29.3% “Tone 2 louder”-points on the psychometric function estimated by the above procedure can be used as “[...] essentially symmetrical estimates of the just-detectable increment and decrement, respectively” (p. 14). Therefore, half the difference between the two values was taken as a measure of loudness variability (Zeng, 1994).

The loudness match and the measure of loudness variability correspond to the point of subjective equality (PSE) and the jnd, respectively, obtained in the classical method of constant stimuli, where the fixed-level standard is presented in the first interval and a comparison tone of varying level is presented in the second interval. In the matching task, the upper track converges on the comparison tone level that is judged to be louder than the target in 70.7% of the trials. The lower track converges on the 29.3% “Comparison louder” point on the psychometric function. Under the assumption that the psychometric function is symmetric about the 50% point, the arithmetic mean of the 70.7% and the 29.3% point is just the PSE (50% point on the psychometric function, also denoted as $x_{0.5}$).

Similarly, in the method of constant stimuli, the just-noticeable-difference is traditionally defined as the difference between the 75% and the 50% point on the psychometric function ($x_{0.75} - x_{0.5}$), or as half the difference between the 75% and the 25 % point. It is obvious that the measure of loudness variability introduced above is equivalent to the latter jnd definition, so that we write

$$\mathrm{jnd}_{Match} = \frac{x_{0.707} - x_{0.293}}{2}, \qquad (6\text{-}1)$$

where the index “Match” is used to distinguish this measure from the jnd obtained in an intensity discrimination experiment (e.g., ΔL_{DL}).

It should be noted that the procedure used for obtaining the measure of loudness variability differs substantially from the usual intensity discrimination procedure (mask in both intervals, increment presented in either first or second interval, feedback). For this reason, it seems questionable whether the two values are compatible.

In each block, only one masker-target level combination was presented. Listeners received the conditions in pseudo-random order with the exception that blocks with the 100-dB SPL masker were always presented at the end of a session.

An experimental session lasted approximately one hour. Listeners took one or two short breaks in each session.

The listeners received both the intensity discrimination task (Experiment 1) and the loudness matching task of the present experiment in each session.

6.3.1.3.1 Relation to previous studies

The method used in this experiment is in many respects the same as in the studies by Zeng (1994) and Plack (1996 a). One notable difference is the position of the sound adjusted by the adaptive procedure. In each run, Zeng and Plack varied the intensity of the masked tone presented in interval 1 while the intensity of the tone presented in interval 2 was fixed. To use the standard terms, the comparison tone was presented in interval 1 and the target in interval 2.

As the present experiment was conducted to gain insight into the effects of the masker-target level difference, the setting used by Plack and Zeng has the disadvantage of this difference being different for the upper and the lower track and also varying during the course of each adaptive track. For this reason, it was decided to keep the intensity of the masked tone (the target) fixed and to adjust the level of the comparison tone presented in interval 2. This setting was also used in all older loudness enhancement experiments (e.g., Zwislocki and Sokolich, 1974; Elmasian and Galambos, 1975).

A second difference to the experiments of Zeng and Plack is the use of equal-duration masker and target. This has the advantage that masker and target presented at the same SPL and in isolation will sound equally loud, which is not the case if the tone durations differ (cf. Zwislocki, 1969; Florentine, Buus, and Poulsen, 1996).

6.3.2 Results

Individual loudness matches are displayed in Fig. 27. As discussed above, loudness enhancement is defined here as the difference between the level of a comparison tone that is perceived as equally loud as the target on the one hand and target level on the other hand ($L_C - L_T$). Positive values of the comparison-target level difference correspond to loudness enhancement, negative values to loudness decrement. This is the definition used by Elmasian, Galambos, and colleagues (e.g., Elmasian and Galambos, 1975). Zwislocki and colleagues (e.g., Irwin and Zwislocki, 1969) defined loudness enhancement to be the difference between the sound pressure level of the comparison tone matching target loudness in forward masking and the sound pressure level of the comparison tone matching target loudness in quiet.

The results for listener BS, who had produced extremely large jnd's at large masker-target level differences in Experiment 1, again strongly deviated from the data produced by the remaining listeners in conditions where a low-level target was combined with an intense masker. The very large amount of loudness enhancement displayed in Fig. 27 for the 25-dB SPL target combined with the 85-dB SPL masker even underestimates the 'true' loudness match because in several trials, requested comparison tone level exceeded the maximum level difference $L_C - L_T$ of 45 dB the test equipment could deliver. Due to these difficulties, the listener was only tested in a limited set of conditions (in quiet, 25-dB SPL target plus 70-dB and 85-dB SPL masker, 55-dB SPL target plus 85-dB SPL masker, and 85-dB SPL target plus 85-dB SPL masker). Apart from the low-level target/intense masker conditions, the loudness matches produced by BS did not differ substantially from the results of the remaining listeners. The data for BS were excluded from the statistical analyses reported below.

For all listeners except BS, loudness enhancement was a non-monotonic function of the masker-target level difference at a target level of 25 dB SPL, as the maximum amount of loudness enhancement was found at values of $\Delta L_{M\text{-}T} = L_M - L_T$ between 15 and 45 dB, with individual maxima of 4.0 to 9.0 dB. The 10-dB SPL masker resulted in loudness decrement.

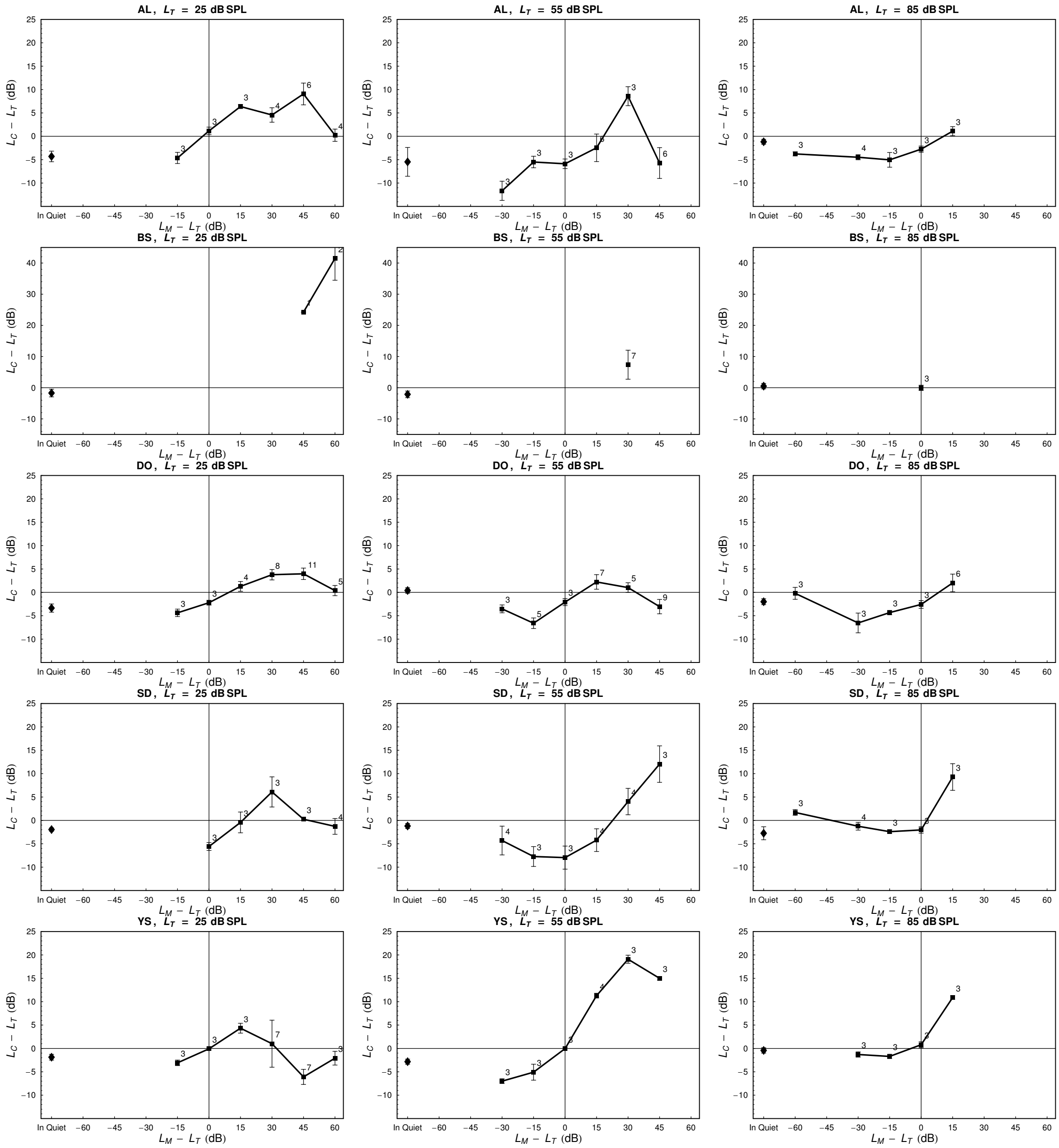

Fig. 27: Individual loudness matches ($L_C - L_T$) as a function of the masker-target level difference and target level L_T. Rows represent listeners, columns target levels. Diamonds: in quiet. Squares: in forward masking. Numbers denote the number of runs a data point is based on. Error bars show ± 1 SEM. Note the different scale of the y-axis used for listener BS.

Mid-difference humps were also present at the 55-dB SPL target level for all listeners except SD. At masker-target level differences between 15 and 30 dB, the maximum amounts of enhancement were observed (2.2 to 19.1 dB). At the 55-dB SPL target level, decrement and enhancement were more pronounced than at the remaining two target levels for listeners SD and YS.

The 100-dB SPL masker caused loudness of the 85-dB SPL target to be enhanced.

Some listeners produced negative matches for 25-dB or 55-dB SPL targets combined with intense maskers (AL, DO, YS). In this context, it is interesting to note that loudness matches in quiet were also frequently negative, as it has previously been found in experiments studying the time order error (cf. Hellström, 1985). Still, the loudness match for the 25-dB SPL target at a masker-target level difference of 45 dB were below the matches in quiet for listener YS. A related issue is the difference between matches in quiet and for $L_M = L_T$ observed in several cases. Such differences are also present in the previous loudness matching data, e.g., Elmasian and Galambos (1975).

Generally, inter- and intra-listener variability was larger than for the difference limens measured in Experiment 1. To give an example, maximum loudness enhancement at the 55-dB SPL target level was 19.1 dB for listener YS but only 2.2 dB for listener DO. Some listeners also produced widely varying matches in a single condition, as for example YS for the 55-dB SPL masker/25-dB SPL target combination.

Repeated-measures analyses of variance (ANOVAs) were conducted using a maximum-likelihood approach (SAS PROC MIXED; Littell et al., 1996). The restricted maximum-likelihood method was used to estimate the covariance parameters and the Satterthwaite method for computing denominator degrees of freedom for approximate *F*-tests of fixed effects. The “compound symmetry heterogeneous” (CSH) was selected to model the covariance structure.

In the mean data (Fig. 28; note that the data from listener BS were excluded), loudness enhancement was a non-monotonic function of $\Delta L_{M\text{-}T}$ at the 25-dB SPL target level. A one-factorial ANOVA conducted for the data obtained at this target level indicated a significant effect of $\Delta L_{M\text{-}T}$, $F(5, 5.84) = 7.2$, $p < 0.018$. The observation of a mid-difference hump was confirmed by a significant quadratic trend, $F(1, 11.5) = 8.71$, $p < 0.013$. The linear trend was also significant, $F(1, 10.4) = 11.8$, $p < 0.007$. The cubic trend did not reach significance.

For the 55-dB SPL target, loudness enhancement was also a non-monotonic function of $\Delta L_{M\text{-}T}$, so that the maximum loudness change was observed for the 30 dB level difference between masker and target. At $\Delta L_{M\text{-}T} = -30$ dB, loudness decrement was rather large, but unfortunately more negative level differences were not presented at this target level so that it remains unclear at which value of $L_M - L_T$ loudness decrement can be expected to decrease again. As masker level is reduced towards detection threshold, however, the loudness change effected by the masker must necessarily vanish as the task eventually becomes equivalent to loudness matching in quiet. Note that for two listeners, loudness decrement was smaller at the −30 than at the −15 dB level difference.

A one-factorial ANOVA conducted for the forward-masked data at the 55-dB SPL target level confirmed a significant effect of $\Delta L_{M\text{-}T}$, $F(5, 4.43) = 6.74$, $p < 0.037$. The cubic trend was significant, $F(1, 4.02) = 12.58$, $p < 0.024$. This result is compatible with the predictions of the similarity model (Fig. 19), according to which the loudness change should be maximum at both positive and negative intermediate values of $\Delta L_{M\text{-}T}$ (e.g., −30 dB and +30 dB). The quadratic trend was not significant, $F(1, 4.51) = 0.01$. There was a significant linear trend, $F(1, 4.02) = 12.58$, $p < 0.024$.

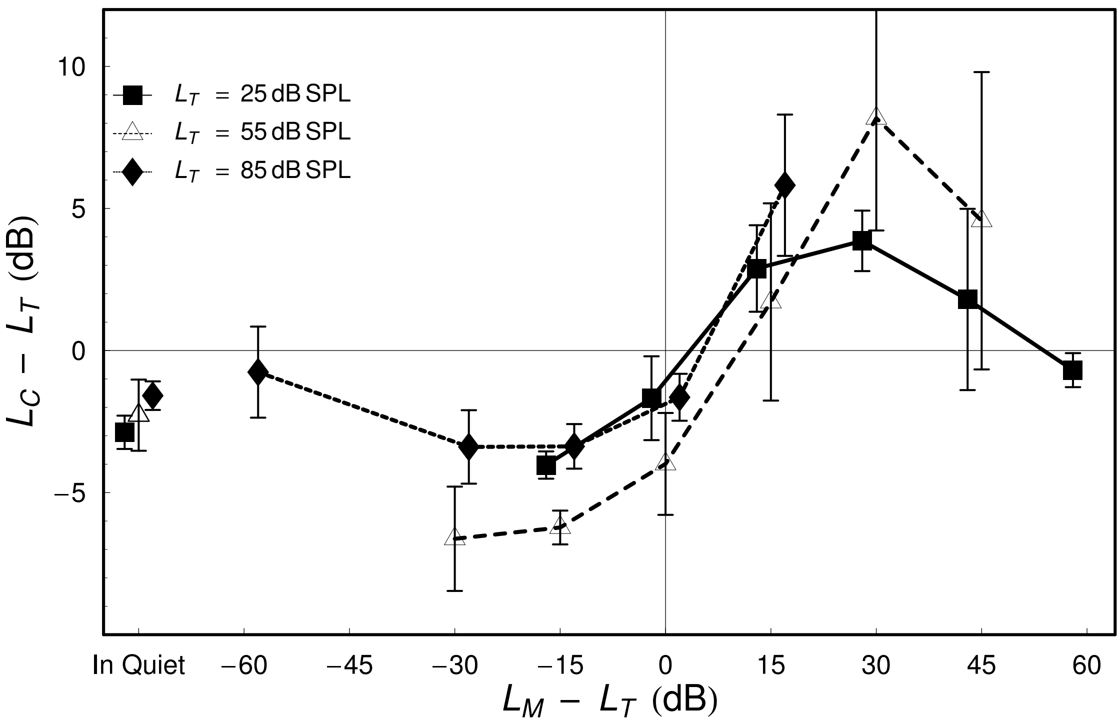

Fig. 28: Mean loudness matches ($L_C - L_T$) as a function of the masker-target level difference $\Delta L_{M\text{-}T}$ for four listeners, data from listener BS excluded. Boxes: 25-dB SPL target. Triangles: 55-dB SPL target. Diamonds: 85-dB SPL target. To avoid cluttering, lines were shifted by 2 dB on the x-axis. Error bars show ± 1 SEM.

At the 85-dB SPL target level, mean loudness decrement was maximal at the −30 and −15 dB level difference between masker and target. Loudness matches at the −60 dB and 0 dB level differences corresponded approximately to the match in quiet. At $\Delta L_{M\text{-}T}$ = 15 dB, loudness was enhanced. The effect of $\Delta L_{M\text{-}T}$ was marginally significant at this target level, $F(4, 4.02) = 6.32$, $p < 0.051$. There was a significant quadratic trend, $F(1, 4.22) = 10.41$, $p < 0.03$, confirming the statistical validity of the *U*-shaped pattern. The linear and cubic trends did not reach significance.

For a level difference of 0 dB ($L_M = L_T$), the similarity model predicts the loudness change to be 0. In the mean data, $L_C - L_T$ was slightly negative at all three target levels in this condition, however, and the forward-masked matches were approximately identical to the matches in quiet at the 25-dB SPL and 85-dB SPL target level, while it was more negative than in quiet for the 55-dB SPL target. An L_T (25, 55, 85 dB SPL) × Condition (In Quiet, $\Delta L_{M\text{-}T}$ = 0 dB) ANOVA was used to examine these differences. Neither the effect of target level ($F[2, 11.2] = 2.78$), nor the effect of Condition ($F[1, 12.4] = 0.69$), nor the L_T × Condition interaction ($F[2, 11] = 1.2$) were significant.

Data for the −15, 0, and 15 dB masker-target level differences were obtained at each target level. The effect of L_T on the masker-induced loudness changes in these conditions seems to be a shift along the *y*-axis related to the more negative loudness match for the 55-dB SPL target obtained in the $L_M = L_T$ condition. In an L_T (25, 55, 85 dB SPL) × $\Delta L_{M\text{-}T}$ (−15, 0, 15 dB) ANOVA, the effect of target level was indeed significant, $F(2, 20.1) = 4.63$, $p < 0.022$. There was also a significant effect of $\Delta L_{M\text{-}T}$, $F(2, 9.97) = 17.5$, $p < 0.001$.

Mean loudness enhancement caused by a masker 15 dB above target level was most pronounced for the 85-dB SPL target. The $L_T \times \Delta L_{M\text{-}T}$ interaction did not reach statistical significance, however, $F(4, 9.8) = 0.63$.

Mean loudness enhancement was more pronounced at the 55-dB SPL than at the 25-dB SPL target level. Yet, in an L_T (25, 55, 85 dB SPL) × $\Delta L_{M\text{-}T}$ (−15 to 45 dB) ANOVA, the $L_T \times \Delta L_{M\text{-}T}$ interaction was not significant, $F(4, 7.13) = 1.23$. The failure to reach significance can be attributed to the very large variability of loudness enhancement at values of $\Delta L_{M\text{-}T}$ between 15 dB and 45 dB that is due to the large inter-subject variability in these conditions (Fig. 28). The effect of $\Delta L_{M\text{-}T}$ was significant, $F(4, 6.76) = 15.86$, $p < 0.002$, while there was no significant effect of L_T, $F(1, 7.8) = 0.06$.

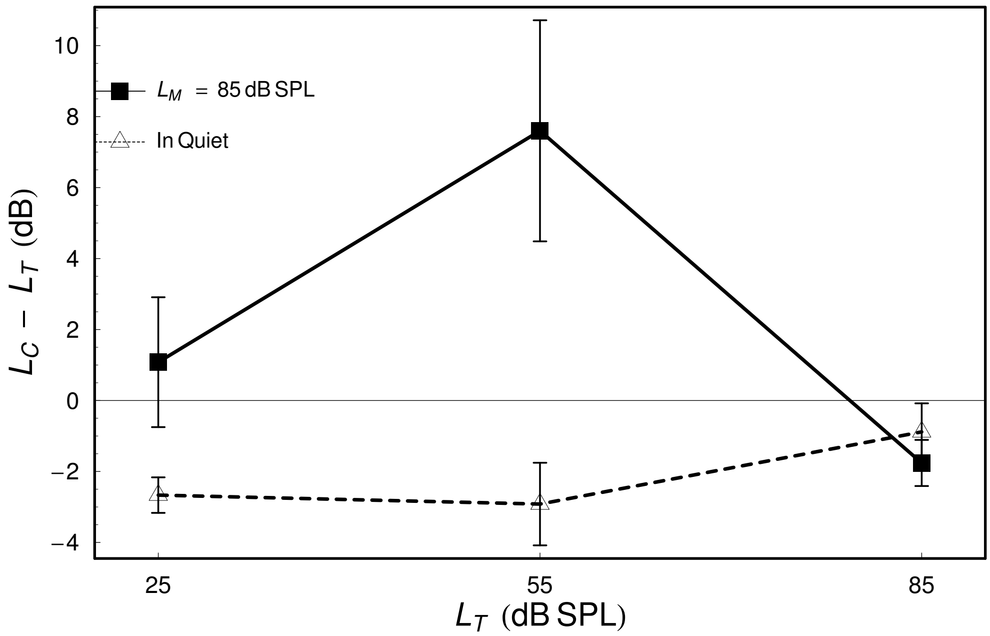

Fig. 29: The 'midlevel hump' in loudness enhancement: mean comparison-target level difference for four listeners (data from BS excluded) plotted as a function of target level L_T. Boxes: 85-dB SPL forward masker. Triangles: in quiet. Error bars show ± 1 SEM.

In Fig. 29, the data obtained with the 85-dB SPL masker are plotted to allow a comparison with the experiments by Zeng (1994) and Plack (1996 a), who displayed their results in this 'midlevel hump' form.

6.3.3 Loudness variability jnd_{Match}

The effect of forward masking on the measure of loudness variability, jnd_{Match} (half the distance between the 70.7% and the 29.3% "Comparison louder" point on psychometric function) differs considerably from the intensity jnd's obtained in Experiment 1.

Only listener DO produced a very small mid-difference hump (3 dB) at the 55-dB SPL target level, while in the same condition, his difference limens increased monotonically with $\Delta L_{M\text{-}S}$ in Experiment 1 (Fig. 30).

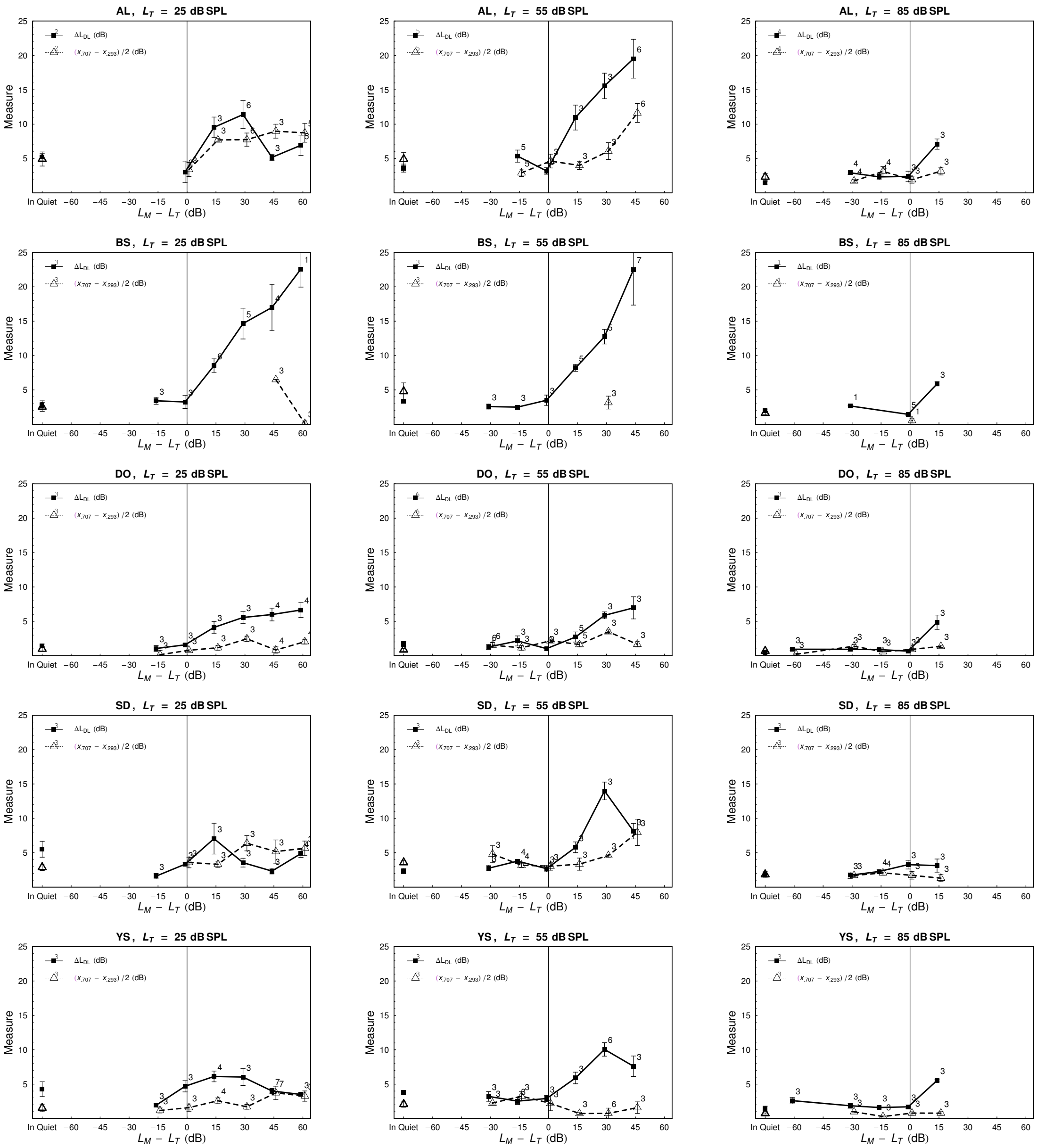

Fig. 30: Individual loudness variability (jnd$_{Match}$, triangles) as a function of the masker-target level difference and target level. For comparison, difference limens from Experiment 1 are also shown (squares). Rows represent listeners, columns target levels. Numbers denote the number of runs a data point is based on. Error bars show ± 1 SEM.

The pronounced mid-difference humps observed for listeners SD and YS at the 55-dB SPL standard level in Experiment 1 were not present for jnd$_{Match}$ obtained in Experiment 2. Instead, loudness variability increased with masker-target level difference. Similarly, the mid-difference humps found at the 25-dB SPL standard level in Experiment 1 for three listeners did not show for jnd$_{Match}$.

Generally, at masker levels smaller than or equal to target level, the two values were approximately identical. For positive masker-target level differences, jnd$_{Match}$ showed either virtually no (listeners DO and YS, all listeners at the 85-dB SPL target level) or only a smaller increase than ΔL_{DL}. In the latter case, no decrease at large values of $\Delta L_{M\text{-}T}$ was observed.

Across all data, the correlation between jnd_{Match} and ΔL_{DL} was significant but very small, $r = 0.464$, $p < 0.001$, $N = 66$.

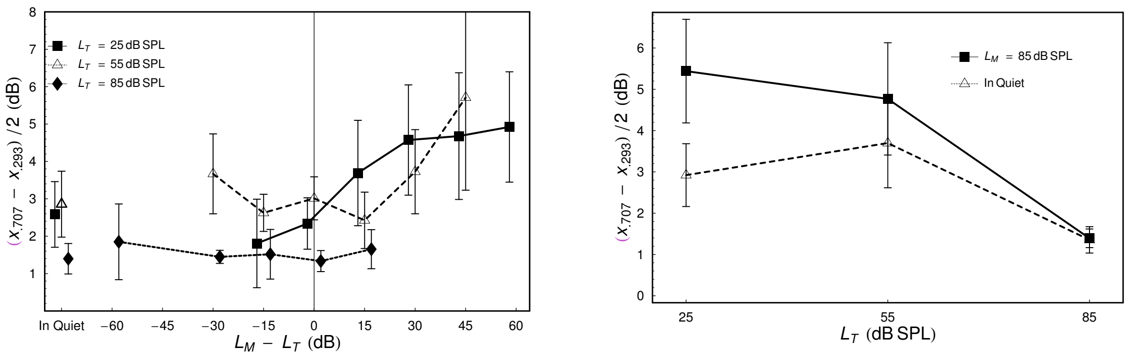

Fig. 31: Left panel: Mean loudness variability (jnd_{Match}) as a function of the masker-target level difference for four listeners, data from listener BS excluded. Boxes: 25-dB SPL standard. Triangles: 55-dB SPL standard. Diamonds: 85-dB SPL standard. To avoid cluttering, lines were shifted by 1 dB along the x-axis.
Right panel: Mean loudness variability (jnd_{Match}) observed with the 85 dB SPL masker and in quiet. Error bars show ± 1 SEM.

In the mean data (Fig. 31; data from BS are again excluded), there was a monotonic increase of jnd_{Match} with $\Delta L_{M\text{-}T}$ at the 25-dB SPL target level. For the 55-dB SPL target, jnd_{Match} was a *U*-shaped function of $\Delta L_{M\text{-}T}$. It was close to the value in quiet for masker-target level differences between − 15 and 15 dB, but elevated at the remaining values of $\Delta L_{M\text{-}T}$. For the 85-dB SPL target, masker level had virtually no effect. Surprisingly, for the data obtained with the 85-dB SPL masker, there was no mid-level hump (Fig. 31, right panel). Instead, jnd_{Match} decreased with target level. Therefore, the data differ not only from the intensity discrimination difference limens measured in Experiment 1, but also from the loudness variability data reported by Zeng (1994) and Plack (1996 a), who presented a fixed 90-dB SPL masker. It should be noted that Plack (1996 a) also reported a rather flat jnd_{Match} function for one of his listeners.

6.3.4 Discussion

The observed masker-induced loudness changes are in good accordance with the predictions of the similarity model, i.e., the mergence hypothesis extended by the aspect of perceptual similarity between masker and target.

The loudness change effected by the masker was a non-monotonic function of the masker-target level difference for all listeners (except BS) at the 25-dB SPL standard level, and for all listeners but SD at the 55-dB SPL target level. The mean effect of the masker was more pronounced at the intermediate than at the lowest target level, although this difference was not statistically significant. In the quantitative version of the similarity model presented below, larger loudness changes caused by the masker are predicted at higher target levels due to the shape of the loudness function for pure tones.

Incompatible with the similarity model are the frequent negative loudness matches $L_C - L_T$ observed for the $L_T = L_M$ condition, as the model predicts a 0 dB loudness change in this condition. This finding must be viewed in relation to the fact that loudness matches in quiet were also negative, however. The level of the conditioner matching loudness of the target was smaller than target level even if no masker was present (time order error; Hellström, 1985). In fact, there was no significant difference between the loudness matches in quiet and for the case of equal-level masker and target. Put differently, mean loudness enhancement defined as the difference between the level of the comparison tone matching the forward-masked and the unmasked target, respectively (Irwin and Zwislocki, 1971), was not significantly different from zero if masker and target level were identical. Still, in the individual data, the loudness match in quiet frequently differed from the loudness match in the $L_M = L_T$ situation.

It remains unclear, why the pattern of results obtained for the measure of loudness variability, jnd_{Match}, differed so clearly from the intensity jnd data obtained in Experiment 1. As discussed in Chapter 6.3.1.3, jnd_{Match} and ΔL_{DL} certainly do not measure exactly the same thing, but this does still not explain why no midlevel hump for jnd_{Match} was found as in previous experiments. One possible, but speculative explanation would be that the masker introduces a response bias in the 2I, 2AFC experiment that results in an increase in ΔL_{DL}. This idea will be discussed in greater detail in Chapter 7.2.4.2. Briefly, consider a signal-detection model of the decision in an intensity-discrimination experiment. The listener is assumed to make an observation of target tone intensity in interval 1 and in interval 2 and to base the response on the difference between these observations. If it is now assumed that the masker has a stronger effect on the observation made in interval 1 than on the observation made in interval 2, an intense masker combined with a less intense standard will cause a bias towards responding “First tone louder” if the listener does not adjust his or her decision criterion accordingly. Such a bias results in a drop in performance (cf. Green and Swets, 1966). The important difference between the 2I, 2AFC intensity discrimination task and the loudness matching procedure used to measure jnd_{Match} is that in the latter task, a bias towards responding, e.g., “First tone louder” would result in comparison tone level being increased. More specifically, the adaptive procedure adjusts mean comparison tone level so that across the two interleaved tracks, the proportion of “First tone louder” and “Second tone louder” responses is identical. In other words, the response bias induced by the masker is compensated for by the adaptive procedure. Still, intensity resolution defined as half the difference between the 70.7% and the 20.3% “Comparison tone louder”-points on the psychometric function could vary with the experimental parameters. If one uses the terms of the method of constant stimuli, the jnd increase observed in the 2I, 2AFC intensity discrimination experiment reflects both the constant error *and* intensity resolution in the case of a response bias, while jnd_{Match} measures, by definition, the intensity resolution only. Therefore, the concept of a response bias caused by the masker could explain why jnd_{Match} was smaller than ΔL_{DL} in the vast majority of conditions (Fig. 30). As it will be demonstrated more formally in Chapter 7.2.4.2, the main argument *against* this explanation is the lack of jnd elevations in Experiment 1 at *negative* masker-standard level differences. As a response bias should also be introduced by maskers presented at levels *below* standard level, considerable jnd elevations can be expected in this condition if response bias were indeed the cause of the masker-induced drop in performance. For this reason, response bias does not seem to provide an explanation for the differences between ΔL_{DL} and jnd_{Match}. An intensity discrimination experiment using the method of constant stimuli could be used to measure both the response criterion and intensity resolution. Such data would allow testing the hypothesis that the masker introduces a response bias.

6.4 Experiment 3: Masker duration

Schlauch et al. (1997) found that the use of different durations for masker and standard, respectively, reduced the jnd elevation caused by a forward masker and argued that this finding could be attributed to the decreased perceptual similarity between masker and target.

In terms of the similarity model, any manipulation rendering masker and standard less similar perceptually should reduce both loudness enhancement and forward-masked jnd's. In Experiment 3, a 100-ms masker was combined with a 30-ms standard, and intensity jnd's as well as loudness matches were obtained for some of the conditions also presented in Experiments 1 and 2.

The similarity model predicts two effects.

First, the loudness of 100-ms tone well above threshold can be expected to exceed the loudness of a 30-ms tone by a factor of 1.35 according to the neural integration model by Zwislocki (1969). This should result in stronger loudness enhancement at small masker-target level differences.

On the other hand, compared with the same-duration masker condition, the larger perceptual distance should result in masker loudness receiving less weight at all masker-target level combinations, so that the increase in loudness enhancement should be rather small.

6.4.1 Method

6.4.1.1 Stimuli, apparatus, procedure

Except for the use of a 100-ms instead of a 30-ms masker, the experimental setting was identical to Experiment 1 and Experiment 2.

For listener BS, who had shown a pattern of results clearly deviating from the remaining subjects (no indication of a non-monotonic jnd function, very large jnd and loudness enhancement obtained with intense maskers combined with the 25-dB SPL target), it was hypothesized that making mask and target perceptually different by using different durations might remove some of these anomalies. Therefore, the complete set of conditions from Experiment 1 was tested again with the longer masker. Due to lack of time and because BS had frequently requested comparison levels that exceeded the maximum output of the test equipment in Experiment 2, loudness matches were obtained only for the 85-dB SPL masker combined with the 25-dB SPL target.

For listener DO and SD, only conditions in which decreased perceptual similarity was expected to have the most pronounced effects were included.

6.4.1.2 Listeners

Three listeners from the preceding experiments participated.

6.4.2 Difference limens

As expected, the 100-ms masker resulted in a decrease of ΔL_{DL} at the maximum masker-target level difference for listener BS, so that even a slight indication of a mid-difference hump was present at the 25-dB SPL standard level (Fig. 32). At smaller values of $\Delta L_{M\text{-}S}$, difference limens were either smaller than or equal to the jnd's obtained with the 30-ms masker. These findings support the notion that in the same-duration condition, the listener was unable to ignore the masker due to perceptual similarity.

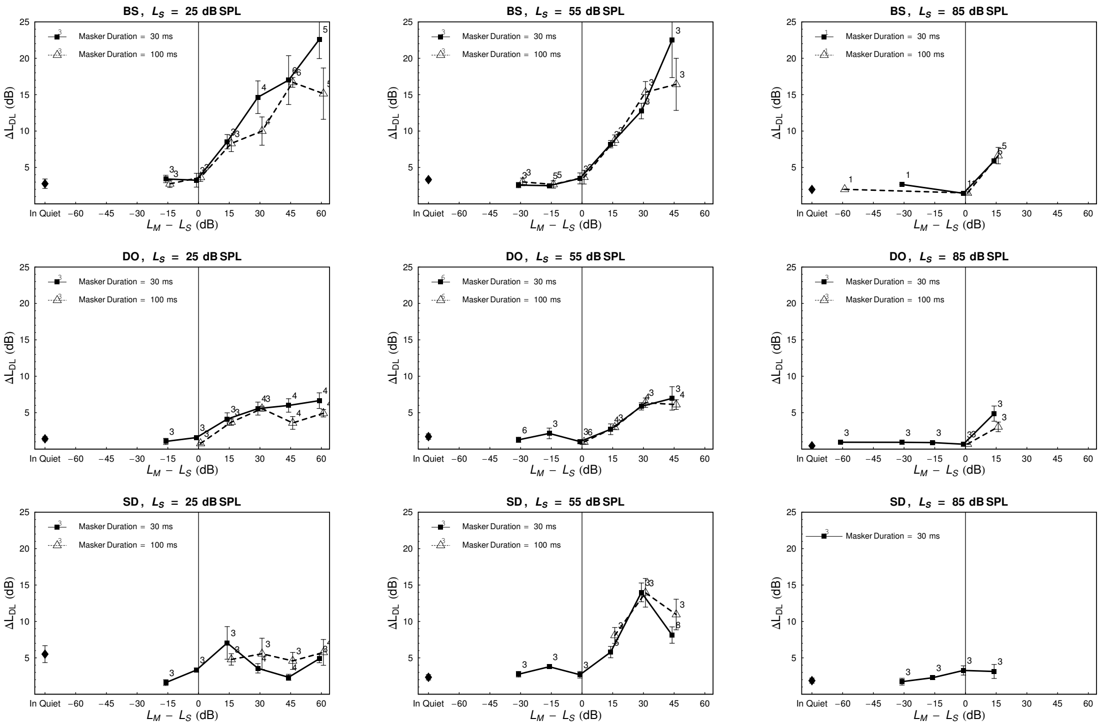

Fig. 32: Jnd's *plotted as a function of the masker-standard level difference. Rows represent listeners, columns standard levels. Diamonds: in quiet. Squares: 30-ms forward masker (Experiment 1). Open triangles: 100-ms forward masker (Experiment 3). Numbers denote the number of runs a data point is based on. Error bars show ± 1 standard error of the mean (SEM).*

Difference limens obtained with the 100-ms masker were smaller for listener DO, who now also showed a weak indication of a mid-difference hump at the 25-dB SPL standard level.

For listener SD, the 100-ms masker resulted in an **increased** difference limen in the 100-dB SPL masker/55-dB-SPL standard condition, while the jnd function became essentially flat at the low standard level, so that jnd's in quiet and in forward masking were identical in this condition.

6.4.3 Loudness enhancement

As for intensity discrimination, the 100-ms masker resulted in a greatly reduced amount of loudness enhancement for listener BS in the 85-dB SPL masker / 25-dB SPL target condition (Fig. 33).

For listener DO, the loudness change effected by the 100-ms masker was generally smaller than for the 30-ms masker.

At the largest masker level in the 55-dB SPL target condition, the 100-ms masker caused a reduction in the amount of loudness enhancement for listener SD, so that unlike in the 30-ms masker condition, a mid-difference hump was observed. This finding is surprising as SD's difference limen was slightly **larger** for the longer masker duration in this condition. Loudness enhancement was slightly increased by the 100-ms masker at the 25-dB SPL target level.

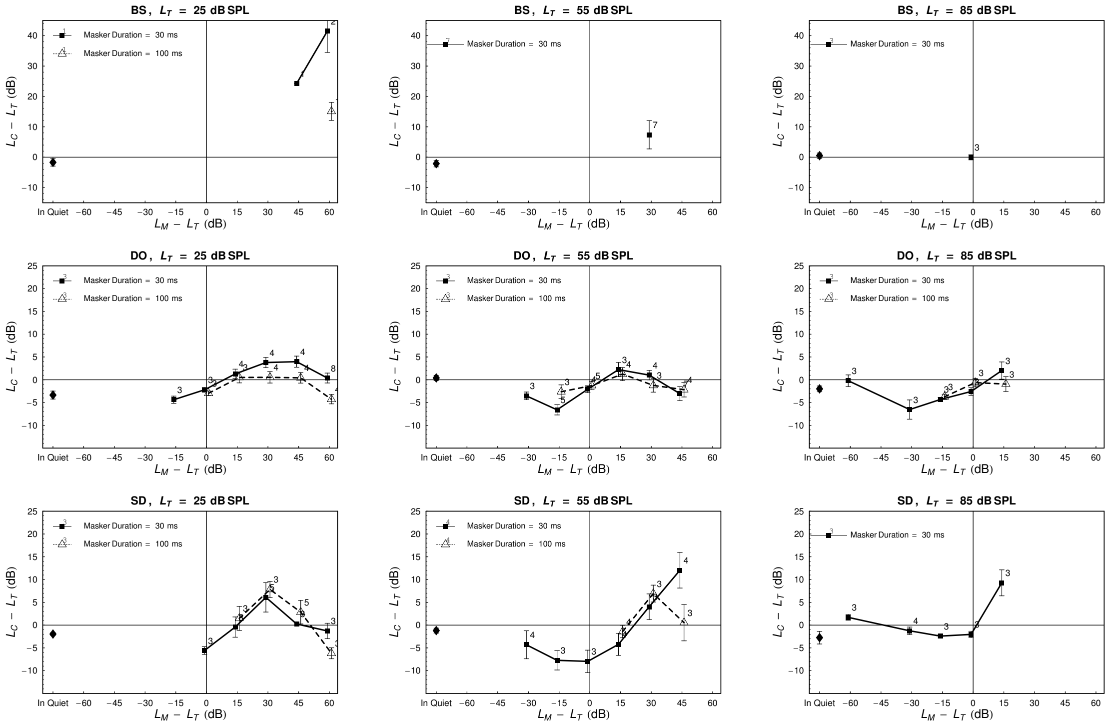

Fig. 33: Loudness enhancement observed with 30-ms and 100-ms maskers as a function of the masker-target level difference. Rows represent listeners, columns standard levels. Diamonds: in quiet. Squares: 30-ms forward masker (Experiment 2). Open triangles: 100-ms forward masker (Experiment 3). Numbers denote the number of runs a data point is based on. Error bars show ± 1 standard error of the mean (SEM). Note the different scale of the y-axis used for listener BS.

6.4.4 Discussion

The effects of presenting a masker longer in duration than the standard are compatible with the predictions of the similarity model that predicts a decrease in the weight assigned to masker loudness in this condition, resulting in less loudness enhancement and thus less loudness variability. The effects of masker duration were clearly smaller than in the experiments by Schlauch et al. (1997). A straightforward explanation would be that the ratio between long and short masker duration was only 3 in the present experiment, but 25 in the experiments by Schlauch et al., where the two durations were 250 ms and 10 ms, respectively. It can be concluded that it would have been advantageous to test a more extreme range of durations in the present experiment.

6.5 Auxiliary experiment: Intensity discrimination using a two-tone masker, listener BS

The 100-ms masker presented in Experiment 3 reduced, but did not completely remove the strong jnd elevation observed for BS with the 25-dB SPL standard and intense maskers.

When interviewed concerning his strategy, BS reported great difficulties in ignoring the masker in this situation. It should be noted, that not all tracks presenting the 85-dB SPL masker/25-dB SPL standard combination resulted in large difference limens. As Fig. 34 shows, some of the jnd's were quite small, as well for the 30-ms as for the 100-ms masker. There was no evidence for a practice effect.

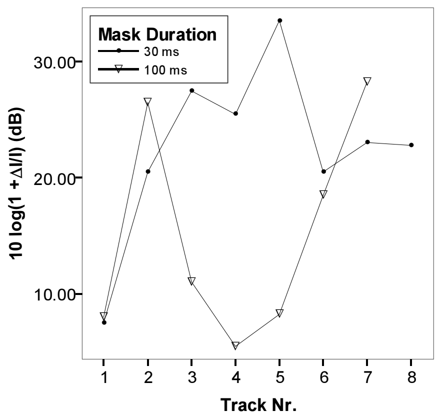

Fig. 34: Intensity discrimination difference limens produced by listener BS in tracks presenting the 85-dB SPL masker / 25-dB SPL standard combination. Dots: 30-ms masker (Experiment 1). Triangles: 100-ms masker (Experiment 3).

Schlauch et al. (1997) found that a two-tone masker consisting of a 1-kHz and a 4.133 kHz component greatly reduced jnd's in the intense masker plus low level standard condition for a listener in one of their experiments, who had produced even larger difference limens than BS. The 4.133-kHz component should cause no peripheral masking for the 1-kHz standard, but can be expected to strongly decrease the perceptual similarity between masker and standard.

Intensity jnd's were measured for the two-tone masker in the three masker-standard level combinations that had resulted in the largest jnd's in Experiment 1. Except for the different masker, the same experimental setup and procedure as in Experiment 1 was used. The two-tone masker consisted of a 1-kHz and a 4.133-kHz tone burst sharing the same temporal envelope (5-ms $\cos^2$ ramps, 25-ms steady-state duration). The sound pressure level of the two components was identical so that overall sound pressure level was 3 dB greater than the level of the 1-kHz tone. In the following, masker level is specified as the sound pressure level of the 1-kHz component.

As seen in Fig. 35, the two-tone masker (stars) resulted in a further reduction of the difference limens compared to the 100-ms masker condition (triangles). Unfortunately, at the 25-dB SPL target level, no data were collected with intermediate two-tone masker-target level differences, so that it remains unclear whether a mid-difference hump would have been observed in this condition.

Taken together, the results from the 30-ms, 100-ms, and two-tone masker conditions confirm the hypothesis that the exceptionally large difference limens and amounts of loudness enhancement BS produced in Experiment 1 and Experiment 2 were due to difficulties in distinguishing between the sensations elicited by masker and target, respectively.

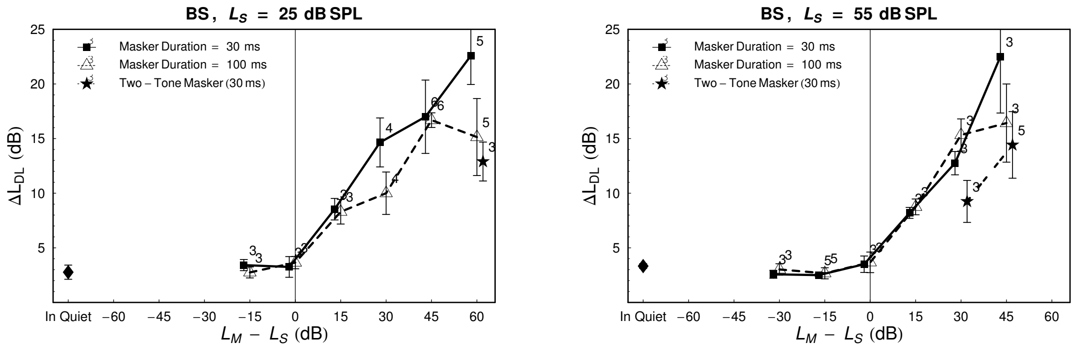

Fig. 35: Intensity jnd's *for listener BS plotted as a function of the masker-standard level difference. Each panel represents one standard level. Diamonds: in quiet (Experiment 1). Stars: two-tone masker. Squares: 30-ms forward masker (Experiment 1). Open triangles: 100-ms forward masker (Experiment 3). Numbers denote the number of runs a data point is based on. Error bars show ± 1 standard error of the mean (SEM).*

6.6 Relation between loudness enhancement and intensity jnd's

The similarity model predicts loudness enhancement and the jnd elevation caused by the forward masker to be correlated, as increased loudness variability is the cause of reduced intensity resolution according to the loudness enhancement hypothesis.

Only two previous experiments measured both loudness enhancement and difference limens for the same listeners and using the same stimuli. With a 90-dB, 100-ms forward masker, Zeng (1994; see his Fig. 3) found a very similar dependence of mean loudness enhancement and jnd_{Match} (the measure of loudness variability obtained in the loudness matching experiment) on target level, although he reported no correlation coefficient. A midlevel hump was present for both measures. Intensity jnd's reported for the same listeners in Zeng and Shannon (1995) also exhibited essentially the same pattern (Fig. 5 in Zeng, 1994).

Plack (1996 a) used the same procedure as Zeng (1994) and presented both forward and backward maskers. Loudness enhancement and jnd_{Match} were significantly correlated. Spearman's rank-correlation coefficient r_S was 0.658 under backward masking. For the forward masking situation, the correlation was negative for one listener. If the data from this listener were excluded, the rank-correlation coefficient was 0.765.

Using the data from Experiments 1, 2, and 3 of this thesis, it is now possible to answer the question as to whether a correlation between loudness enhancement on the one hand and loudness variability and intensity jnd's on the other hand is present only in the 'midlevel hump setting', where masker level is fixed (90-dB SPL), or if it is also observed if a wide range of masker levels is used.

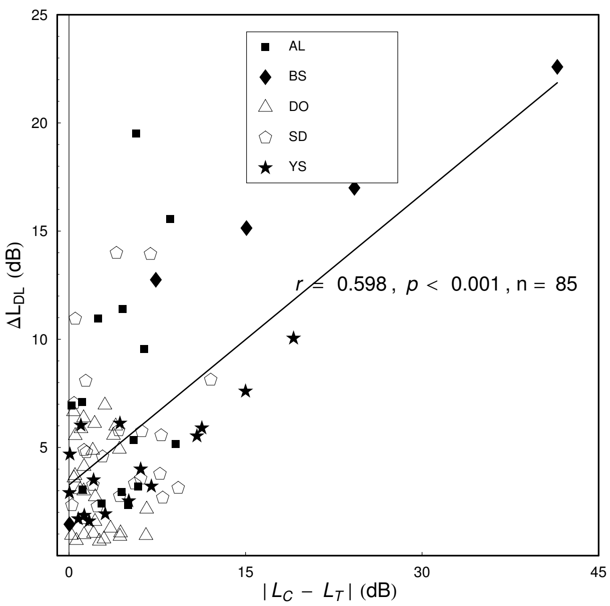

Fig. 36: Correlation between the absolute amount of loudness change $|L_C - L_T|$ and ΔL_{DL} for the five listeners who participated in both Experiment 1 and Experiment 2. Symbols denote listeners; the line shows the best fitting linear regression line.

Recall that the similarity model predicts the jnd to increase with the *absolute value* of $L_C - L_T$, as both loudness enhancement and loudness decrement should increase loudness variability. For the complete set of data points obtained in Experiment 1 to 3, the Pearson product-moment correlation coefficient differed significantly from zero, $r = 0.598$, $N = 85$, $p < 0.001$ (two-tailed). The data are displayed in Fig. 36. The proportion of variance accounted for by linear regression of ΔL_{DL} on $|L_C - L_T|$ was $R^2 = 0.36$, which is rather weak but comparable to the correlations reported by Plack (1996 a).

Although the correlation present in the complete data set supports the similarity model, individual correlations between the absolute value of the loudness change and ΔL_{DL} were significant for listener YS ($r = 0.778$, $n = 16$, $p < 0.001$) and BS ($r = 0.909$, $n = 5$, $p < 0.033$) only (Fig. 37, left column).

For listener DO, there was a non-significant *negative* correlation between the absolute value of the loudness change and ΔL_{DL}. As seen in the center column of Fig. 37, negative loudness changes corresponded to low values of ΔL_{DL} for this listener, so that the correlation between the jnd and the 'raw' values of $L_C - L_T$ (going from negative to positive values) was significant, $r = 0.466$, $n = 26$, $p < 0.013$. For listener AL, the correlation between the same variables was significant only if the data point with the largest value of ΔL_{DL} was excluded ($L_T = 55$ dB SPL, $L_M = 100$ dB SPL); the correlation coefficient ($r = 0.625$, $n = 13$, $p < 0.023$) displayed in the center column of Fig. 37 reflects this condition. For listener SD, use of the 'raw' values of $L_C - L_T$ rather than the absolute values resulted in a stronger but still not significant correlation. Across all listeners, the correlation between the loudness change $L_C - L_T$ and ΔL_{DL} was significant, $r = 0.640$, $p < 0.001$, $N = 85$.

Now it could be argued that the appropriate measures to correlate are ΔL_{DL} and the loudness change relative to the match *in quiet.* The similarity model predicts loudness variability due to loudness changes induced by the masker to be the cause of impaired intensity resolution. Loudness differences between target and conditioner observed in quiet can be attributed to trace drift or a response bias, so that the loudness variability argument does not apply in this case. Therefore, loudness enhancement according to the Zwislocki et al. definition (L_C in forward masking minus L_C in quiet) was computed for each listener in each condition. The original rather than the absolute value of the so-computed loudness enhancement was used. Across all listeners, the correlation between the loudness change relative to the match in quiet and ΔL_{DL} was significant, $r = 0.669$, $p < 0.001$, $N = 85$, which is slightly larger than the correlation computed for the independent variable $L_C - L_T$. Individual correlations displayed in the right column of Fig. 37 were significant for all listeners except SD; note that for AL, the data point with the largest value of ΔL_{DL} was again excluded. In the discussion of the loudness matching data, it was pointed out that for listener SD, the loudness match in quiet differed from the loudness match in the $L_M = L_T$ situation (Fig. 27). The reason for this behavior is unclear, as the similarity model predicts the match at $L_M = L_T$ to be zero. If time-order errors are taken into account, the match can be expected to be identical to the match in quiet. Using the loudness match obtained for masker level equal to target level as the reference value in computation of the amount of loudness enhancement resulted in a significant correlation between the so-computed loudness change and ΔL_{DL} for listener SD ($r = 0.578$, $p < 0.005$, $n = 22$).

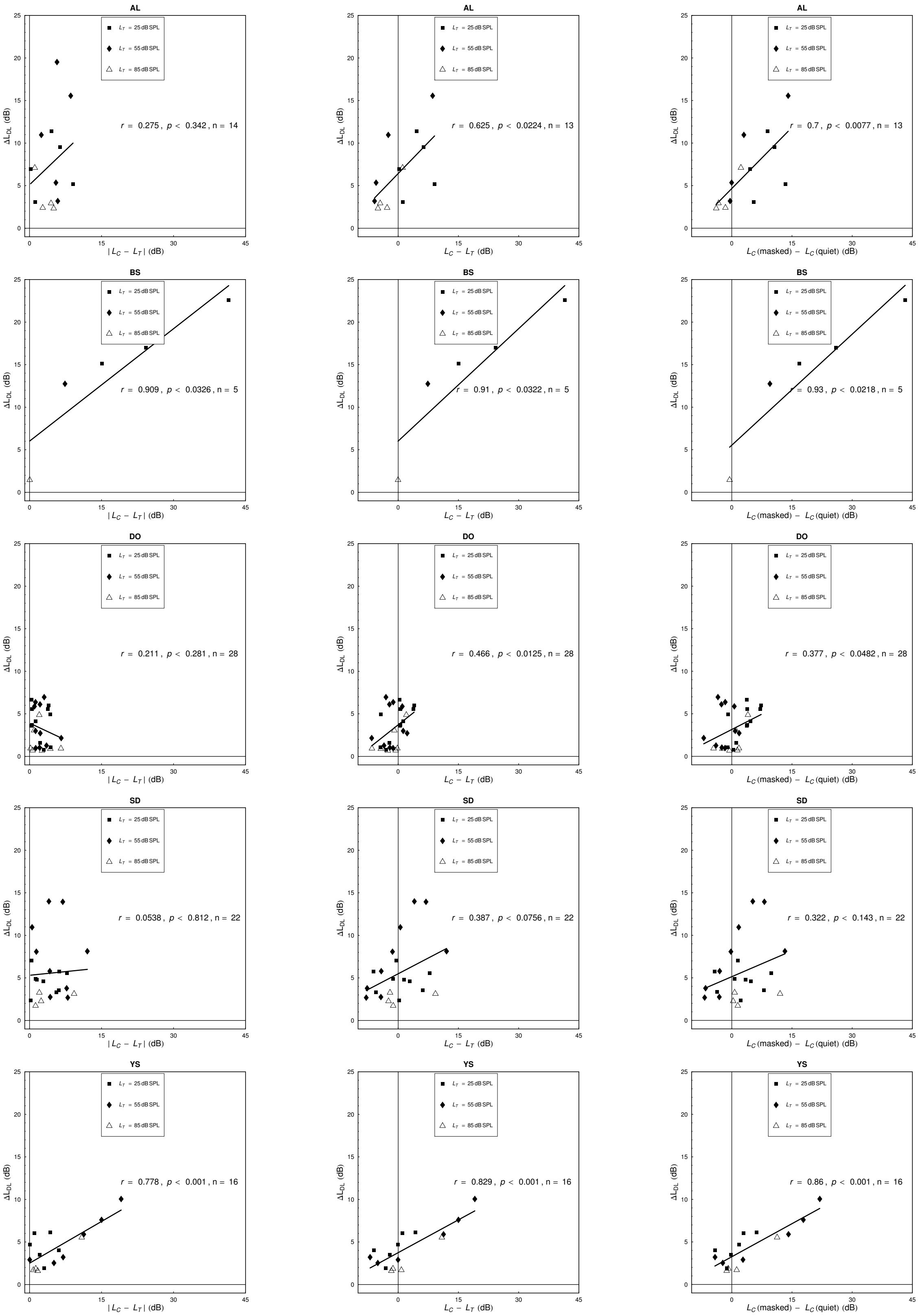

Fig. 37: Individual correlations. Left column: ΔL_{DL} re. absolute amount of loudness change $|L_C - L_T|$. Middle column: ΔL_{DL} re. loudness enhancement $L_C - L_T$. Right column: ΔL_{DL} re. loudness change computed as L_C(forward masked) − L_C(in quiet). Rows represent listeners. Symbols denote target levels. Regression lines are also displayed.

Zeng (1994) and Plack (1996 a) presented conditions with an intense masker only. For the data obtained with the 85-dB SPL masker in Experiments 1 and 2 of the present study, the correlation between ΔL_{DL} and $|L_C - L_T|$ was $r = 0.768$, $N = 21$, $p < 0.002$, which is indeed larger than the correlation computed for the complete data set. Due to the small number of data points per listener, the analysis of individual correlations does not seem informative.

It is important to note, that Plack (1996 a) did not analyze difference limens from a 2I, 2AFC intensity discrimination experiment, but loudness variability jnd_{Match} as obtained in Experiments 2 and 3 of this study. For the present data, the correlation between $L_C - L_T$ and jnd_{Match} was practically zero, $r = 0.083$. This points once again to the pronounced difference between the measures ΔL_{DL} and jnd_{Match}. It remains unclear, why the correlation between jnd_{Match} and loudness enhancement reported by Zeng and Plack was not present in the data from Experiment 2 and 3.

6.6.1 Discussion

The significant correlation between intensity discrimination difference limens and loudness enhancement is evidence supporting the similarity model and also confirms previous findings (Zeng, 1994; Plack, 1996 a).

On the other hand, several results of the correlation analysis are reason enough to question the straightforward relation between loudness enhancement and intensity resolution predicted by the loudness enhancement hypothesis (Carylon and Beveridge, 1993), that is part of the similarity model.

First, the prediction of loudness decrement also causing jnd elevations was not supported. Instead, loudness decrement corresponded to small values of ΔL_{DL}. To account for this finding in terms of the loudness enhancement hypothesis, it would be necessary to assume that a reduction in loudness due to a forward masker has no effect on loudness variability, while loudness enhancement results in an increase in variability. For loudness reduction caused by adaptation in the auditory nerve, arguments supporting this assumption could be found, as physiological studies reported a decrease in the variability of neural counts in the case of adaptation (Smith, 1977). Loudness decrement observed with a 100-ms ISI between masker and target can certainly not be attributed to peripheral mechanisms, however. For central processes such as the 'averaging' between masker and target loudness assumed by the similarity model, it is difficult to find formal arguments for reduced loudness variability associated with loudness decrement as well as for increased variability associated with loudness enhancement. This issue already discussed by Plack (1996 a) is clearly a weak point of the loudness enhancement hypothesis, and consequently also of the similarity model. In Chapter 7, an attempt will be made to explain the effects of loudness enhancement on intensity resolution in terms of a signal detection model.

Even if the independent variable used in the correlation analysis was loudness change going from positive to negative values instead of the absolute value of loudness change, three of the five individual correlations with ΔL_{DL} were rather weak and even not significant for one listener. In all fairness, only part of this problem can be attributed to the variability and instabilities associated with the measurement of difference limens and especially subjective qualities such as loudness. If one compares the individual patterns of ΔL_{DL} and loudness enhancement (Fig. 22, Fig. 27), several instances are obvious that speak against a simple relation between loudness enhancement and intensity resolution. To give an example, the loudness matches produced by listener AL at the 55-dB SPL target level show a pronounced mid-difference hump, while her jnd's increased monotonically with the masker-standard level difference in this condition.

The zero correlation between loudness enhancement and jnd_{Match} again raises the question of why the pattern observed for this measure differs so distinctly from the findings of Zeng (1994), Plack (1996 a), and the difference limens measured in Experiment 1 of this thesis. Insofar, it remains unclear whether the differences between the variability of loudness matches and ΔL_{DL} are additional evidence against the relation between loudness enhancement, loudness variability and intensity resolution, or if the problem can be attributed to a principal difference between the loudness matching data obtained in this thesis and in previous experiments.

To summarize, in the data from Experiments 1 – 3, loudness enhancement was indeed associated with an increase in ΔL_{DL}, although a straightforward linear relation between the two measures was found only for listeners BS and YS.

6.7 General discussion

If one considers loudness enhancement as predicted by the 'mergence' component of the similarity model, the data clearly support the hypothesis that perceptual similarity is an important factor. In Experiment 2, the mid-difference humps predicted by the model were observed at both the low and the intermediate target level, where the range of masker-target level differences was sufficiently large to allow a test of the hypothesis. The longer masker duration presented in Experiment 3 also resulted in slightly smaller amounts of loudness enhancement, especially at large masker-target level differences. This observation is compatible with the hypothesis that perceptual differences in other dimensions than loudness also decrease the weight assigned to masker loudness. The nonzero loudness matches in the case of masker level equal to target level could only partly be attributed to time order error, however. At this point, it is important to note that in a recent paper, which I became aware of after completion of the experiments, Scharf, Buus, and Nieder (2002) suggested that the loudness matches obtained in a three-tone matching task reflect "loudness recalibration" (that is, a reduction in loudness, cf. Marks, 1994) of the comparison tone rather than loudness enhancement of the target tone. These findings will be discussed in Chapter 7.1.9.

The recovery-rate model, the referential encoding hypothesis, and the similarity model predict distinctly different patterns for the effect of the masker-standard level difference at different standard levels on intensity jnd's. The mid-difference humps observed at the 25-dB SPL and at the 55-dB SPL standard level are clearly incompatible with both the recovery-rate model and the referential encoding hypothesis, but support the similarity model. For two of the six listeners in Experiment 1, ΔL_{DL} monotonically increased with the masker-standard level difference at all standard levels. The idea that listener BS experienced great difficulty in separating the sensations elicited by masker and standard, respectively, was supported by the observation of strongly reduced difference limens if a masker of longer duration or a two-tone masker was used. It remains unclear whether the absence of a mid-difference hump for four listeners at the 55-dB SPL standard level can be attributed to the maximum masker-standard level difference of 45 dB being too small for the jnd to decrease again.

The predicted causal relation between loudness enhancement and the difference limen, i.e., the prediction of ΔL_{DL} being a monotonically increasing function of loudness enhancement, was only partly supported. As expected, there was a significant correlation between loudness enhancement and

ΔL_{DL}. On the other hand, the individual data can be described as a linear relation between the two measures only for two of the five listeners. Loudness decrement was not associated with a jnd elevation, as the loudness enhancement hypothesis in its simplest form would suggest. The results emphasize the need for a formal model of the effects of non-simultaneous masking on both loudness and intensity resolution. Such a model will be discussed in Chapter 7.2.

The distinctly different patterns observed for the measure of loudness variability derived from the loudness matching data (jnd_{Match}) and ΔL_{DL} obtained in the intensity discrimination paradigm were completely unexpected. The reason for this finding remains unclear.

7 Quantitative formulation of the similarity model

As demonstrated in the preceding chapter, the similarity model can account for a large proportion of the effects of masker and target level in intensity discrimination and loudness matching tasks.

In the present form, the model can make only qualitative predictions. It would certainly be desirable to have a quantitative model that can be used to derive exact predictions for various experimental conditions.

In the following, a quantitative version of the loudness part of the similarity model is formulated first. It is based on well-known properties of the auditory system (the loudness function for pure tones) but also includes higher-level effects (target-distractor similarity).

In the second part of the chapter, a signal-detection approach is used to discuss possible explanations of the effect of mergence on intensity resolution.

7.1 Loudness

Loudness comparisons between non-simultaneously presented sounds necessarily involve the use of a memory representation for the sound presented first. Durlach and Braida (1969) as well as Massaro (1975) distinguished a sensory representation (auditory trace or preperceptual auditory store) available for several hundred milliseconds only, and a categorical representation. Consider a matching task where masker and target are presented in the first observation interval and a comparison tone in the second interval (Fig. 5). The listener decides whether the target or the comparison tone was louder. It can be assumed that masker-target interaction at peripheral stages does not play a significant role because the masker-target ISI is about 100 ms (Kiang et al., 1965). There are two central mechanisms, however, which could explain the effects of the masker:

1. Mergence (Elmasian, Galambos, and Bernheim, 1980), and
2. Memory confusion (Kallman and Massaro, 1979; Crowder, 1976).

According to the mergence hypothesis, the sensations (loudness values) elicited by masker and target, respectively, are automatically 'merged' into a combined percept. As a result, "*[...] the final percept of the target is approximated by a weighted average of the separate sensations each interactor would produce if presented alone*." (Elmasian, Galambos, and Bernheim, 1980, p. 606). Consequently, comparison tone loudness is compared to this weighted average rather than to 'true' target loudness. If the masker is louder than the target, this will result in the listener being biased towards responding "First tone louder", i.e., comparison tone level has to exceed target level for the loudness of target and comparison tone to match (loudness enhancement).

An alternative mechanism arises in the context of Massaro's (1975) stage-model of auditory information processing (1975) that distinguishes between an initial sensory representation (the "preperceptual auditory store", PAS) followed by a partially categorical representation (synthesized auditory memory, SAM). According to the model, if the ISI between target and comparison is sufficiently large (>500 ms), the initial sensory representation will no longer be available at presentation of the comparison tone, so that the listener has to retrieve a partially categorical representation of target loudness from synthesized auditory memory (SAM) in order to compare it with the loudness of the comparison tone. Now, Kallman and Massaro (1979) assume "memory confusion" to play an important role in SAM. According to this concept, if a mask is presented in the first interval, there is a probability that the listener erroneously retrieves the loudness information associated with the masker instead of the information associated with the target. Put differently, the listener in some proportion of the trials assesses masker loudness instead of target loudness when deciding whether the first or the second tone was louder. Again, if the masker is louder than the target, this will result in the matching procedure converging on a comparison level exceeding target level (loudness enhancement).

What can be assumed about the effects of perceptual similarity? Kallman and Massaro (1979) explicitly assumed masker-target similarity to have an influence on the probability of a memory confusion. Data from a frequency discrimination task supported their hypothesis. Systematic biases were observed for small masker-target frequency differences, but not for large differences (see Chapter 4.3.4).

The mergence hypothesis as formulated by Elmasian, Galambos, and Bernheim (1980) makes no assumptions about the effects of masker-target similarity. In the similarity model, it is assumed that the effect of masker loudness on the weighted average is smaller if masker and target are **perceptually different**. Depending on the spectro-temporal parameters of masker and target, the masker-target ISI, and the presence or absence of additional cues (e.g., a visual cue marking the target tone), the weight assigned to masker loudness will vary.

In the following, discussion will be restricted to the mergence concept.

The model can be summarized as follows:

Masker and target are processed by the auditory system. The results of the transformation are a memory representation N_M for loudness of the masker and a memory representation N_T for loudness of the target. During the silent interval between target tone and comparison tone, the representations of masker and target loudness interact (mergence). Therefore, the initial loudness value of the target is no longer available at presentation of the comparison tone. Instead, the listener will compare a weighted average of masker loudness N_M and target loudness N_T with comparison tone loudness N_C. The weight assigned to masker loudness in computing the weighted average is a function of the perceptual distance between masker and target.

In the quantitative model, the loudness change is predicted by conducting four processing steps:

1. Compute the initial loudness representations of masker and target (N_M and N_T, respectively).
2. Determine masker weight p_M. At this point, similarity in loudness and on other dimensions is effective.
3. Compute the weighted average N_{Merged} between masker and target loudness.
4. Find the comparison tone level corresponding to a loudness value N_C equal to the weighted average N_{Merged} computed in Step 3.

7.1.1 Step 1: Loudness of masker and target

To determine the initial loudness values, the loudness function for pure tones as proposed by Zwislocki (1965) is used.

According to his model, which successfully fits a wide range of magnitude estimation data (cf. Hellman, 1991), loudness of a tone in quiet depends on sound pressure P of the tone and on sound pressure P_{Th} at detection threshold as

$$\mathrm{N} = K\,[(P^2 + 2.5\,P_{Th}^{\,2})^{\theta} - (2.5\,P_{Th}^{\,2})^{\theta}], \qquad (7\text{-}1)$$

where K is a scale constant. According to the review of Hellman (1991), the mean exponent θ computed on the basis of 78 ratio scaling experiments is 0.3 (SD = 0.045, Min. = 0.185, Max. = 0.425). This value agrees with ISO/*R* 131-1959. A change in threshold (e.g., induced by simultaneous masking or hearing loss) alters the function at low levels, while leaving it relatively unaffected at high levels (loudness recruitment; cf. Scharf, 1978). Tones differing in, e.g., duration or frequency will produce different loudness values even at higher sound pressure levels (e.g., Hellman, 1976; Zwislocki, 1969; Florentine, Buus, and Poulsen, 1996). This effect can be modeled by choosing different values for the scale parameter K.

The model rests on the assumptions that total loudness within one critical band (CB) depends on the total power present in that band and that loudness is a power function of sound pressure. A further assumption made by Zwislocki is the presence of an intrinsic physiological noise that determines the threshold of hearing in the given critical band. He emphasized that the auditory system is capable of selectively listening to the tone alone, so that in the model, loudness of the intrinsic noise is subtracted from total loudness. The effect of the intrinsic noise is represented by the 2.5 $P_{Th}^{\,2}$ terms in Eq. 7-1.

An alternative loudness model was proposed by Zwicker (1958, 1963; cf. Scharf, 1978; Zwicker and Fastl, 1999). His model of the "specific loudness" of a tone uses excitation patterns to calculate loudness of pure tones as well as of spectrally more complex stimuli. For pure tones, the predicted loudness function is very similar to the Zwislocki function (cf. Hellman and Meiselman, 1990; Zwicker and Fastl, 1999, p. 224). If complex stimuli are presented, the Zwicker model must be used.

According to the Zwislocki model, the sole effect of the forward masker on the initial loudness value of the target can be an elevated detection threshold, which will result in an altered loudness function according to Eq. 7-1. For this reason, the loudness values for masker and target presented in interval 1 can be computed as if the two tones were presented in isolation, except for a potential threshold elevation caused by either tone. Using Eq. 7-1, masker loudness is written as

$$\mathrm{N}_M = K_M\,[(P_M^{\,2} + 2.5\,P_{ThM}^{\,2})^{\theta} - (2.5\,P_{ThM}^{\,2})^{\theta}], \qquad (7\text{-}2)$$

where K_M is a scale parameter, P_M is masker pressure and P_{ThM} is pressure at detection threshold. In the same way, target loudness is modeled as

$$\mathrm{N}_T = K_T\,[(P_T^{\,2} + 2.5\,P_{ThT}^{\,2})^{\theta} - (2.5\,P_{ThT}^{\,2})^{\theta}], \qquad (7\text{-}3)$$

where P_T is target pressure and P_{ThT} pressure at detection threshold.

Note that sound pressure level measured in dB SPL is related to sound pressure as

$$L = 10 \log_{10}\left(\frac{P}{P_0}\right), \qquad (7\text{-}4)$$

where P_0 is reference sound pressure (20 µPa, cf. Appendix II.3).

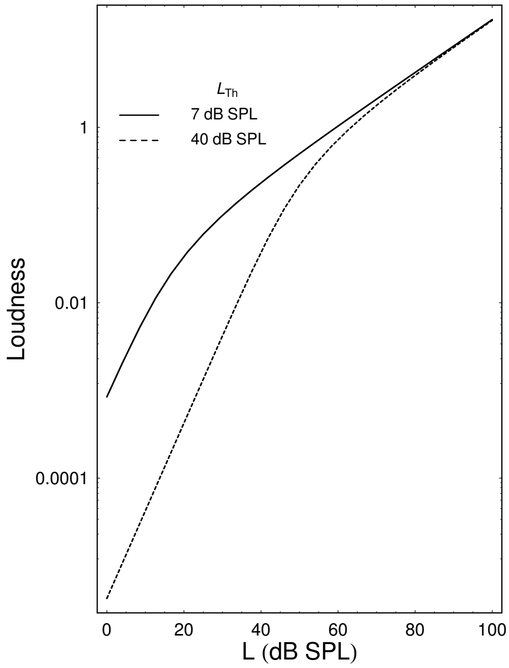

Fig. 38: Loudness in assigned numbers according to Eq. 7-1 as a function of the sound pressure level L of a pure tone. Solid line: detection threshold level L_{Th} = 7 dB SPL. Dashed line: L_{Th} = 40 dB SPL. Parameters: K = 11.2, θ = 0.3.

If the masker elevates target threshold, this will have an effect on the loudness values computed for the target. Equivalently, masker loudness will be altered if the target is intense enough to act as a backward masker. It has been demonstrated that backward masking is less pronounced than forward masking in a detection task (e.g., Oxenham and Moore, 1995). Moreover, even the effects of forward masking were moderate or small for most listeners in Experiment 1-3 (Table 1), and in other experiments using comparable stimuli (e.g., Zeng, Turner, and Relkin, 1991; Carlyon and Beveridge, 1993). Consequently, it is assumed that the effects of backward masking are negligible. Fig. 38 shows two loudness functions computed according to Eq. 7-1.

7.1.2 Step 2: Masker weight

Given the masker and target loudness values computed in the first step, it is now necessary to model the mergence process.

The important question is What determines the amount of mergence? According to the similarity hypothesis, less mergence will occur if the representations of masker and target are 'separated' in one or more dimensions. In fact, it has been demonstrated that increasing the difference in frequency (Zwislocki and Sokolich, 1974) or presenting the masker contralaterally (Elmasian and Galambos, 1975) reduces the amount of loudness enhancement. The latter two effects can be understood in terms of peripheral mechanisms (e.g., masker and target differing in frequency are processed in different auditory channels). For the similarity model, however, it is not important whether the interaction between masker and target is reduced because each tone is processed in different peripheral structures, or if similarity has an effect on processing at a more central location, as it must be the case for similarity in loudness or duration.

Temporal separation between masker and target in interval 1 should also increase the 'perceptual distance' between the memory representations. In fact, a long masker-target ISI reduces the effect of the masker (Zwislocki and Sokolich, 1974; Zeng and Turner, 1992).

For the concept of memory confusion, the effects of decreasing perceptual similarity can be understood simply by assuming that, e.g., the loudness representation of a contralaterally presented masker has information about the ear it was presented to tagged to it, so that the probability of a confusion with the loudness representation of the target presented to the opposite ear is reduced. For the mergence concept, it is less intuitive to explain why automatic weighted averaging of masker and target loudness should be more pronounced in this case. An admittedly highly speculative explanation arises if one follows the equally speculative suggestion by Elmasian, Galambos, and Bernheim (1980) that "*overlapping in the brain*" was responsible for the effect. According to this idea, if the loci of the neuronal representations are spatially proximal, a stronger amount of interaction/mergence can be expected. For similarity in frequency, such a 'neuronal distance effect' would certainly make sense, as different frequencies are processed in different peripheral structures. Additionally, the human primary auditory area (AI) was reported to be tonotopically organized (e.g., Romani et al., 1982; Liegeois-Chauvel et al., 1995; Pantev et al., 1995).

Similarly, for contralaterally presented sounds, different parts of the auditory periphery are involved.

As to similarity in loudness, one could argue that there are also reports of an *ampliotopic* organization in auditory cortex (Brugge and Merzenich, 1973; Suga, 1977; Pantev et al., 1991; Houtilainen et al., 1995). Therefore, loudness similarity might also be correlated with 'neuronal spatial proximity'.

Other similarity effects are difficult to explain by "neural proximity", however. To give an example, to account for the effect of a masker differing in duration from the standard, one would have to assume a "duratiotopic" structure involved in the processing.

Turning to the quantitative model again, the effect of loudness similarity is modeled to be independent of similarity in other dimensions. The weight assigned to masker loudness is written as

$$p_M = p_{Max} f(N_M, N_T), \qquad (7\text{-}5)$$

where the function $f(N_M, N_T)$ represents the effects of perceptual similarity in the loudness dimension. The effect of similarity in the remaining dimensions is represented by p_{Max}, which is the maximum amount of mergence that will be effective if masker loudness equals target loudness. For example, if masker and target are presented contralaterally, p_M will be maximal for $N_M = N_T$. Still, masker weight in this condition can be modeled to be smaller than for the case of ipsilateral presentation by choosing a smaller value of p_{Max}. The value of p_{Max} depends on the listener, the stimuli and the temporal configuration used.

The function $f(N_M, N_T)$ is chosen in such a way that

$$0 \le f(N_M, N_T) \le 1 \text{ and } f(N, N) = 1. \qquad (7\text{-}6)$$

Above that, it is required that $f(N_M, N_T)$ decreases monotonically with the absolute value of the difference between N_M and N_T, approaching 0 at large differences. It also seems reasonable to assume a variant of Weber's law, i.e., masker and target loudness differing by, e.g., 10% should result in the same value of p_M independent of target loudness. A simple means to achieve this behaviour is to use a logarithmic transformation of the loudness values. Given that the requirements outlined above are met, the choice of $f(N_T, N_M)$ is certainly arbitrary. The specific function introduced here can be deducted from a 'memory confusion' model (cf. Appendix I):

$$\mathrm{f}(N_T, N_M) = \left(1 - \mathrm{Erf}\left(\frac{|\log_{10} N_M - \log_{10} N_T|}{\sqrt{2}\sigma_{Sim}}\right)\right), \qquad (7\text{-}7)$$

where $\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_{z=0}^{x} e^{-z^2} dz$. For the sake of simplicity, σ_{Sim} is assumed to be independent of loudness. Therefore, as $\log_{10} N_M - \log_{10} N_T$ is equal to $\log_{10}(N_M/N_T)$, the value of $f(N_T, N_M)$ depends only on the ratio between masker and target loudness. The 'similarity parameter' σ_{Sim} determines how fast p_M decreases with the difference between $\log_{10} N_M$ and $\log_{10} N_T$. The effect of similarity in the loudness dimension is maximal at small values of σ_{Sim}. Technically speaking, the function $1 - \mathrm{Erf}(x)$ in Eq. 7-7 maps the real numbers between 0 and infinity onto the interval [0, 1].

According to $f(N_T, N_M)$, only a small weight will be assigned to masker loudness if masker and target loudness differ strongly. If they are identical, the maximum amount of mergence (p_{Max}) will be effective. Substituting from Eq. 7-7 into Eq. 7-5 yields

$$p_M = p_{Max}\left(1 - \mathrm{Erf}\left(\frac{|\log_{10} N_M - \log_{10} N_T|}{\sqrt{2}\sigma_{Sim}}\right)\right). \qquad (7\text{-}8)$$

It is required that the weights p_M and p_T assigned to masker and target loudness, respectively, sum to unity,

$$p_T = 1 - p_M. \qquad (7\text{-}9)$$

There is no a-priori reason why masker loudness should not receive more weight than target loudness ($p_{Max} > 0.5$); even though the memory confusion approach suggests that the maximum weight assigned to masker loudness should be 0.5, which corresponds to random guessing.

Predicted masker weight is plotted in Fig. 39 (left panel) as a function of the relative difference between masker and target loudness. If masker loudness is zero, masker weight p_M is also zero. Obviously, such behavior is reasonable because this condition is equivalent to the situation *without* masker. We note that according to the model, an intense masker with loudness, e.g., 50% above target loudness ($\frac{N_M - N_T}{N_T} = 50\%$) has a stronger effect on the weighted loudness average than a soft masker with loudness 50% below target loudness ($\frac{N_M - N_T}{N_T} = -50\%$).

In the right panel of Fig. 39, p_M is plotted as a function of the masker-target level difference $\Delta L_{M\text{-}T} = L_M - L_T$ at low, intermediate and high target levels. It is obvious that the model predicts the effect of similarity in loudness to be most pronounced at the smallest target level, as p_M decreases faster with the absolute value of $\Delta L_{M\text{-}T}$ at low than at intermediate and high levels.

The reason for this behavior lies in the form of the loudness function, which is steeper at low levels (Fig. 38) and results in the loudness change caused by, e.g., a 5 dB increase in masker level to be larger than in the flatter region of the loudness function. The slope variation has been attributed to basilar-membrane compression being effective at intermediate levels, but not at levels near threshold (e.g., Yates, 1990; Schlauch, DiGiovanni, and Ries, 1998).

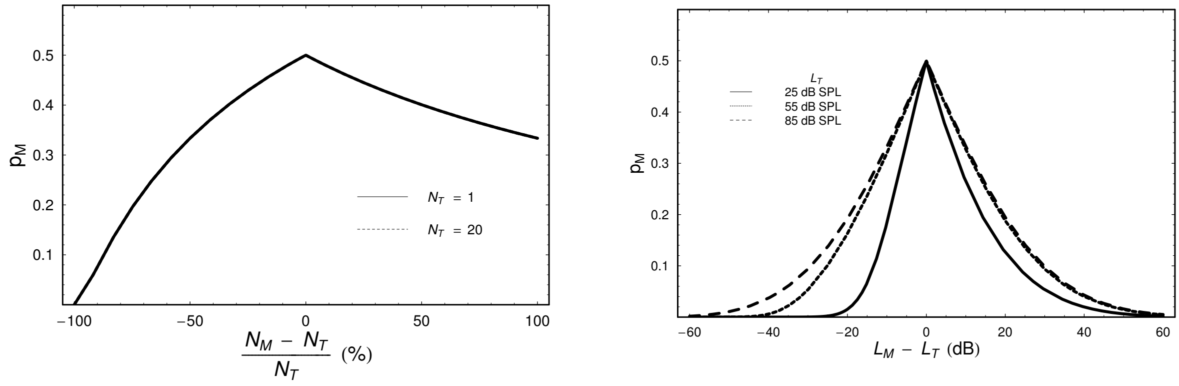

Fig. 39: Left panel: Masker weight p_M plotted as a function of the relative difference (expressed in percent) between masker and target loudness. The two curves lying on top of each other are for $N_T = 1$ and $N_T = 20$. Parameters: $p_{Max} = 0.5$, $\sigma_{Sim} = 1$.
Right panel: Masker weight p_M as a function of $L_M - L_T$ at low (solid line), intermediate (short-dashed line), and high target levels (long-dashed line). Parameters: $p_{Max} = 0.5$, $\sigma_{Sim} = 0.7$, $\theta = 0.3$, $L_{ThM} = L_{ThT} = 10$ dB SPL, $K_T = K_M$.

Some direct measurements of BM vibration also indicate a steeper amplitude growth at high levels (e.g., Ruggero, Rich, and Recio, 1996), but this finding is less well established. Note that the slope of the Zwislocki loudness function does not increase at high levels, as such a pattern is not found in most psychophysically measured loudness functions. Exceptions are magnitude estimation data by Viemeister and Bacon (1988), and loudness functions derived from loudness matching data by Florentine, Buus, and colleagues (Buus, Florentine, and Poulsen, 1997; Buus and Florentine, 2001; Florentine, Epstein, and Buus, 2001). The loudness model proposed by Moore, Glasberg, and Baer (1997) also predicts a decrease in slope at high levels. Certainly, a different loudness function could be incorporated into the model, but the Zwislocki (1965) function provides a good description of most psychophysical data collected during the last decades.

7.1.3 Step 3: Mergence

The weighted average between target and masker loudness is modeled to be

$$N_{Merged} = p_M N_M + (1 - p_M) N_T. \qquad (7\text{-}10)$$

It seems reasonable to require that merged loudness always lies between masker and target loudness,

$$\mathrm{Min}(N_M, N_T) \le N_{Merged} \le \mathrm{Max}(N_M, N_T). \qquad (7\text{-}11)$$

In other words, merged loudness should not be smaller than loudness of the softer sound and should also not exceed loudness of the louder sound. This constraint is met if $0 \le p_{Max} \le 1$.

7.1.4 Step 4: Loudness match

The sound pressure P_C of the comparison tone eliciting a loudness sensation N_C equal to the weighted average N_{Merged} can be found by solving the equation

$$N_{Merged} = K_C\,[(P_C^{\,2} + 2.5\,P_{ThC}^{\,2})^{\theta} - (2.5\,P_{ThC}^{\,2})^{\theta}] \qquad (7\text{-}12)$$

for P_C. In Eq. 7-12, K_C is a scale parameter, P_C is masker pressure and P_{ThC} is the pressure at detection threshold.

Eq. 7-12 can be solved for P_C analytically. The parameters P_T, P_M, P_{ThT}, P_{ThM}, P_{ThC}, K_T, K_M, K_C and θ (which is assumed to be 0.3) are known a priori. Only the parameters p_{Max} and σ_{Sim} have to be estimated when fitting the model. Note that if $K_T = K_M = K_C$, the scale factors cancel out.

In the following, we use sound pressure level rather than pressure (Eq. 7-4).

7.1.5 Model predictions

Fig. 40 displays predicted loudness enhancement ($L_C - L_T$) as a function of the masker-target level difference for three different target levels. In the simulation, detection thresholds were 10 dB SPL, and the scale parameters of the loudness functions for masker, target and comparison tone were identical, so that they canceled out. The parameter values $p_{Max} = 0.5$ and $\sigma_{Sim} = 0.7$ were used.

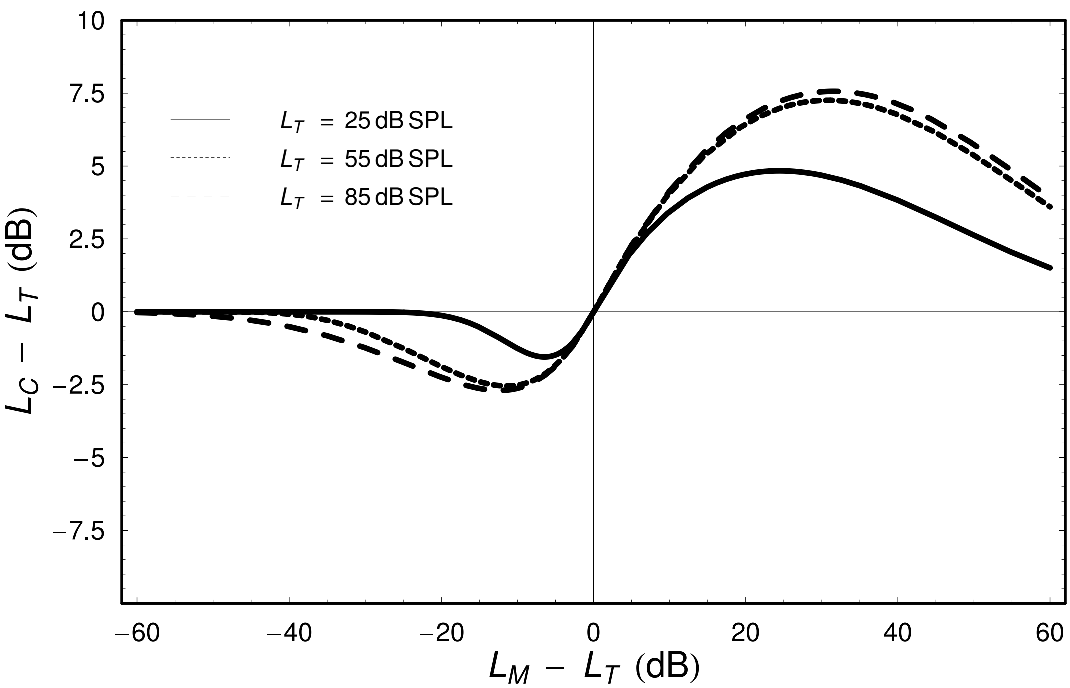

Fig. 40: Predicted loudness enhancement ($L_C - L_T$) as a function of the masker-target level difference. Solid line: L_T = 25 dB SPL. Short-dashed line: L_T = 55 dB SPL. Long-dashed line: L_T = 85 dB SPL. Parameters: $p_{Max} = 0.5$, $\sigma_{Sim} = 0.7$, $\theta = 0.3$, $L_{ThM} = L_{ThT} = L_{ThC}$ = 10 dB SPL, $K_T = K_M = K_C$.

Predicted loudness enhancement is a non-monotonic function of the masker-target level difference at each target level; both enhancement and decrement are maximal at intermediate masker-target level differences. Note that presenting a masker-target level difference of more than 20 dB is precluded for the 85-dB SPL target. Additionally, for extremely intense maskers (> 110 dB SPL), adaptation in the auditory periphery can be expected to cause additional effects, which are ignored in Fig. 40.

Due to the mechanisms discussed in Chapter 7.1.2, the effect of the masker increases with target level and is smallest at the lowest target level.

What are the effects of changes in p_{Max} and σ_{Sim}? As seen in Fig. 41, a smaller value of p_{Max} results in reduced loudness change while the inflection points in the loudness enhancement function (the location of the mid-difference humps for enhancement and decrement) remain at the same position (solid line versus short-dashed line). A smaller value of σ_{Sim} causes the position of the humps to be closer to $\Delta L_{M\text{-}T}$ = 0 dB (solid line versus long-dashed line). As in this situation, pronounced mergence occurs for small masker-target level differences only, the maximal effect of the masker is also smaller.

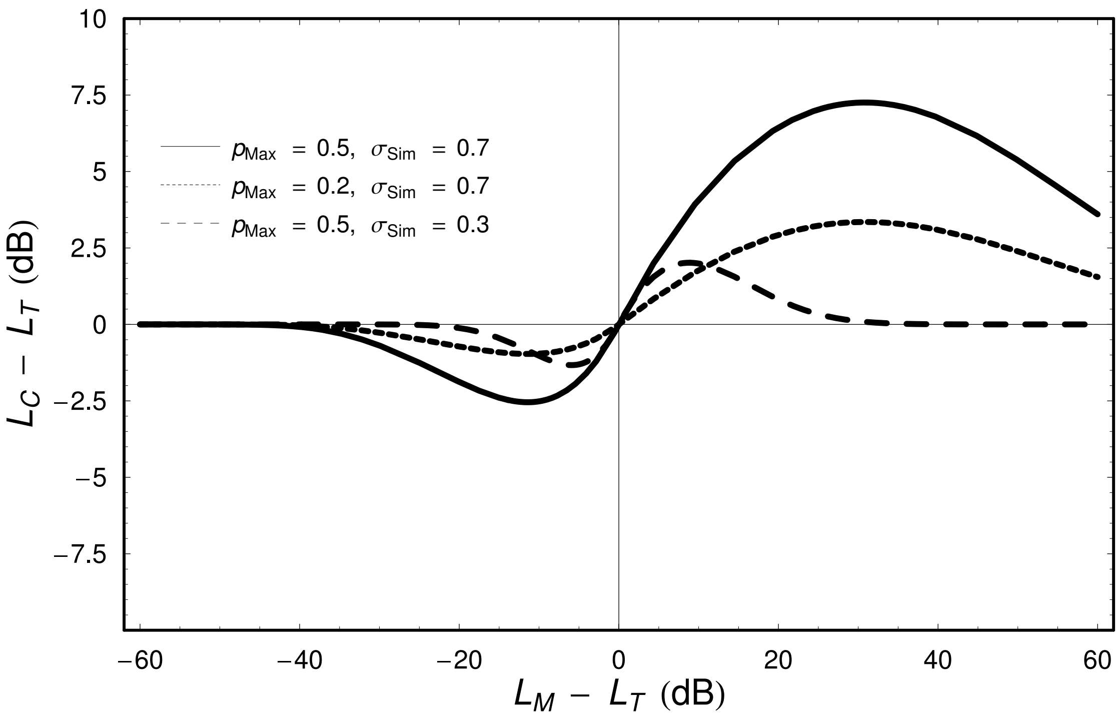

Fig. 41: Predicted loudness enhancement ($L_C - L_T$) as a function of the masker-target level difference for a 55-dB SPL target. Solid line: p_{Max} = 0.5, σ_{Sim} = 0.7. Short-dashed line: p_{Max} = 0.2, σ_{Sim} = 0.7. Long-dashed line: p_{Max} = 0.5, σ_{Sim} = 0.3. Remaining parameters: θ = 0.3, $L_{ThM} = L_{ThT} = L_{ThC}$ = 10 dB SPL, $K_T = K_M = K_C$.

It is somewhat difficult to predict whether increasing the perceptual distance between masker and target by, e.g., increasing the ISI, should have an effect on the maximum value of masker weight (p_{Max}) only, or if the similarity parameter σ_{Sim} should also become smaller so that the position of the mid-difference hump is shifted. An argument for the former prediction would be that even if masker weight is reduced in the $N_M = N_T$ condition due to, e.g., an increased temporal separation between masker and target, presenting a masker 10% louder than the target should cause the same relative change in masker weight as for a short ISI. Put differently, the effect of loudness similarity should be independent of p_{Max}.

7.1.6 Comparing model predictions and data

The present version is the simplest form of the model. We are now in a position to qualitatively compare the model predictions to the data obtained in previous studies and in Experiment 2 of the present study.

Following this evaluation, a necessary modification of the model will be introduced.

To summarize Chapter 3 and the results from Experiment 2, the most important findings concerning loudness enhancement are:

Observation 1. If masker level is greater than target level, loudness of the target tone is increased: the level of the comparison tone adjusted by the listener to match the loudness of the target tone is greater than target tone level (loudness enhancement: $L_C - L_T > 0$; Fig. 6; Elmasian and Galambos, 1975; Elmasian, Galambos, and Bernheim, 1980; Zwislocki and Sokolich, 1974).

Observation 2. If masker level is smaller than target level, loudness of the target tone is decreased (loudness decrement: $L_C - L_T < 0$, Fig. 6).

For positive level differences $\Delta L_{M\text{-}T}$ (masker more intense than target), the model correctly predicts loudness enhancement, while a reduction of loudness follows for negative level differences.

Observation 3. Loudness decrement is less pronounced than loudness enhancement (Experiment 2; Elmasian and Galambos, 1975; Elmasian, Galambos, and Bernheim, 1980).

Compatible with the data, enhancement is predicted to be more pronounced than decrement. According to Eq. 7-10, the change in target loudness caused by the masker is

$$\Delta N_T = N_{Merged} - N_T = (N_M - N_T) \cdot p_M. \quad (7\text{-}13)$$

The slope of the function relating loudness (plotted on a linear scale) and sound pressure level increases with level. Therefore, the value $|N_M - N_T|$ is smaller for a negative masker-target level difference (e.g., $\Delta L_{M\text{-}T} = -15$ dB) than for a positive level difference of the same absolute value ($\Delta L_{M\text{-}T} = +15$ dB).

Observation 4. Loudness decrement vanishes as masker level approaches detection threshold (Elmasian and Galambos, 1975; Elmasian, Galambos, and Bernheim, 1980).

As masker loudness N_M approaches zero, p_M also approaches zero (Eq. 7-7). I.e., the loudness match is predicted to be the same as in quiet for masker levels below threshold.

Observation 5. For a fixed target level, loudness enhancement is a non-monotonic function of the masker-target level difference $\Delta L_{M\text{-}T} = L_M - L_T$ (Experiment 2).

This pattern is correctly predicted for masker levels exceeding target level as well as for masker levels smaller than target level.

Observation 6. Loudness enhancement is more pronounced at intermediate than at low target levels (Experiment 2).

The level dependence of p_M discussed in Chapter 7.1.2 results in loudness enhancement being less pronounced for the 25-dB SPL than for the 55-dB SPL target. Out of concern for the listeners, masker-target level differences exceeding +15 dB have not been presented combined with an intense target (e.g., 85 dB SPL). Therefore, it remains unclear if the model is correct in predicting enhancement to be slightly more pronounced at the highest target level than at intermediate levels. In the data from Experiment 2 of this study, which are the only results to date that allow comparing the effects of masker level at different target levels, decrement was less pronounced at L_T = 85 dB SPL than at L_T = 55 dB SPL. The model, on the other hand, predicts decrement to be slightly more pronounced at the highest target level.

Observation 7. For a fixed intense masker, loudness enhancement is maximal at intermediate levels of the to-be-matched tones (Fig. 42, left panel; Zeng, 1994; Plack, 1996).

This "midlevel hump" is also predicted by the similarity model (Fig. 42, right panel).

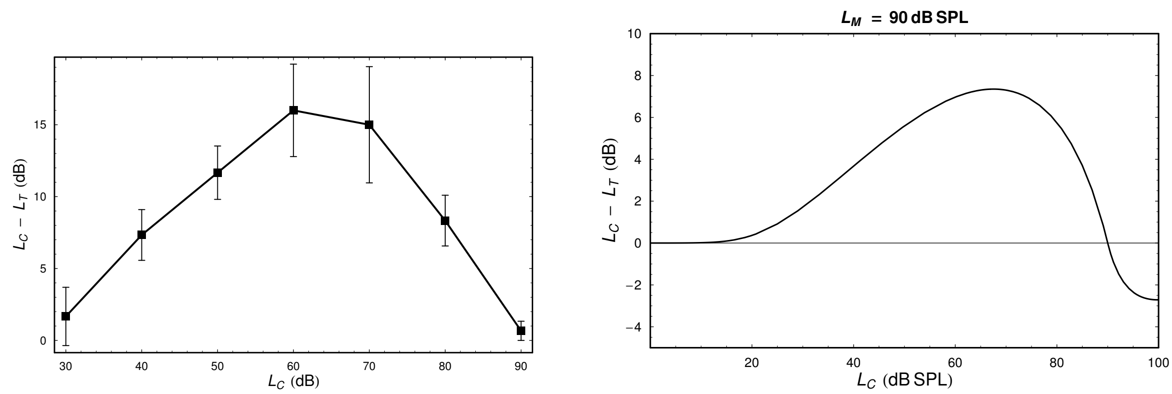

Fig. 42: Left panel: Mean loudness enhancement ($L_C - L_T$) reported by Zeng (1994) as a function of comparison tone level. Masker level: 90 dB SPL. Three listeners. Error bars show ± 1 SEM.
Right panel: Predicted loudness enhancement ($L_C - L_T$) as a function of comparison tone level for a 90-dB SPL masker. Parameters: p_{Max} = 0.5, σ_{Sim} = 0.7, θ = 0.3, $L_{ThM} = L_{ThT} = L_{ThC}$ = 10 dB SPL, $K_T = K_M = K_C$.

Before fitting the model to psychophysical data, one problem remains to be discussed.

7.1.7 Modification of Step 4: Accounting for the loudness match in quiet

Listeners frequently produce a nonzero difference between target level L_T and comparison level L_C even if no masker is present, as experiments studying the time order error have demonstrated (see Hellström, 1985, for a review). Equally important, loudness matches ($L_C - L_T$) are not necessarily zero if masker and target are **identical** (same level, duration, spectrum etc.), so that their loudness is also identical ($N_M = N_T$; e.g., Elmasian and Galambos, 1975).

A simple way to account for the bias observed in quiet would be to correct the value of L_C according to Eq. 7-12 by subtracting the value of ($L_C - L_T$) obtained in quiet. Alternatively, if no loudness match in quiet is available, L_C could be corrected so that the predicted comparison level for $N_M = N_T$ matches the observed value. A problem with this simple solution is that the correction takes place in physical domain (sound pressure level) and not in the domain of subjective sensation (loudness).

According to the trace drift hypothesis (Michels and Helson, 1954; Hellström, 1985), the bias in quiet (time order error) is due to N_T and N_M drifting towards adaptation level during the silent interval between target and comparison tone. In a first approximation, this is equivalent to a drift of the merged loudness N_{Merged}. In other words, comparison tone loudness is compared to the merged loudness modified by trace drift, which can be formalized by introducing a bias factor b to Eq. 7-12 so that

$$b\,N_{Merged} = K\,[(P_C^2 + 2.5\,P_{ThC}^2)^\theta - (2.5\,P_{ThC}^2)^\theta]. \quad (7\text{-}14)$$

If $b = 1$, no trace drift is present. A value of $b < 1$ corresponds to a negative value of $L_C - L_T$ in quiet, i.e., loudness of the tones presented in interval 1 drifted towards lower values. Eq. 7-14 is a first approximation in the sense that trace drift of masker loudness during the silent interval between masker and target is ignored. As the masker-target interval is normally much smaller than the target-comparison tone interval, this simplification seems reasonable.

Selecting b to account for the loudness match obtained without masker allows correcting for an effect that is not due to forward masking. In this situation, b can be interpreted as a fixed parameter that is known a-priori. If b is selected to account for the empirical loudness match obtained at $N_M = N_T$, or if a value of b minimizing the residual sum of squares is estimated when fitting the model, this amounts to introducing an additional unknown parameter. As trace drift was found to depend on target level (Hellström, 1985), a separate value of b is used for each target level.

7.1.8 Fitting the model

Previous studies presented either a fixed-level target (e.g., 70 dB SPL) and varied masker level (e.g., Elmasian and Galambos, 1975; Elmasian, Galambos, and Bernheim, 1980), or varied comparison tone level and used a masker level of about 90 dB SPL (Zeng, 1994; Plack, 1996).

In the experiments by Elmasian, Galambos and colleagues, the difference between masker and target level ($\Delta L_{M\text{-}T} = L_M - L_T$) was not varied over a range large enough to allow conclusions about the absence or presence of similarity effect in loudness enhancement. Two of their experiments provide information about loudness decrement, however.

As $L_M - L_T$ was varied over a range of 60 dB in the ‘midlevel hump’ experiments conducted by Zeng (1994) and Plack (1996 a), their data can be used to evaluate the predictions of the similarity model concerning the effect of the loudness difference between masker and target.

In order to gain a more detailed insight into the effects of both masker and target level, loudness enhancement was measured as a function of the masker-target level difference at different target levels in Experiments 2 and 3. These data provide a direct test of the predictions displayed in Fig. 40.

In the following, fits of the model to a number of data sets are presented. For previous studies, fitting the model involved reading off the observed amounts of loudness enhancement from the figures presented in the papers. For this reason, the data points used for fitting can be exact to 1-2 dB only.

The model was fitted using functions internal to the computer algebra system Mathematica Version 5 (NonlinearRegress[], which employs a Levenberg-Marquardt algorithm, and NMinimize[])

The exponent θ of the loudness function was set to 0.3 for all data sets. In some of the studies, detection thresholds (L_{ThT}, L_{ThM}, and L_{ThC}) were reported. For the remaining cases, they were estimated on the basis of thresholds reported for similar stimuli (e.g., Zeng, Turner, and Relkin, 1991). The effect of tone duration on detection threshold was accounted for by Eq. 152 in Zwislocki (1965, p. 57), which gives the ratio between threshold intensity of a tone burst and threshold intensity of a continuous tone. The effect of tone duration on the scale parameter K of the loudness function was computed using the temporal loudness summation model by Zwislocki (1969, Eq. 15, p. 436). Without loss of generality, the scale parameter K_T effective for target loudness was set to unity.

7.1.8.1 Fixed target level

Fig. 43 displays fits to two data sets by Elmasian and Galambos (1975). They contrasted loudness enhancement for ipsilateral, diotic, and contralateral presentation of the masker and studied forward as well as backward masking. All stimuli were 20-ms, 5-kHz pure tones, the target was presented monaurally.

Detection threshold was estimated to be 28 dB, which is the arithmetic mean of three individual thresholds reported for the same stimuli by Elmasian and Galambos (1975) and Elmasian, Galambos, and Bernheim (1980). No loudness matches in quiet were reported in the paper. Therefore, the bias parameter b was so chosen that the model exactly predicted the loudness match observed at $L_M = L_T$. Goodness of fit was summarized by calculating the coefficient of determination

$$R^2 = 1 - \sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2 , \qquad (7\text{-}15)$$

where y_i denotes the value of $L_C - L_T$ observed at data point i, $\hat{y}_i$ denotes the value of $L_C - L_T$ predicted for data point i, and $\bar{y}$ is mean observed loudness enhancement (Kvålseth, 1985).

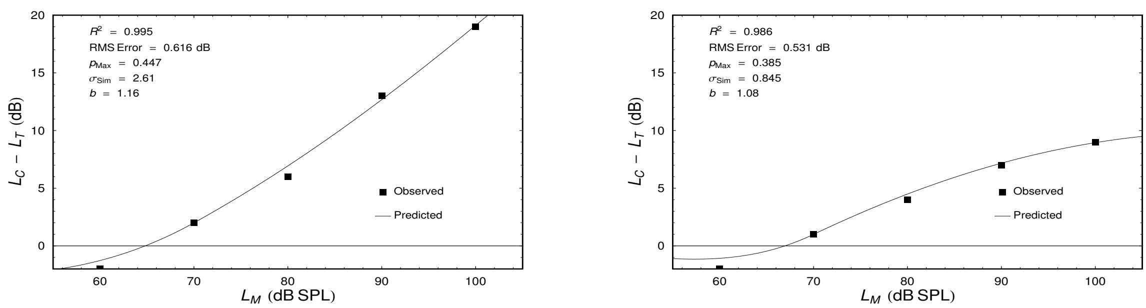

Fig. 43: Best fits to mean data (8 subjects) by Elmasian and Galambos (1975). Condition: forward masker. Left column: ipsilateral masker. Right column: contralateral masker. Target level was 70 dB SPL, masker level was varied in 10 dB steps, and listeners adjusted comparison tone level. Filled boxes: observed. Lines: predicted. The insets show R^2, RMS error, and the best fitting parameter values. Fixed parameters: $\theta = 0.3$, $L_{ThM} = L_{ThT} = L_{ThC} = 28$ dB SPL (estimated), $K_T = K_M = K_C$. The bias parameter b was selected so that predicted loudness enhancement matches the observed value at $L_M = L_T$ (70 dB SPL).

The model produced an excellent fit to the data obtained with the ipsilateral masker ($R^2 >$ 0.99), which is of course not very surprising as five data points were fitted using three free parameters. The best-fitting value of p_{Max} was 0.447, σ_{Sim} was estimated to be 2.61. The bias parameter b was estimated to be 1.16, reflecting the non-zero loudness match in the $L_M = L_T$ condition.

As discussed in Chapter 7.1.2, contralateral presentation of the masker should result in uniformly smaller masker weights than in the ipsilateral condition because localization provides a distinction between masker and target that is independent of loudness similarity. Contrary to expectation, the values of the estimated parameters indicate that the difference between the two conditions was due to a different value of σ_{Sim} (0.845) rather than of p_M (0.385; Fig. 43, right panel). Fitting the data obtained with the contralateral masker using the value of σ_{Sim} estimated for the ipsilateral condition only slightly increased RMS error from 0.531 to 1.24 dB (Fig. 44). Given the fact that the inter-individual variability in Elmasian and Galambos' data was much larger for the dichotic than for the ipsilateral condition, it seems that the estimates computed from these data do not allow drawing a definitive conclusion.

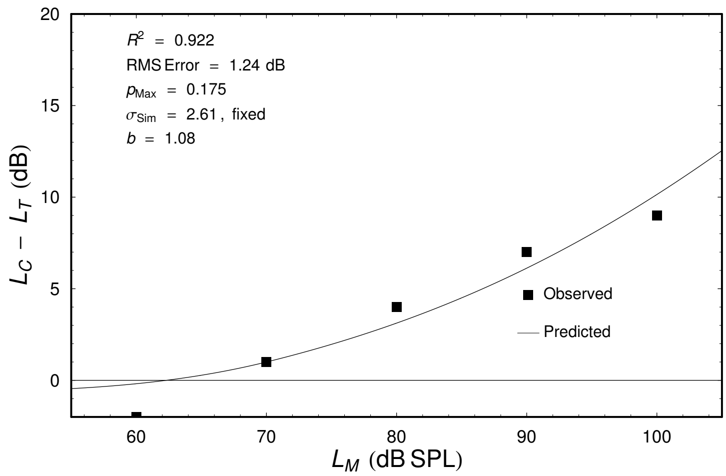

Fig. 44: Best fit to mean data obtained in the dichotic condition (Elmasian and Galambos, 1975). The same parameters as in Fig. 43 were used, except that σ_{Sim} was set to the value estimated in the ipsilateral condition (2.61).

In their Experiment 6, Elmasian and Galambos (1975, p. 232) presented data for one listener that received masker levels both below and above target level. Loudness decrement as well as loudness enhancement was observed.

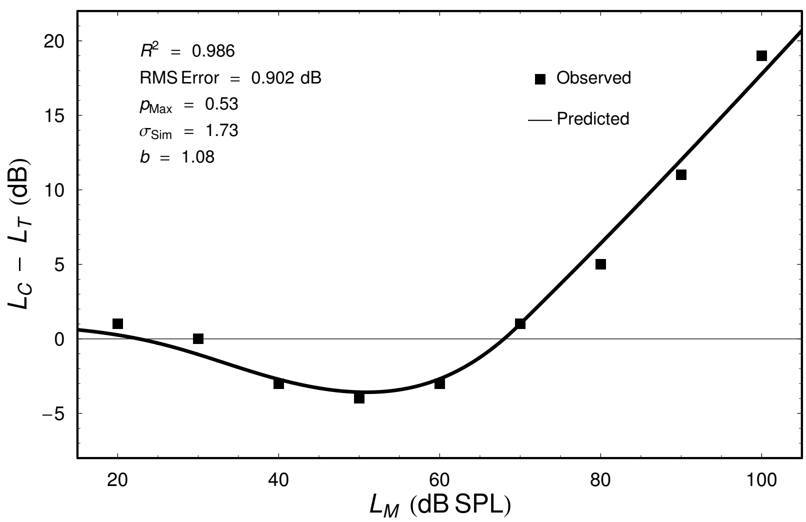

Fig. 45: Elmasian and Galambos (1975): one subject, ipsilateral forward masker, 70 dB SPL target. Observed (squares) and predicted (line) loudness enhancement plotted as a function of masker level. Fixed parameters: $\theta = 0.3$, $L_{ThM} = L_{ThT} = L_{ThC}$ = 28 dB SPL, $K_T = K_M = K_C$. The bias parameter b was selected to account for the match in quiet ($L_C - L_T$ = 1 dB).

Detection threshold in quiet was reported to be 28 dB SPL. In quiet, the level of the comparison tone matching loudness of the 70-dB SPL target was 71 dB SPL. The bias parameter *b* was chosen to account for this match, i.e., *b* entered the model as a fixed parameter. Again, the similarity model was successful in fitting the data, $R^2 = 0.986$, even though only two free parameters were used (Fig. 45). The best-fitting parameter values ($p_{Max} = 0.53$, $\sigma_{Sim} = 1.73$) were comparable to the estimates obtained for the ipsilateral condition discusse above.

7.1.8.2 Fixed masker level

We now turn to the two 'midlevel hump' experiments by Zeng (1994) and Plack (1996 a). In these studies, a fixed-level intense masker was used. The level of the tone presented in the second interval (the comparison tone) was fixed. The level of the masked tone (the target) was varied by the adaptive procedure.

The masker in Zeng's experiment differed in duration from target and comparison tone. The latter were 25-ms, 1-kHz tones, while masker duration was 100 ms.

The same listeners also participated in an experiment by Zeng and Shannon (1995), who reported mean detection threshold for the 25-ms tones to be 17 dB SPL. This value seems unusually high for the type of stimuli used. Nevertheless, the value of 17 dB SPL was used for fitting. According to Zwislocki (1965, p. 57), the threshold for a 100-ms 1-kHz tone can be expected to be 5.25 dB below the threshold for a 25-ms, 1-kHz tone. This value is compatible with results by Florentine, Fastl, and Buus (1988, Fig. 2) who summarized nine studies and reported the mean threshold difference for 1-kHz tones of 100-ms and 30-ms duration, respectively, to be 4 to 5 dB. Thus, L_{ThM} was set to 11.75 dB SPL.

Even at levels well above threshold, loudness of a 100-ms tone exceeds loudness of a 25-ms tone (e.g., Florentine, Buus, and Poulsen, 1996). The loudness integration model by Zwislocki (1969, p. 436) predicts the ratio to be 1.418, so that the scale factor K_M (Eq. 7-2) was set to $1.418 \cdot K_T$, while K_T and K_C were set to unity.

In Zeng's experiment, the level of the third tone was fixed in each run and the level of the masked tone (presented in interval 1) was adjusted by the adaptive procedure. In the following, the terminology normally used in loudness enhancement experiments is adopted: the masked tone is termed 'target' and the tone presented in interval 2 is termed 'comparison tone' (Fig. 5).

Zeng (1994) reported no loudness matches in quiet. Therefore, the bias parameter *b* was a free parameter.

As comparison tone level and not target tone level was the independent variable in the experiment and Eq. 7-12 can not be solved analytically for target pressure P_T, a "brute force" approach had to be used to find the parameter values minimizing the sum of squared residuals. The parameter p_{Max} was varied between 0.2 and 1.5 with a step size of 0.1, σ_{Sim} was varied between 0.1 and 2.5 with a step size of 0.1, and *b* was varied between 0.7 and 1.3 with a step size of 0.05. The sum of squared residuals (observed L_T minus predicted L_T) was computed for each of the 4900 combinations. The parameter combination resulting in the minimum sum of squares was selected as the best fit.

Observed and predicted loudness enhancement is plotted as a function of L_C in Fig. 46. The model produced a reasonable fit ($R^2 = 0.877$) to the mean data and correctly predicted the midlevel hump. Predicted loudness enhancement was larger than observed for the 80-dB SPL and 90-dB SPL comparison tone, but smaller at intermediate comparison tone levels. In other words, while the observed loudness enhancement function was essentially symmetric about the intermediate comparison tone level (60 dB SPL), the model predicted the function to be negatively skewed. The root-mean-square (RMS) error of 2.11 dB is not excessively large if one considers the inter-individual variability present in the data. A fit to the individual data points rather than the mean data resulted in exactly the same best-fitting values of p_{Max} and σ_{Sim}.

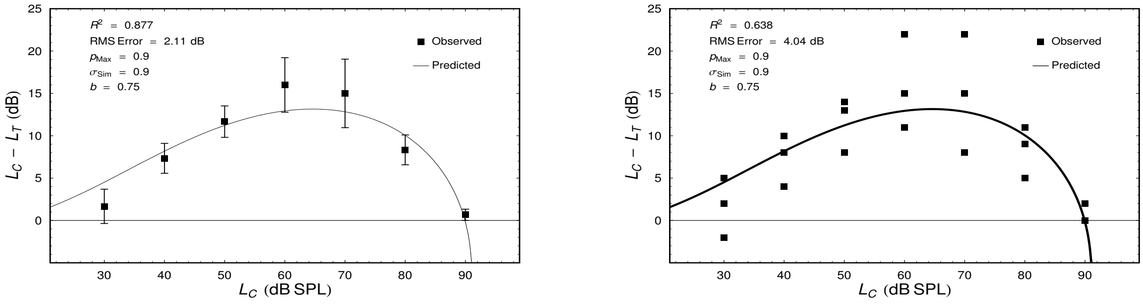

Fig. 46: Best fits to mean data (3 subjects, left column) and all data points (right column) obtained in the loudness matching experiment by Zeng (1994). Masker level was fixed at 90 dB SPL, comparison tone level was varied and target tone loudness was matched to comparison tone loudness. 1-kHz tones; 100 ms masker, 25 ms target. Fixed parameters: $\theta = 0.3$, $L_{ThM} = 11.75$ dB SPL, $L_{ThT} = L_{ThC} = 17$ dB SPL, $K_M = 1.418 \cdot K_T$, $K_T = K_C = 1$. The best-fitting value of the bias parameter b was used. Error bars show ± 1 SEM. The insets show R^2, RMS error and best fitting parameter values.

Note that the estimated value of p_{Max} was 0.9. As masker weight and target weight sum to unity (Eq. 7-9), such a value means that masker loudness received considerably *more* weight than target loudness at target levels within approximately a 10 dB range around masker level. Certainly, it is not altogether impossible that the listeners adopted such strategy, but is also not plausible why the listeners should have not only been unable to ignore the masker, but weighted it even stronger than the target.

Inspection of fits obtained with various parameter values suggested that the non-zero loudness match in the $L_M = L_C = 90$ dB SPL condition and the resulting value of b smaller than unity were the cause of the large estimate of p_{Max}. For comparison, the model was fitted again with the bias parameter b set to unity. The RMS error of this fit (Fig. 47) was larger than above, but the estimated value of p_{Max} was 0.677 for the fit to the mean data now, which seems to be a more realistic value.

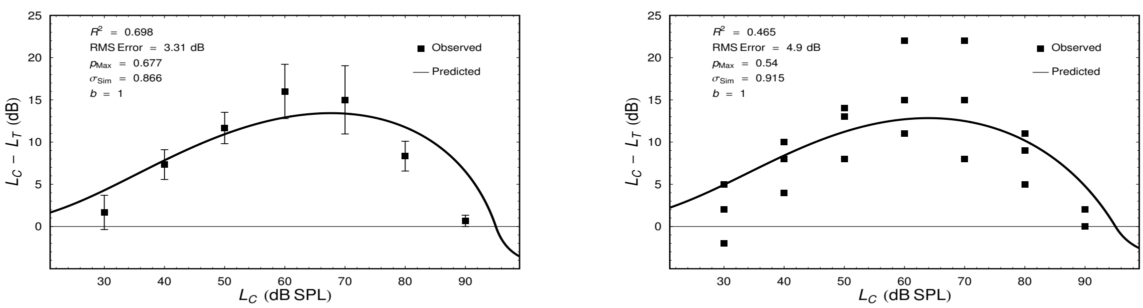

Fig. 47: Best fits to mean data (3 subjects, left column) and all data points (right column) obtained in the loudness matching experiment by Zeng (1994), with the bias parameter b set to 1.0. All other settings as in Fig. 46.

The estimated values of σ_{Sim} were smaller than for the data by Elmasian and Galambos (1975) fitted in the preceding chapter.

For the data by Plack (1996 a), who employed essentially the same setting as Zeng (1994), the model produced an excellent fit to the mean data in both the forward masking (R^2 = 0.924, Fig. 48, left panel) and in the backward masking condition (R^2 = 0.956, Fig. 48, right panel). Note that the negative skewness of the loudness enhancement function predicted by the model was actually observed by Plack.

The best-fitting value of the bias parameter b was used because no loudness matches in quiet were reported. As no detection thresholds were given, threshold level was assumed to be 8 dB SPL for the 30-ms tones, corresponding to results by Zeng, Turner, and Relkin (1991) and from Experiment 1 of this thesis. Threshold and loudness for the 110-ms masker were again assumed to deviate from target loudness according to the equations provided by Zwislocki (1965, 1969). The same fitting procedure as for the Zeng (1994) data was used.

As to the difference between forward and backward masking, the best fitting parameter values indicate that the slightly smaller loudness enhancement in backward masking was due to a difference in p_{Max} rather than σ_{Sim}. Maximum masker weight p_{Max} was 36% smaller in the backward masking condition, while σ_{Sim} changed by 2% only.

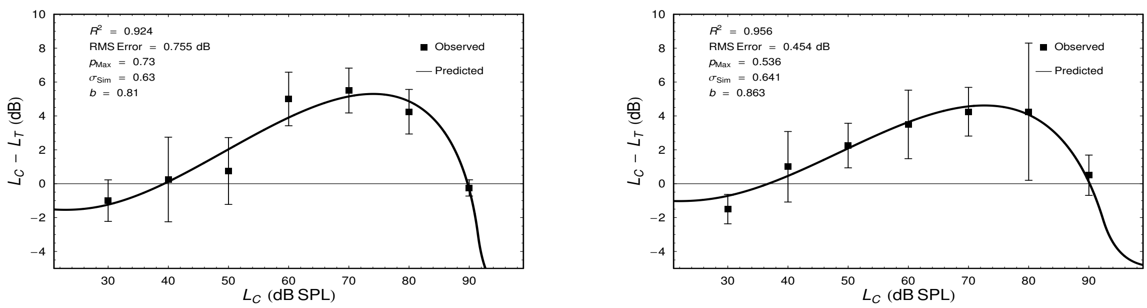

Fig. 48: Best fits to mean data (4 subjects) obtained in the loudness matching experiments by Plack (1996 a). Left panel: forward masking. Right panel: backward masking. Masker level was fixed at 90 dB SPL, comparison tone level was varied in 10 dB steps, and target tone loudness was matched to comparison tone loudness. The insets show R^2, RMS error and best fitting parameter values. Fixed parameters: $\theta = 0.3$, L_{ThM} = 2.75 dB SPL (estimated), $L_{ThT} = L_{ThC}$ = 8 dB SPL (estimated), $K_M = 1.418 \cdot K_T$, $K_T = K_C = 1$. The best-fitting value of the bias parameter b was used.

7.1.8.3 Masker and target level varied

The data obtained in Experiment 2 of the present study allow testing the predictions of the model concerning the effects of both the masker-target level difference and target level.

The bias parameter b entered the model as a fixed parameter as it could be computed from the matches in quiet at each target level. In other words, the data were fitted using two free parameters only. A separate value of b was used for each target level, as the time order error varies with level (Hellström, 1985). Average detection threshold for the four listeners was used (cf. Table 1).

The RMS error of 1.55 dB (Fig. 49, upper panels) can be considered small compared to the inter-individual variability. The proportion of variance accounted for by the model was R^2 = 0.867 and the best fitting values were p_{Max} = 0.606 and σ_{Sim} = 0.708. Due to the considerable inter-individual variation, a fit to all data points rather than to the means was rather weak (R^2 = 0.467), but the parameter estimates were similar to those obtained for the mean data, p_{Max} = 0.614 and σ_{Sim} = 0.717. In the lower panel of Fig. 49, predicted masker weight p_M is plotted as a function of $\Delta L_{M\text{-}T}$ at the three target levels.

As discussed above, the predictions of the model are compatible with almost all aspects of the data, despite the greater-than-predicted loudness decrement at a target level of 55 dB SPL.

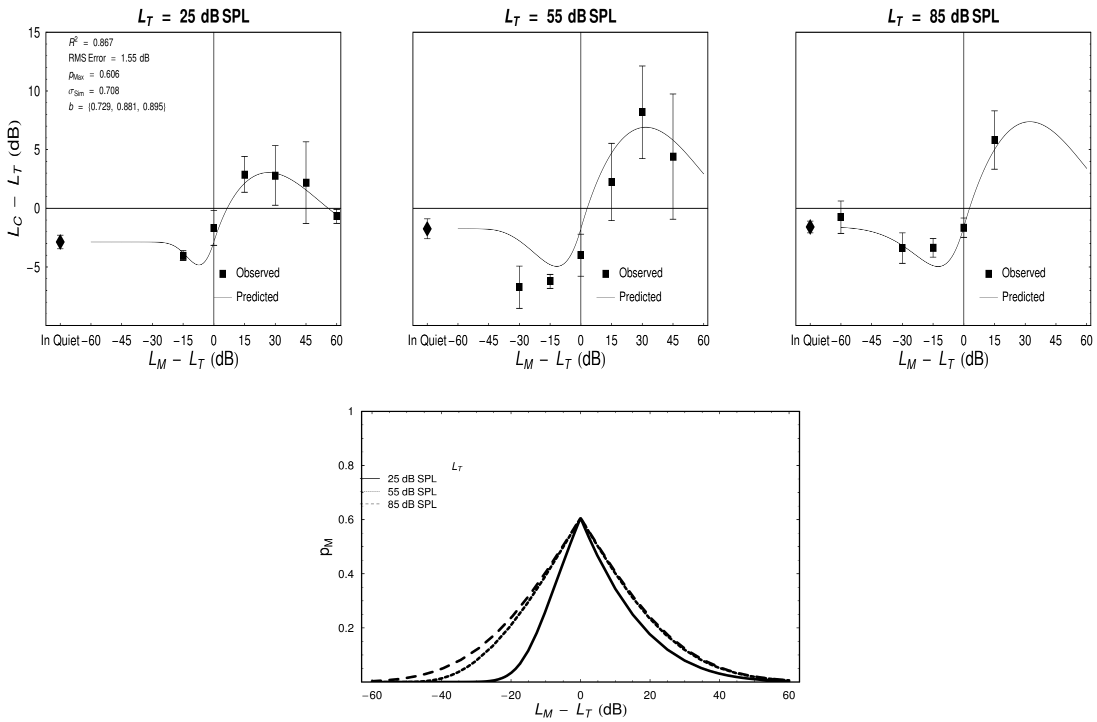

Fig. 49: Upper panels: Best fits to mean data (4 subjects) from Experiment 2. The inset shows R^2, RMS error and best fitting parameter values. Fixed parameters: $\theta = 0.3$, $L_{ThM} = L_{ThT} = L_{ThC}$ = 6.59 dB SPL, $K_M = K_T = K_C$. From the matches in quiet, a separate value of b was estimated for each target level, so that the nonlinear fit used two free parameters only (p_{Max} and σ_{Sim}). Error bars show ± 1 SEM. Lower panel: corresponding p_M.

For fits to the individual data, the bias parameter b was selected to account for the loudness match observed in the $L_M = L_T$ condition at each target level, as matches in quiet and at $L_M = L_T$ differed considerably for some listeners. The reason for these differences is unknown (cf. the discussion in Chapter 6.6). Individual detection thresholds as displayed in Table 1 were used for the parameters $L_{ThM} = L_{ThT} = L_{ThC}$.

For listener AL (Fig. 50, first row), the fit was reasonable at the 25-dB SPL and the 85-dB SPL target levels, but considerably weaker at the intermediate target level. Note the large variability in the loudness matches obtained in the latter condition.

For listener DO (second row), it seems that the predicted effect of loudness similarity was stronger than observed at the 25-dB SPL target level and weaker than observed at the 55-dB target level. For the 25-dB SPL target, the predicted loudness enhancement peak was located at too small a masker-target level difference, and at too high a value of $\Delta L_{M\text{-}T}$ for the 55-dB SPL target.

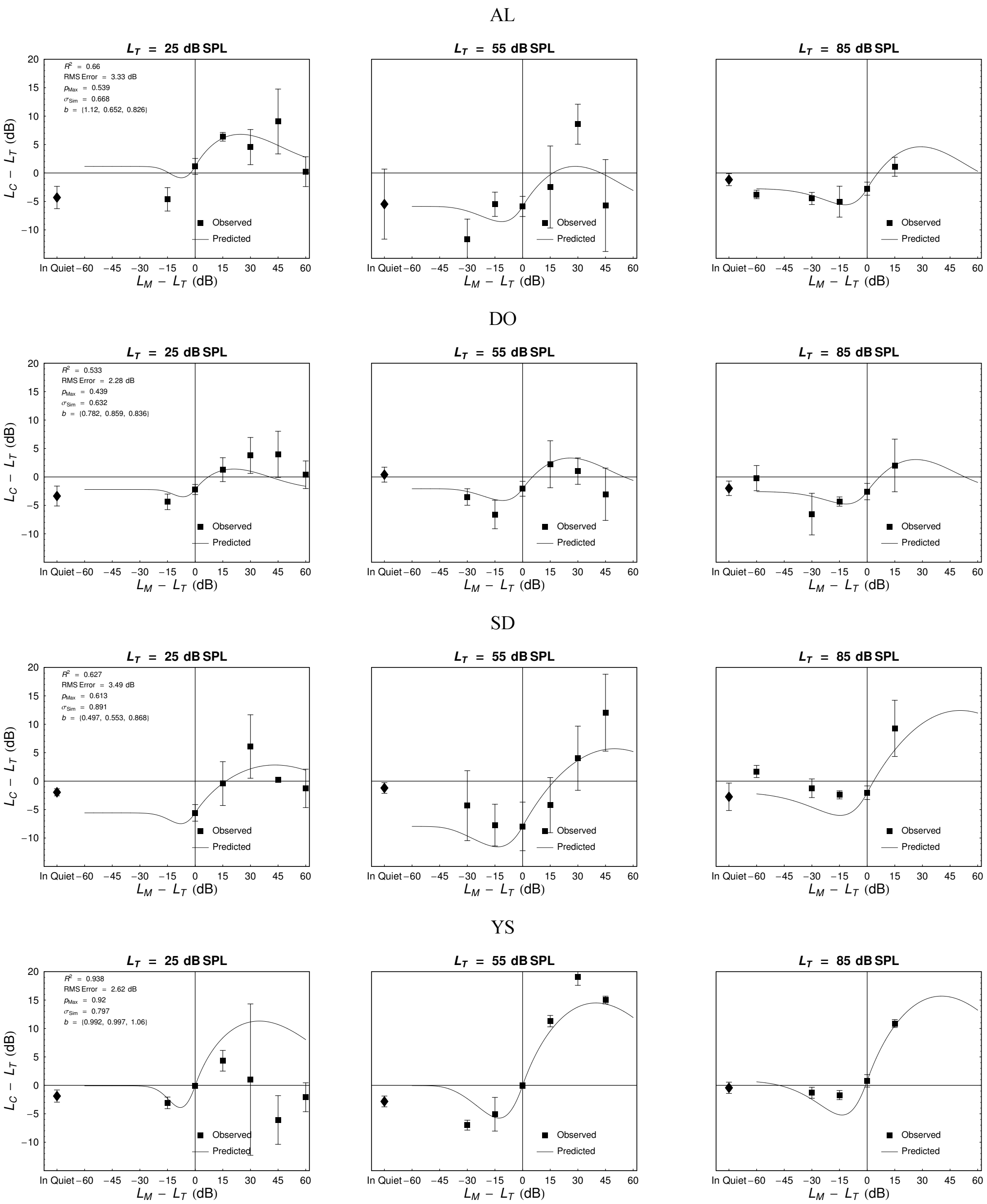

Fig. 50: Best fits to individual data from Experiment 2. Fixed parameters: $\theta = 0.3$, $K_M = K_T = K_C$. Individual detection threshold as displayed in Table 1 were used for the parameters $L_{ThM} = L_{ThT} = L_{ThC}$. The bias parameter b was selected to account for the loudness match observed in the $L_M = L_T$ condition at each target level. Weighted nonlinear regression was used for listener YS (see text). Error bars show ± 1 SD.

For masker level equal to target level, listener SD (third row) produced loudness matches much more negative than the matches in quiet at the lower levels. Additionally, at the intermediate target level, the loudness match at $L_M = L_T$ was more negative than for masker levels *below* target level. For this reason, the mismatch between observed and predicted and observed loudness enhancement seems to be due to a shift of the predicted loudness enhancement function. If the minima of this function were aligned with the observed minima, a much better fit would result. It remains unclear, why the minimum loudness match was observed for equal masker and target level rather than for a masker level below target level.

For listener YS (lower row), the observed matches at L_T = 25 dB SPL were *negative* at masker-target level differences of +30 and +45 dB. Certainly, this pattern is incompatible with the model. The variability at these data points was very large, especially at $\Delta L_{M\text{-}T}$ = +45 dB because YS produced both positive and negative matches in this condition (Fig. 51). To reduce the influence of the most variable data points, the fit displayed in Fig. 50 was obtained using a weighted nonlinear regression procedure in which each observation was weighted by its inverse variance. RMS error and R^2 were also computed using the $1/\sigma^2$ weights. At the higher target levels, the fit was reasonable. The observation of negative matches for positive masker-target level differences might be explained by a contrast effect. Plack (1996 a) also reported negative matches for low-level targets combined with a 90-dB SPL masker for one the listeners in his study and proposed that the listener "[…] may have 'overcompensated' for the intense masker by reducing her estimate of the loudness of the masked tone […]" (p. 1029). Note that the estimate of p_{Max} was again considerably greater than 0.5. Fitting the data for YS with the value of p_{Max} = 0.606 obtained for the mean data resulted in only a small increase in RMS error from 2.62 to 2.98 dB, however.

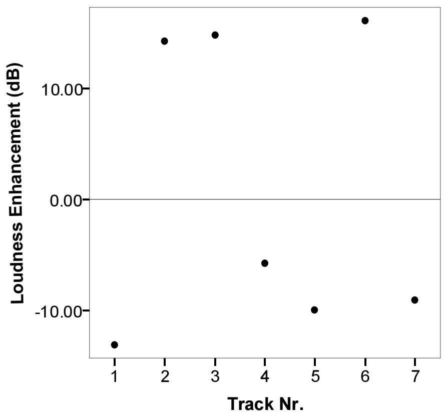

Fig. 51: Loudness matches produced by listener YS for the 25-dB SPL target combined with the 55-dB SPL masker.

The parameter estimates obtained for the mean data from Experiment 2 can be used to compute the loudness function of the target tone according to Eq. 7-10 at different masker levels (Fig. 52). All forward-masked functions show a characteristic inflection point at $L_T = L_M$ that represents the shift from loudness enhancement at target levels smaller than masker level to loudness decrement at target levels greater than masker level. With the 85-dB SPL masker (long-dashed line), the midlevel hump in loudness enhancement is predicted, as loudness of the masked tone is above loudness in quiet at intermediate target levels only.

7.1.9 Discussion

Fits of the quantitative version of the similarity model to data from experiments using a wide variety of stimuli and also different procedures were reasonable to excellent, even though only as few as only two free parameters were used in some cases. The similarity model correctly predicts the most important findings, as for instance the mid-difference humps.

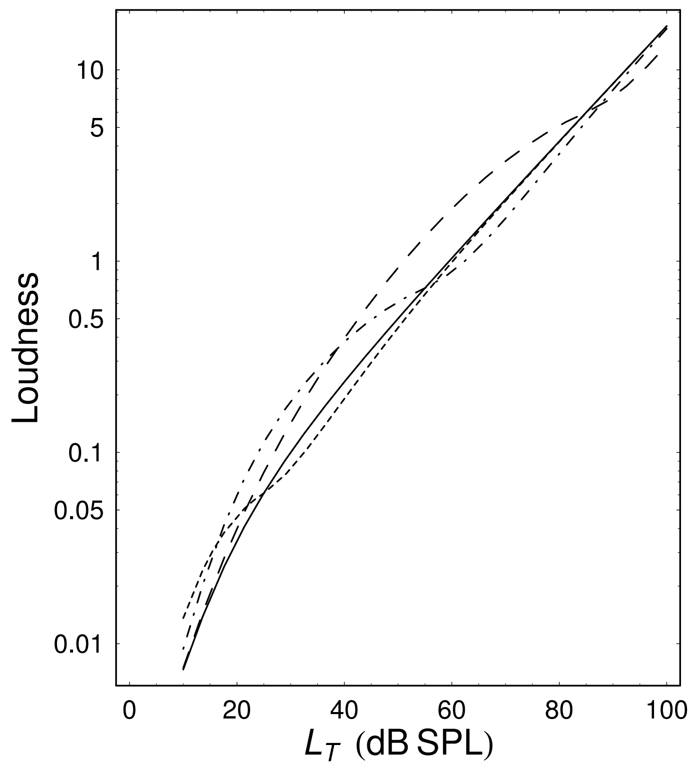

Fig. 52: Loudness functions of a non-simultaneously masked tone as predicted by the similarity model. Solid line: in quiet. Short-dashed line: 25-dB SPL forward masker. Dot-dashed line: 55-dB SPL forward masker. Long-dashed line: 85-dB SPL forward masker. The parameter estimates obtained for the mean data from Experiment 2 were used, p_{Max} = 0.606, σ_{Sim} = 0.708. Fixed parameters: θ = 0.3, $L_{ThM} = L_{ThT} = L_{ThC}$ = 6.59 dB SPL (Table 1), $K_T = K_M$ = 11.2, P_0 = 20 μPa.

It can also explain why the effects of the masker are least pronounced at the **lowest** target level, as this observation can be attributed to the slope of the loudness function. The estimates of the parameters p_{Max} and σ_{Sim} obtained for the different data set were of the same order of magnitude.

Discrepancies between observed and predicted values seemed in most cases to be related to the presence of non-zero loudness matches in quiet and for maskers equal in level to the target.

Clearly, accounting for the rather complex pattern of findings outlined in the time-order error literature (cf. Hellström, 1985) is beyond the scope of the similarity model designed to explain loudness perception in the presence of non-simultaneous masking. Inclusion of a bias parameter representing trace drift provided a pragmatic solution to this problem. Still, the observation of loudness matches in quiet differing from matches in the condition of equal-level masker and target presents a problem for the model, which predicts loudness enhancement $L_C - L_T$ to be identical to the match in quiet in the latter condition.

An additional critical issue are the best-fitting values of p_{Max} considerably greater than 0.5 that were computed for the data by Zeng (1994) and for listener YS from Experiment 2. These estimates are difficult to justify, as they indicate that masker loudness received **more** weight than target loudness at small differences between masker and target level. This conclusion must be qualified by the observation that using smaller values of p_{Max} resulted in only a small increase in RMS error. This finding indicates that the number of data points was too small to allow for very precise parameter estimates.

To conclude, the similarity model was successfully formalized by combining well-established properties of the auditory periphery (loudness function) and higher-level effects (target-distractor similarity).

At this point, it is important to note that in a recent paper, which I became aware of after completion of the experiments and the modeling described in Chapters 6 and 7, Scharf, Buus, and Nieder (2002) argue that the loudness matches obtained in a three-tone matching do not reflect an effect of the masker on the loudness of the target tone. Scharf et al. used a comparison tone much higher in frequency than the target and found *no* evidence for loudness enhancement. They interpreted this finding as to show that "loudness recalibration" (cf. Marks, 1994), or "induced loudness reduction (ILR)", as Scharf et al. termed it, is the cause of the effects found in three-tone matching tasks using target and comparison tones of the same frequency. The loudness recalibration experiments showed that a moderately strong tone (e.g., 80 dB SPL) *reduces* the loudness of a weaker tone (e.g., 60 dB SPL) following with an ISI of about 1 s (e.g., Mapes-Riodan and Yost, 1999). The loudness reduction was found only for test tones similar in frequency to the inducer tone (or "masker"). Based on the observation by Nieder et al. (2003) that -contrary to previous assumptions (Mapes-Riordan and Yost, 1999; Scharf, 2001)- loudness recalibration is in effect even for tone durations smaller than 100 ms, Scharf et al. (2002) argued that with masker, target, and comparison tone presented at the same frequency, the masker did not enhance loudness of the target (the second tone in the trial) but rather *reduced* loudness of the comparison tone (the third tone), resulting in positive comparison-target level differences produced in the loudness matches. This would explain "loudness enhancement" as defined in the three-tone matching task, i.e., comparison tone level adjusted to higher levels than the target. To explain why the target tone is not subject to loudness reduction, however, the assumption has to be made that "[…] a weak tone presented in close temporal proximity to a stronger tone somehow is protected from ILR." (Scharf et al., 2002, p. 809). While the reason for such a pattern remains unclear, some evidence for this idea was provided by Arieh and Marks (2003), who also used a three-tone matching task with conditioner and target at 2500 Hz and the comparison tone at 500 Hz. They found virtually no loudness reduction (but also no clear loudness enhancement) for conditioner-target ISIs smaller than 200 ms, but loudness reduction of up to 13 dB at longer ISIs. Taken together, these recent data make it necessary to re-interpret the data obtained in three-tone loudness matching tasks. It remains to be shown whether the entire loudness enhancement" data (for example, the effects observed in Experiment 2 of this work) can be explained by loudness recalibration. For instance, Nieder et al. (2003) found virtually no effect of conditioner level (80 dB SPL versus 95 dB SPL) on the loudness reduction induced in 60-dB SPL and 70-dB SPL target tones. In contrast, in Experiment 2 of this study, mean loudness matches for a 55-dB SPL target obtained with a 70-dB SPL and an 85-dB SPL conditioner, respectively, differed by 7 dB (Fig. 28). Equally important, the observation of loudness *decrement*, i.e., the loudness of the target being matched by a comparison tone *lower* in level than the target, could be explained by loudness recalibration only if an inducer tone less intense than the test tone produced an *increase* rather than a reduction in test tone loudness. Mapes-Riordan and Yost (1999) found virtually no effect of a 40-dB SPL inducer on the loudness of an 80-dB SPL test tone, but the the level difference between inducer and test tone may have been too large to produce loudness decrement. To resolve these issues, additional experiments are necessary.

7.2 Intensity resolution

The loudness enhancement hypothesis by Carlyon and Beveridge (1993), which is part of the similarity model, assumes that loudness changes induced by a forward or backward masker lead to increased loudness variability, so that an increase in the just-noticeable difference results.

While this idea seems plausible, it would be preferable to have a formal model of why loudness enhancement causes a jnd elevation. Existing models of the relation between loudness and intensity resolution discussed in Chapter 4.2.1 fail to account for the large (midlevel) jnd elevations observed under non-simultaneous masking.

The idea of mergence influencing not only target loudness effective in a loudness matching task, but also the decision variable in an intensity discrimination experiment represents a model where a common mechanism underlies both masker-induced loudness changes and jnd elevations. This common mechanism (mergence) is the inclusion of masker information. A study by Stellmack and Viemeister (2000) provided direct evidence for the inclusion of masker level information in the decisions made in an intensity discrimination task. The authors estimated the relative weight assigned to forward and backward maskers in a cued one-interval intensity discrimination task. Stimuli were 10-ms, 1-kHz tone pulses. In the first observation interval, two identical pulses (the cues) were presented at a level of 75 dB SPL. The second interval also contained two pulses. In each trial, the level of each of the two pulses was drawn independently and at random from a normal distribution with an expected value of 75 dB SPL. In one condition, listeners decided whether the first pulse in interval 2 was smaller or higher in level than the cues. They were instructed to ignore the second pulse (the 'backward masker'). In the other condition, the target tone was the second pulse in interval 2 ('forward masking'). From the binary responses, Stellmack and Viemeister computed relative weights for target tone and masker (Berg, 1989; Richards and Zhu, 1994). The weight assigned to the masker was greater than zero even at the longest masker-target ISI (128 ms). It was maximum at ISIs between 4 and 8 ms, slightly smaller at an ISI of 2 ms, and otherwise decreased with the ISI. Masker weight was smaller in the backward than in the forward masking condition. It should be noted that the just described procedure differed from a normal intensity discrimination experiment, as masker level fluctuated randomly from trial to trial. In fact, the masker-induced change in percent correct was smaller in a condition where masker level was fixed and identical to mean target level (see Fig. 2 in Stellmack and Viemeister, 2000). The authors also measured difference limens for 10-ms, 1-kHz tone pulses presented at 75 dB SPL and combined with forward or backward maskers at masker-target level differences of −10, 0, or +10 dB. The jnd elevation (relative to the jnd in quiet) only very slightly decreased with the masker-target ISI, which stands in contrast to the pronounced ISI dependence of masker weight found in the first experiment by Stellmack and Viemeister. This emphasizes the notion that the condition with randomly fluctuating masker level is not directly comparable to the usual condition of a fixed-level masker.

In this chapter, a signal-detection model will be used to derive predictions regarding the effect of mergence on intensity discrimination performance. In a two-interval intensity discrimination task, according to the SDT framework, inclusion of masker intensity information influences the decision variable, which is the difference between the random variables representing the observations of standard-plus-increment and standard, respectively. Additionally, sensation noise associated with the masker contributes to the variance of the decision variable. Finally, the masker could cause a response bias resulting in a drop in performance in a 2I, 2AFC task (cf. Green and Swets, 1966).

First, a general framework for modeling the decision process in an intensity discrimination experiment where mergence is effective is introduced.

In the second step, the distributions of the decision variables and the relation between the physical parameters of the stimuli under test and the hypothesized internal continua are specified according to two alternative models of intensity discrimination.

Finally, predictions of the two resulting decision models are discussed for three conditions that differ in the assumed characteristics of masker weight p_M.

7.2.1 Signal-detection framework for modeling intensity discrimination

According to the well-known decision model based on continuous random variables (e.g., Thurstone, 1927; Green and Swets, 1966; Durlach and Braida, 1969), the presentation of each stimulus elicits a sensation (a value on the internal continuum). The decision is based on these transformations of the acoustic waveform of each stimulus, and the sensory effects are modeled as real-valued random variables.

In the two-interval forced-choice intensity discrimination task, the standard is presented in one interval and the standard-plus-increment in the other interval. So we have two possible orders of presentation: the increment occurs either in the first interval (denoted by $<S+I, S>$) or in the second interval ($<S, S+I>$).

Following Green and Swets (1966, p.68*f*), we assume that the listener makes an observation x_1 of the intensity of the tone presented in the first interval and an observation x_2 of the tone intensity presented in the second interval. He or she uses a decision variable such that forced-choice decisions are based on the difference in magnitude $(x_1 - x_2)$. Specifically, if $(x_1 - x_2) > C$, the listener will respond that the increment occurred in interval 1. If $(x_1 - x_2) \leq C$, he or she will respond that the increment occurred in interval 2. According to the usual Gaussian decision model (cf. Green and Swets, 1966, p. 66), the observations x_1 and x_2 are normally distributed with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively. Table 2 displays the four proportions observed in the experiment, together with the predicted values.

Response	*State of the World*	
	$<S+I, S>$	$<S, S+I>$
"Increment in Interval 1"	P('Incr. in Interval 1'|$<S+I, S>$) = $P(x_1 - x_2 > C) = \int_{C}^{+\infty} \mathrm{PDF}[N(\mu_{S+I} - \mu_S, \sigma_{Diff}), z]\, dz$	P('Interval 1'|$<S, S+I>$) = $P(x_1 - x_2 > C) = \int_{C}^{+\infty} \mathrm{PDF}[N(\mu_S - \mu_{S+I}, \sigma_{Diff}), z]\, dz$
"Increment in Interval 2"	P('Interval 2'|$<S+I, S>$) = $P(x_1 - x_2 \leq C) = \int_{-\infty}^{C} \mathrm{PDF}[N(\mu_{S+I} - \mu_S, \sigma_{Diff}), z]\, dz$	P('Interval 2'|$<S, S+I>$) = $P(x_1 - x_2 \leq C) = \int_{-\infty}^{C} \mathrm{PDF}[N(\mu_S - \mu_{S+I}, \sigma_{Diff}), z]\, dz$

Table 2: Signal Detection Model of the 2I, 2AFC level discrimination task.

In Table 2, PDF[$N(\mu, \sigma)$, x] denotes the probability density function of the normal distribution with mean μ and standard deviation σ evaluated at the point x.

The standard deviations of S and $S+I$, respectively, need not to be equal. The standard deviation of the decision variable $(x_1 - x_2)$ is

$$\sigma_{Diff} = \sqrt{\sigma^2_1 + \sigma^2_2 + 2\, r(x_1, x_2)\, \sigma_1\, \sigma_2} \qquad (7\text{-}16)$$

where $r(x_1, x_2)$ is the correlation coefficient. In the following, x_1 and x_2 are generally assumed to be independent, so that $r(x_1, x_2) = 0$. With this value, the above equation simplifies to

$$\sigma_{Diff} = \sqrt{\sigma^2_1 + \sigma^2_2} = \sqrt{\sigma^2_S + \sigma^2_{S+I}}. \qquad (7\text{-}17)$$

If the increment occurs with equal probability in both intervals, it follows that the proportion of correct responses, P(Corr), is the arithmetic mean of two conditional probabilities

$$\text{P(Corr)} = \tfrac{1}{2}\, P(\text{'Interval 1'}|\langle S{+}I, S\rangle) + \tfrac{1}{2}\, P(\text{'Interval 2'}|\langle S, S{+}I\rangle). \qquad (7\text{-}18)$$

Using Table 2, we find

$$\text{P(Corr)} = \tfrac{1}{4}\left(2 + \text{Erf}\left(\frac{C + \mu_{S+I} - \mu_S}{\sqrt{2}\sigma_{Diff}}\right) - \text{Erf}\left(\frac{C - (\mu_{S+I} - \mu_S)}{\sqrt{2}\sigma_{Diff}}\right)\right). \qquad (7\text{-}19)$$

It can be shown that P(Corr) is maximum if $C = 0$, i.e., if the decision is unbiased. With this value of C, Eq. 7-19 reduces to

$$\text{P(Corr}|\, C = 0) = \tfrac{1}{2}\left(1 + \text{Erf}\left(\frac{\mu_{S+I} - \mu_S}{\sqrt{2}\sigma_{Diff}}\right)\right) \qquad (7\text{-}20)$$

Given the structure of a 2I, 2AFC task where the increment occurs with equal probability in both intervals and the fact that feedback is provided in most experiments, it is usually assumed that an observer will adopt this optimum unbiased cutoff.

In the model, there are two distributions of $x_1 - x_2$: one with mean $\mu_{S+I} - \mu_S$ corresponding to the order of presentation $\langle S{+}I;\, I\rangle$ and the other with mean $-(\mu_{S+I} - \mu_S)$ corresponding to the order of presentation $\langle S;\, S{+}I\rangle$. They are separated by $2 \cdot (\mu_{S+I} - \mu_S)$. The standard deviation $\sigma_{Diff} = \sqrt{\sigma^2_S + \sigma^2_{S+I}}$ is just $\sqrt{2}\sigma_S$ in the equal-variance case. If $C = 0$, the separation of the means divided by the standard deviation is, by definition, the SDT measure of sensitivity

$$d'_{2AFC} = 2\sqrt{2} \cdot \text{Erf}^{-1}[2 \cdot P(\text{Corr}) - 1] = \frac{2\,(\mu_{S+I} - \mu_S)}{\sigma_{Diff}}, \qquad (7\text{-}21)$$

which simplifies to

$$d'_{2AFC} = \frac{\sqrt{2}\,(\mu_{S+I} - \mu_S)}{\sigma_S} \qquad (7\text{-}22)$$

in the equal-variance case.

Note that this definition of d' follows the notation by Green and Swets (1966), which emphasizes that d'_{2AFC} predicted for the 2AFC task (equal-variance case, unbiased choice) is $\sqrt{2}$ times d'_{YN} in a yes-no task employing the same standard-plus-increment and standard (see Jesteadt and Bilger, 1974, for a discussion).

The only assumptions concerning the decision variable made so far are that x_1 and x_2 are independent and normally distributed.

The models proposed in the literature use different assumptions regarding the relation between physical tone intensity I on the one hand and the mean and standard deviation of the random variable x representing the sensory effect of the tone on the other hand. It is obvious that the expected value of x should be a monotonic function of tone level. Two possible relations will be discussed in Chapter 7.2.3.

7.2.2 Effects of mergence

According to the SDT framework, if a forward or backward masker is present in one or both of the observation intervals, the effect of the masker on intensity discrimination performance must be due to the masker influencing the decision variable.

In line with the similarity model of loudness matches in the presence of a non-simultaneous masker (Chapter 7.1), it is assumed that the two observations (made in interval 1 and interval 2) forming the decision variable are weighted averages of the responses (e.g., spike counts) to masker and to target, respectively. For example, if a forward masker is presented in the first interval, the observation x_1 made in interval 1 is modeled as a weighted average between the response to the target tone (standard or standard-plus-increment) and the response to the masker.

If x_{1M} denotes the numerical value of the sensory effect of masker intensity and x_{1T} the sensory effect of target tone intensity, the value representing interval 1 is

$$x_{1Merged} = p_{M1}\, x_{1M} + (1 - p_{M1})\, x_{1T}, \qquad (7\text{-}23)$$

where p_{M1} is the weight given to masker intensity in interval 1. According to the similarity model (Eq. 7-8), p_{M1} is a function of masker and signal intensity.

If μ_{1M}, and μ_{1T} are the means of the random variables x_{1M} and x_{1T}, respectively, the expected value of $x_{1Merged}$ is

$$\mu_{1Merged} = p_{M1}\, \mu_{1M} + (1 - p_{M1})\, \mu_{1T}. \qquad (7\text{-}24)$$

For the tones presented in interval 2, $x_{2Merged}$ and $\mu_{2Merged}$ can be computed analogously given the weight p_{M2} assigned to the masker in interval 2. The decision variable is again the difference $x_{1Merged} - x_{2Merged}$. Table 3 displays the resulting means of the values representing the observations made in interval 1 and interval 2, respectively.

State of the world	Observation Interval 1	Interval 2
$<S + I, S>$	$\mu_{1Merged\|<S+I,S>} = p_{M1}\,\mu_M + (1 - p_{M1})\,\mu_{S+I}$	$\mu_{2Merged\|<S+I,S>} = p_{M2}\,\mu_M + (1 - p_{M2})\,\mu_S$
$<S, S + I>$	$\mu_{1Merged\|<S,S+I>} = p_{M1}\,\mu_M + (1 - p_{M1})\,\mu_S$	$\mu_{2Merged\|<S,S+I>} = p_{M2}\,\mu_M + (1 - p_{M2})\,\mu_{S+I}$

Table 3: Means of the observations made in interval 1 and interval 2 if mergence is in effect.

If the increment occurs with equal probability in both intervals, the proportion of correct responses, $P(\text{Corr})$, is again the arithmetic mean of two conditional probabilities (Eq. 7-18). Using Table 3, we find

$$P(\text{Corr}) = \frac{1}{4}\left(2 - \text{Erf}\left(\frac{C - (\mu_{1Merged|<S+I,S>} - \mu_{2Merged|<S+I,S>})}{\sqrt{2}\sigma_{Diff|<S+I,S>}}\right) + \text{Erf}\left(\frac{C - (\mu_{1Merged|<S,S+I>} - \mu_{2Merged|<S,S+I>})}{\sqrt{2}\sigma_{Diff|<S,S+I>}}\right)\right), \quad (7\text{-}25)$$

where $\sigma_{Diff|<S+I,S>}$ and $\sigma_{Diff|<S,S+I>}$ are the standard deviations of the decision variable in the two orders of presentation.

As to the masker weights effective in the two intervals, we distinguish three cases:

1. The masker has the same effects in both intervals ($p_{M1} = p_{M2}$). Masker weight is constant in all trials presenting the same condition.
2. The masker receives different weights in the two intervals ($p_{M2} = m \cdot p_{M1}$). The two masker weights are constant in all trials presenting the same condition.
3. Masker weight is a random variable, so that the weights assigned to masker loudness in the two observation intervals fluctuate from trial to trial.

The variances of the two observations, and consequently also the variance of the decision variable, depend on which of the above three cases is modeled.

If the masker weights are fixed for all trials presenting the same combination of masker and standard level (Cases 1 and 2), as x_{1M} and x_{1T} are independent, the variance of $x_{1Merged}$ is

$$\text{Var}_{1Merged} = p_{M1}^2\,\sigma_{1M}^2 + (1 - p_{M1})^2\,\sigma_{1T}^2, \quad (7\text{-}26)$$

and the variance of $x_{2Merged}$ can be computed in exactly the same way.

If the masker weights are modeled as independent random variables P_{1M} and P_{2M} rather than fixed quantities (Case 3), we can use the formula for the exact variance of the product of two independent variables x and y with expected values $E(x)$ and $E(y)$ and variances $\text{Var}(x)$ and $\text{Var}(y)$, respectively,

$$\text{Var}(x \cdot y) = E^2(x)\,\text{Var}(y) + E^2(y)\,\text{Var}(x) + \text{Var}(x)\,\text{Var}(y) \quad (7\text{-}27)$$

(Goodman, 1960).

Using Eqs. 7-24 and 7-27, we find that

$$\mathrm{Var}_{1Merged} = \sigma_{1T}^{2}\,[(1-p_{M1})^{2} + \sigma_{pM1}^{2}] + \sigma_{1M}^{2}\,(p_{M1}^{2} + \sigma_{pM1}^{2}) + \sigma_{pM1}^{2}(\mu_{1M}^{2}+\mu_{1T}^{2}), \qquad (7\text{-}28)$$

where p_{M1} and σ_{pM1} denote the mean and the standard deviation of P_{M1}, respectively, and all of the relevant random variables are assumed to be independent. $\mathrm{Var}_{2Merged}$ is computed analogously.

It is evident that mergence has an effect on

a) the expected value of the decision variable ($x_{1Merged} - x_{2Merged}$) and

b) on its variance.

In the next step, the decision variable is specified according to two alternative relations between physical sound intensity and the internal continuum.

7.2.3 Specifying the decision variable: Two approaches

The choice of the decision variable in models of intensity discrimination has been a continuous topic of debate (e.g., Laming, 1986), so that a large number of models has been proposed. The models can be classified into such explicitly recurring to properties of neural responses in the auditory periphery (e.g., McGill and Goldberg, 1968 a, b; Luce and Green, 1972; Viemeister, 1988) and more ‘abstract’ models (e.g., Durlach and Braida, 1969; Dau, Püschel, and Kohlrausch, 1996).

Following Laming (1986), it must be emphasized that an SDT model based on continuous random variables is useful for describing data from many experiments, but that it is difficult to decide between different possible distributions of the decision variables and different possible relations between the physical parameters of the stimuli under test and the hypothesized internal continua on the basis of the existing psychophysical and physiological data. Moreover, once a specific decision variable is selected, the predictions of an SDT model are invariant for a large number of transformations of this variable (cf. Laming, 1986; Falmagne, 1986).

It therefore seems reasonable to adopt a pragmatic criterion by requiring that a model of intensity discrimination between pure tones presented in quiet be able to account for three observations:

1. Intensity resolution measured as *d’* is proportional to the difference in level between standard and standard-plus-increment (ΔL; Buus and Florentine, 1991; Buus, Florentine, and Zwicker, 1995; see Green, 1993, for a discussion). In other words, the proportion of correct responses is a normal integral with respect to ΔL. More precisely, Buus and colleagues fitted psychometric functions of the form $d' = a\,\Delta L^{k}$, and found the mean value of the exponent k to be 0.94 (values ranged from 0.73 to 1.29 for data from conditions with difference limens ΔL_{DL} ranging from 0.4 to 10.89).

2. Weber’s law applies within a single frequency channel (cf. Florentine and Buus, 1981; Delgutte, 1987; Viemeister, 1988).

3. The near-miss to Weber's law is observed for pure tones in quiet (e.g., Jesteadt, Wier, and Green, 1977).

The near-miss is most frequently attributed to the spread of excitation due to the mechanical properties of the basilar membrane (cf. Plack and Carlyon, 1995). For modeling the effects of a forward masker, this effect is not of essential importance. Above that, the variation of the jnd in quiet is small compared to the jnd elevations observed in non-simultaneous masking. Out of these reasons, intensity discrimination in the presence of mergence will be modeled without accounting for the near-miss. Certainly, the level dependence of intensity resolution in quiet could easily be incorporated into an extended version of the model.

7.2.3.1 Durlach and Braida (1969)

A simple model accounting for *d'* being proportional to ΔL is the "preliminary theory of intensity resolution" by Durlach and Braida (1969). This model is 'abstract' in the sense that it does not explicitly refer to physiological processes such as neural counts. Durlach and Braida use the standard Gaussian decision model, in which there is a continuous decision axis X and each presentation of a stimulus of intensity I elicits one particular value of X. The conditional probability density function $p(X|I)$ is Gaussian with mean μ_I and standard deviation σ. The authors further assume that the mean μ_I is a monotonic transform of stimulus intensity

$$\mu_I = \alpha(I) = k \log_{10} I, \qquad (7\text{-}29)$$

where k is a constant, and that the standard deviation σ is independent of sound intensity. Durlach and Braida explain the variance effective in the decision by an "internal-noise model" (cf. Chapter 4.1.2.1), according to which the variance can be partitioned into "sensation noise" (i.e., variability associated with the transformation of the acoustic waveform into a value on the internal continuum) and "memory noise" (variability associated with the memory processes involved). For the present discussion, memory noise will be ignored.

Given the foregoing formulation of the decision process, the predictions of the Durlach and Braida (1969) model can be analyzed.

For intensity discrimination in quiet and a 2I, 2AFC task, there are two observations, x_1 (interval 1) and x_2 (interval 2). The means of these random variables are

$$\begin{aligned} \mu_1 &= k \log_{10} I_1, \\ \mu_2 &= k \log_{10} I_2, \end{aligned} \qquad (7\text{-}30)$$

respectively. I_1 and I_2 denote sound intensity in the first and second interval, respectively. The following relations can be used:

$$\begin{aligned} I_S &= 10^{\frac{L_S}{10}} I_0 \\ I_{S+I} &= 10^{\frac{L_S+\Delta L}{10}} I_0 , \end{aligned} \qquad (7\text{-}31)$$

where L denotes sound pressure level measured in dB SPL and I_0 is reference intensity that determines the unit of the intensity scale; most frequently $I_0 = 10^{-12}\ W/m^2$ is used.

It follows that

$$\mu_S = k\,(L_S + \log_{10} I_0),$$
$$\mu_{S+I} = k\,(L_S + \Delta L + \log_{10} I_0). \qquad (7\text{-}32)$$

Inserting these values into Eq. 7-20 (note that $C = 0$ is assumed) and converting P(Corr) to d' results in

$$d' = \frac{\sqrt{2}\,k\,\Delta L}{10\,\sigma}, \qquad (7\text{-}33)$$

where σ is the standard deviation of each observation (x_1 and x_2). As the standard deviation is assumed to be level independent, d' is independent of standard level L_S. The Durlach and Braida model thus predicts Weber's law, which is reasonable for the performance of a single channel. Equally important, d' is a linear function of ΔL.

The near miss to Weber's law could be accounted for by expressing the standard deviation of the decision variable as a decreasing linear function of tone level,

$$\sigma(L_S) = a - b\,L_S. \qquad (7\text{-}34)$$

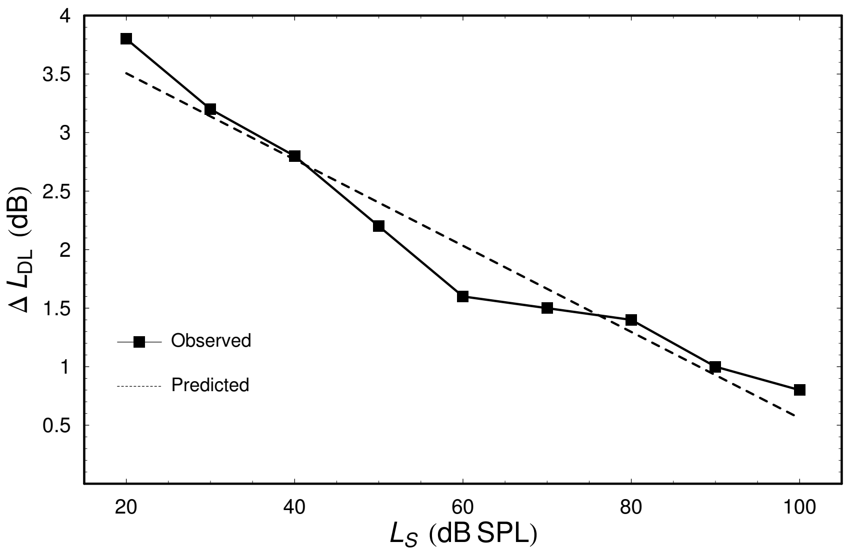

Fig. 53: Best fit of the Durlach and Braida model (Eqs. 7-33 and 7-34) to the intensity discrimination data obtained by Zeng, Turner, and Relkin (1991, Fig. 2). Parameters: $a = 5.51$, $b = 0.048$, $k = 10$, $I_0 = 1$.

For the data by Zeng, Turner, and Relkin (1991), who measured intensity discrimination for 1-kHz, 30-ms pure tones in quiet, the best fitting function is

$$\sigma(L_S) = 5.51 - 0.048\,L_S, \qquad (7\text{-}35)$$

i.e., the standard deviation of the decision variable is a slowly decreasing function of standard level (Fig. 53). As the unit of the scale of the decision variable is irrelevant, k was arbitrarily set to 10 and I_0 to unity, so that $\mu = L$, i.e., the value on the internal continuum is equal to sound-pressure-level measured in dB SPL.

7.2.3.2 Staggered-threshold model

Viemeister (1988) proposed a physiologically oriented decision model in which empirically obtained properties of auditory nerve fibers are used to predict performance in an intensity discrimination task. He used parameters of high-SR auditory nerve fibers reported by Evans and Palmer (1980). At the characteristic frequency (CF), the average spontaneous rate (SR) of 216 cat auditory nerve fibers was 55 spikes/s (SE = 1.6 spikes/s). The mean dynamic range (firing rate 1% above SR to 99% of maximum rate) was 41 dB (SE = 0.65 dB). In terms of the maximum observed firing rate r_{Max} minus the spontaneous rate, the dynamic range was 150 spikes/s (SE = 3.3 spikes/s). Using these parameters, Viemeister modeled the rate-level functions by Gaussian ogives of the form

$$r(L) = \mathrm{SR} + \mathrm{CDF}[N(\mu, \sigma), L]\,(r_{Max} - \mathrm{SR}), \qquad (7\text{-}36)$$

where CDF[$N(\mu, \sigma)$, L] is the cumulative density function of the normal distribution with mean μ and standard deviation σ evaluated at the point L. For the fibers described by Evans and Palmer (1980), mean firing rate depends on detection threshold L_{Th} and sound pressure level L as

$$r(L, L_{Th}) = 130.5 + 75\,\mathrm{Erf}(-1.75 + 0.085\,L - 0.085\,L_{Th}). \qquad (7\text{-}37)$$

To compute a measure of sensitivity, an estimate of the variability of the firing rate is needed. Teich and Khanna (1985) measured mean and variance of the spike count of cat auditory nerve fibers in response to continuous pure tones and noise, using counting intervals of 51.2 ms. They found the mean-to-variance ratio to be approximately 2, independent of level, i.e.,

$$r(L)/\mathrm{Var}(L) = 2, \qquad (7\text{-}38)$$

where $r(L)$ denotes the mean firing rate and Var(L) the variance of the firing rate elicited by a tone of level L. This pattern is compatible with a dead-time modified Poisson process plus a non-linear receptor function (Lachs, Al-Shaik, Saia, and Teich, 1984; Winslow and Sachs, 1988). Comparable findings were reported by Young and Barta (1986, Fig. 6), who measured responses to continuous noise and to 200-ms pure tones in noise near masked threshold. In their data, the mean-to variance ratio was approximately 2, except for very small spike counts. Javel and Viemeister (2000) observed an approximately constant mean-to variance ratio of 3 (50-ms sinusoids). Delgutte (1988, Fig. 6), on the other hand, found the variance of spike counts in response to 50-ms tone burst at the CF to be half the mean count only for mean spike counts smaller than 6. Above this region, the mean-to-variance ratio increased with spike count.

For modeling the performance of a single fiber with threshold L_{Th}, Viemeister assumed the decision variable to be the difference between the number of spikes elicited by standard-plus-increment and standard, respectively.

For reasonably large spike counts ($n > 30$), the Gaussian approximation can be used so that

$$d'(L_S, \Delta L) = \frac{T \cdot [r(L_s + \Delta L, L_{Th}) - r(L_s, L_{Th})]}{\sqrt{\mathrm{Var}(L_S, T) + \mathrm{Var}(L_s + \Delta L, T)}}, \qquad (7\text{-}39)$$

where L_S is standard level, ΔL the level increment, T tone duration, and $\mathrm{Var}(L, T)$ the variance of the spike count elicited by a tone of level L during the counting interval T. Note that the decision is assumed to be unbiased.

Using Eq. 7-38, the above equation simplifies to

$$d'(L_S, \Delta L) = \sqrt{2T}\frac{r(L_s + \Delta L, L_{Th}) - r(L_s, L_{Th})}{\sqrt{r(L_s + \Delta L, L_{Th}) + r(L_s, L_{Th})}}. \qquad (7\text{-}40)$$

The rate-level and variance-level functions of a single unit with threshold at 10 dB SPL are displayed in the left panel of Fig. 54.

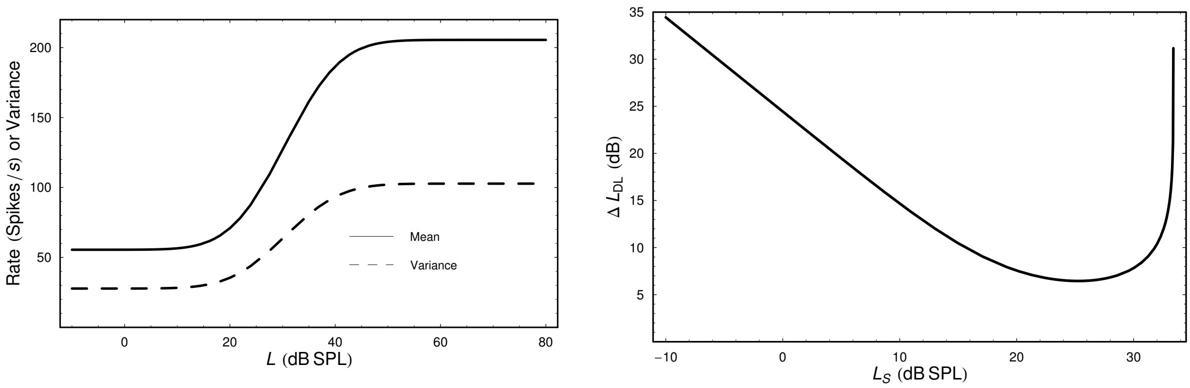

Fig. 54: Characteristics of a single auditory nerve fiber (threshold 10 dB SPL). Left panel: Mean firing rate and its variance as a function of tone level. Right panel: predicted just-detectable level increment (ΔL_{DL}, corresponding to 70.7% correct) for 30-ms pure tones ($T = 0.03$ s) as a function of standard level.

As seen in the right panel of Fig. 54, the fiber is predicted to optimally perform in only a narrow region between 20 and 30 dB SPL, with a slow increase in ΔL_{DL} towards lower levels and a steeper increase towards higher levels. It is obvious that a single fiber is not sufficient to explain the wide dynamic range of human intensity discrimination performance. Therefore, Viemeister (1988) modeled a population of nine fibers in the next step of his analysis. All were high-SR fibers with identical parameters except for threshold. Thresholds were separated by 10 dB, with the lowest threshold at 0 dB SPL and the highest threshold at 80 dB SPL (Fig. 55, left panel).

For this multiunit-analysis, the decision variable was the unweighted sum of the spike counts contributed by the single units. As the responses of all fibers are assumed to be independent, the variance of the decision variable is the sum of the variances and the distribution even more closely follows a normal distribution.

The right panel of Fig. 55 makes clear that the model essentially predicts the sum of spike counts to be a linear function of tone level (except at the extremes of the operating range) and variance of the spike count to be half the mean count. Between 20 and 100 dB SPL, the function can be closely approximated by

$$r_{Approx.}(L, T) = T \cdot (274.69 + 14.88\, L). \qquad (7\text{-}41)$$

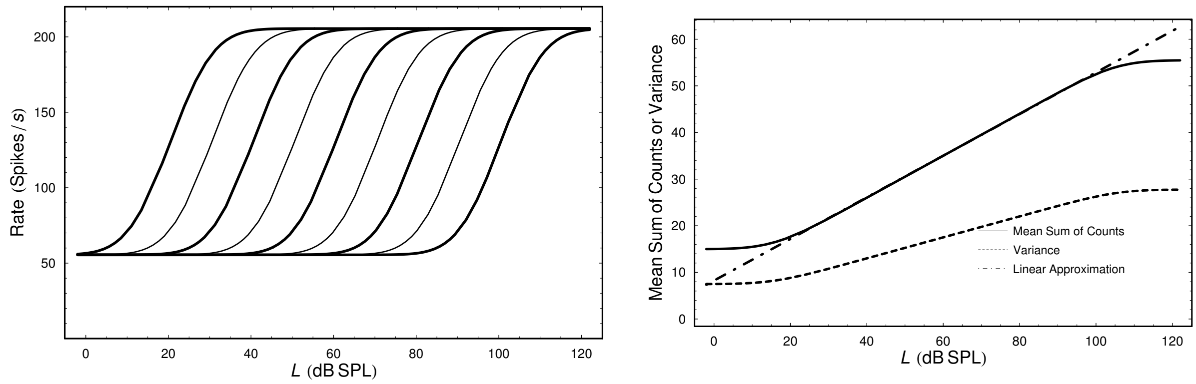

Fig. 55: Population of 9 units. Thresholds are distributed between 0 and 80 dB SPL in 10 dB steps. Left panel: rate-level functions. Right panel: predicted sum of spike counts in response to a 30-ms tone (solid line) and its variance (dashed line) as a function of level. The dot-dashed line shows a linear approximation to the sum of spike counts.

Predicted performance of the nine units is

$$d'_9(L_S, \Delta L) = \sqrt{2T}\,\frac{\sum_{i=1}^{9}\left(r_i(L_s+\Delta L) - r_i(L_s)\right)}{\sum_{i=1}^{9}\left(r_i(L_s+\Delta L) + r_i(L_s)\right)}, \qquad (7\text{-}42)$$

where $r_i(L)$ denotes the rate of fiber i in response to a tone of level L.

Between 20 and 80 dB, predicted performance roughly follows Weber's law (Fig. 56, left panel), although there is a slight increase in ΔL_{DL} with standard level. This pattern is compatible with human intensity discrimination in notched-noise (e.g., Viemeister, 1983), where effects of spread-of-excitation can be assumed to be absent. At each standard level, d' is approximately a linear function of ΔL (Fig. 56, right panel), so that predicted performance is in good accordance with the psychophysical data provided by Buus and Florentine (1991). In the simulations of Viemeister (1988), the use of an optimum decision rule (Green and Swets, 1966, p. 239), according to which performance resulting from n independent observations is $d'_n = \sqrt{\sum_{i=1}^{n}(d'_i)^2}$, had the effect of predicted performance showing a greater deviation from human psychophysical data.

Viemeister (1988) as well as Delgutte (1987), who presented a comparable model, emphasized that the assumed population of fibers in the multiunit model is critical. The simple "staggered threshold scheme" discussed here is certainly not compatible with the distribution of thresholds, spontaneous rates, dynamic ranges, slopes, etc. of auditory nerve fibers (Liberman, 1978; Evans and Palmer, 1980; Winslow and Sachs, 1988). Basing the model on a more realistic distribution resulted in a stronger deviation of model predictions from psychophysical data in the simulations of Viemeister (1988) and Delgutte (1987). Viemeister (1988, p. 233) concluded that "[…] virtually any Weber function can be described […] simply by choosing an appropriate set of rate-intensity functions." As discussed above, however, for the present purposes, the predicted relation between mean and variance of the neural counts at a given standard level is more important than the question whether Weber's law, the near-miss to Weber's law or a different Weber function is predicted. Although psychophysical and especially physiological data are limited, the existing empirical and theoretical results seem to support the assumption of a linear increase of spike count variance with mean spike count. If this proportionality holds for each group of fibers, *d'* computed for a sum of spike counts will also be approximately proportional to ΔL, independent of the actual neural population involved. Therefore, at each standard level, the dependence of performance on ΔL can be accounted for by the staggered-threshold model given that ΔL is not excessively large.

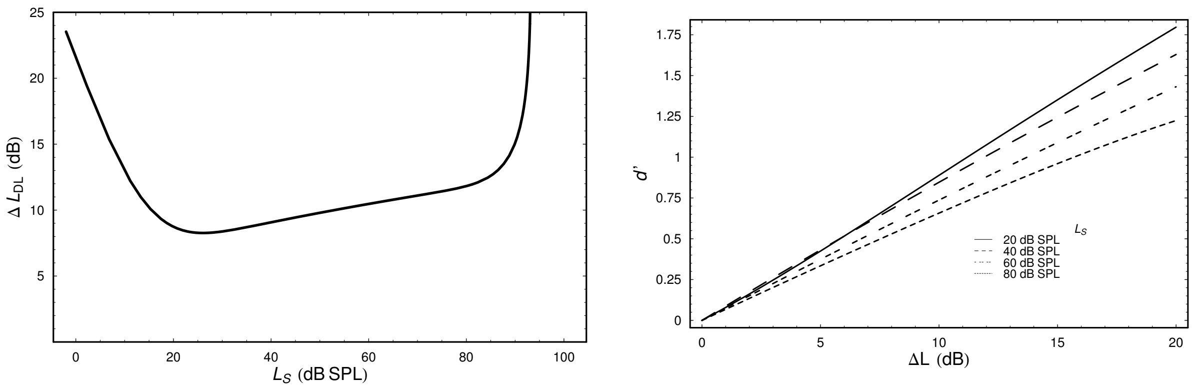

Fig. 56: Staggered-threshold model: population of nine units. Left panel: predicted level increment ΔL_{DL} corresponding to 70.7%-correct for 30-ms pure tones as a function of standard level. Right panel: predicted d' at different standard levels as a function of ΔL.

7.2.4 Model predictions

The predictions of the signal-detection model including mergence can now be evaluated for the three variants of masker weight combined with the two relations between physical sound pressure and internal continuum.

7.2.4.1 Masker weight fixed quantity, identical in both intervals

The simplest model results if the masker weights p_{M1} and p_{M2} effective in interval 1 and interval 2, respectively, are assumed to be identical and fixed quantities rather than random variables.

Note that in terms of the similarity model (Eq. 7-8), 'identical masker weight' could be defined as the same values of p_{Max} and the similarity parameter σ_{Sim} being effective in both intervals. As masker weight p_M depends not only on p_{Max} and σ_{Sim}, but also on the ratio between masker and target loudness, a separate masker weight could still be computed for both intervals in this case, so that $p_M(L_M, L_S)$ would be effective in the interval containing the standard and $p_M(L_M, L_{S+I})$ in the interval containing the standard-plus-increment. If the increment is small, however, a simplification can be used by letting $p_M(L_M, L_{S+I}) = p_M(L_M, L_T)$. In the following, the simplification is adopted, so that $p_{M1} = p_{M2}$.

Table 4 displays the resulting means and variances for intervals 1 and 2 that were determined using Table 3 and Eq. 7-26. Note that the standard 2I, 2AFC intensity discrimination task is modeled, where the same masker is presented in both intervals.

State of the world		Observation	
		Interval 1	Interval 2
<S, S + I>	Mean	$p_M\,\mu_M + (1 - p_M)\,\mu_S$	$p_M\,\mu_M + (1 - p_M)\,\mu_{S+I}$
	Variance	$p_M^2\,\sigma_M^2 + (1 - p_M)^2\,\sigma_S^2$	$p_M^2\,\sigma_M^2 + (1 - p_M)^2\,\sigma_{S+I}^2$
<S + I, S>	Mean	$p_M\,\mu_M + (1 - p_M)\,\mu_{S+I}$	$p_M\,\mu_M + (1 - p_M)\,\mu_S$
	Variance	$p_M^2\,\sigma_M^2 + (1 - p_M)^2\,\sigma_{S+I}^2$	$p_M^2\,\sigma_M^2 + (1 - p_M)^2\,\sigma_S^2$

Table 4: Equal-weight case ($p_{M1} = p_{M2}$): Means and variances of the observations made in interval 1 and interval 2.

We use Eq. 7-25 to compute P_{Merged}(Corr) in this situation:

$$P_{Merged}(\mathrm{Corr}) = \frac{1}{4}\cdot\left(2 + \mathrm{Erf}\left(\frac{C + (\mu_{S+I} - \mu_S)\,(1 - p_M)}{\sqrt{2}\sigma_{Merged}}\right) - \mathrm{Erf}\left(\frac{C - (\mu_{S+I} - \mu_S)\,(1 - p_M)}{\sqrt{2}\sigma_{Merged}}\right)\right), \qquad (7\text{-}43)$$

where σ_{Merged} is the standard deviation of the decision variable,

$$\sigma_{Merged} = \sqrt{\mathrm{Var}_{1Merged} + \mathrm{Var}_{2Merged}} = \sqrt{2\,p_M^2\,\sigma_M^2 + (1 - p_M)^2\,\sigma_{S+I}^2 + (1 - p_M)^2\,\sigma_S^2}. \qquad (7\text{-}44)$$

It can be shown that the optimum cutoff is still the same as without masker, namely $C = 0$. With this value of C, Eq. 7-43 simplifies to

$$P_{Merged}(\mathrm{Corr}|C = 0) = ½\left(1 + \mathrm{Erf}\left(\frac{(\mu_{S+I} - \mu_S)\,(1 - p_M)}{\sqrt{2}\sigma_{Merged}}\right)\right) \qquad (7\text{-}45)$$

How does the masker influence intensity discrimination according to Eq. 7-45? Note that no direct representation of masker level appears in the equation. The inclusion of masker information in the decision increases the standard deviation σ_{Merged} of the decision variable, however, because masker variance contributes to total variance. Above that, the difference $(\mu_{S+I} - \mu_S)$ is multiplied by a factor $(1 - p_M)$. Compared to the situation in quiet (Eq. 7-20), this results in P(Corr) being smaller for the same standard and standard-plus-increment. If p_M is computed according to the similarity model (Eq. 7-8), masker level has an effect on the weight assigned to masker information.

7.2.4.1.1 Equal-weight case: Durlach and Braida model

Given the preceding analysis, the effects of mergence can now be evaluated for the Durlach and Braida model.

For the sake of simplicity, no correction for the near-miss to Weber's law is made, so that the standard deviation of the decision variable is independent of level. From Eqs. 7-32 and 7-45 (unbiased choice) it follows that

$$P_{Merged}(\mathrm{Corr}) = ½\left(1 + \mathrm{Erf}\left(\frac{k\,\Delta L\,(1 - p_M)}{20\sqrt{1 - 2\,p_M\,(1 - p_M)}\,\sigma}\right)\right),\ 0 \le p_M < 1, \qquad (7\text{-}46)$$

where σ is the standard deviation of each observation.

Converting to *d'*, we write

$$d'_{Merged} = \frac{\sqrt{2}\, k\, \Delta L\, (1 - p_M)}{10\sqrt{1 - 2\, p_M\, (1 - p_M)}\, \sigma}. \qquad (7\text{-}47)$$

If we compare *d'* in quiet (Eq. 7-33) to d'_{Merged}, it is obvious that mergence results in sensitivity being smaller by a factor of $\frac{(1 - p_M)}{\sqrt{1 - 2\, p_M\, (1 - p_M)}}$. This factor monotonically decreases with p_M so that performance is maximal for $p_M = 0$. As the standard deviation σ is independent of level, masker intensity has no effect unless p_M depends on masker level, as assumed by the similarity model.

To illustrate the effect of mergence, the standard deviation was so chosen that the difference limen of 3.2 dB at 30 dB SPL reported by Zeng, Turner, and Relkin (1991; 1-kHz, 25-ms pure tones) is predicted in quiet. At such a low sound pressure level, the effects of spread of excitation should be negligible so that the observed jnd can be interpreted as the performance of a single auditory channel. The fixed parameters used in the analysis were $k = 10$, $\sigma = 4.15$ dB, and $L_S = 30$ dB SPL. The value ΔL_{DL} corresponding to $P(\text{Corr}) = 0.707$ (2-down, 1-up adaptive procedure) can be found by solving the equation

$$1/\sqrt{2} = \tfrac{1}{2}\left(1 + \text{Erf}\left(\frac{k\, \Delta L_{DL}\, (1 - p_M)}{20\sqrt{1 - 2\, p_M\, (1 - p_M)}\, \sigma}\right)\right) \qquad (7\text{-}48)$$

for ΔL_{DL}. This yields

$$\Delta L_{DL} = \frac{0.7707\sqrt{1 - 2\, p_M\, (1 - p_M)}\, \sigma}{(1 - p_M)} \qquad (7\text{-}49)$$

The jnd depends only on p_M, as σ is a constant. For small values of p_M, the jnd elevation is moderate. Fig. 57 (left panel) displays the predicted jnd increase with the weight p_M given to masker intensity.

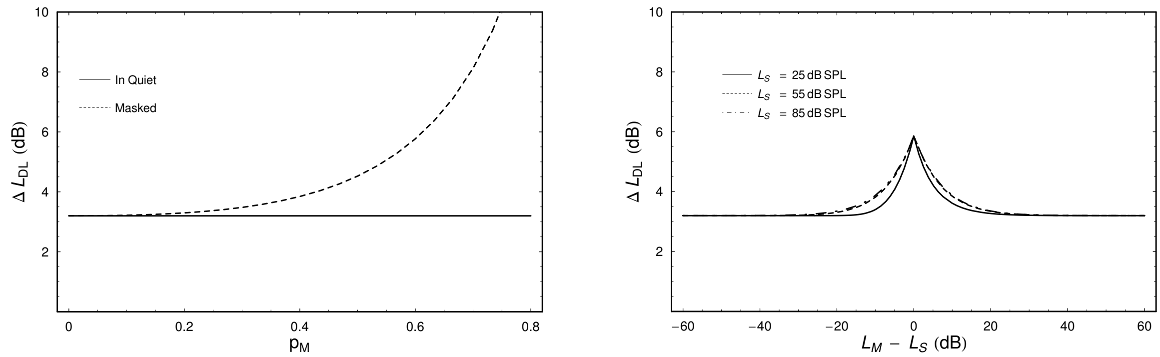

Fig. 57: Equal-weight case, Durlach and Braida model.
Left panel: Predicted jnd (ΔL_{DL} corresponding to 70.7% correct) for a 30-dB SPL standard. Solid line: ΔL_{DL} in quiet ($p_M = 0$). Dashed line: ΔL_{DL} as a function of masker weight p_M. Parameters: $k = 10$, $\sigma = 4.15$ dB, $C = 0$. For constant p_M, the predicted decrease in performance is independent of standard and masker level.
Right panel: ΔL_{DL} plotted as a function of the masker-standard level difference for three standard levels, with masker weight computed according to the similarity model. Parameters: $k = 10$, $\sigma = 4.15$ dB, $C = 0$, $p_{Max} = 0.606$, $\sigma_{Sim} = 0.708$, $\theta = 0.3$, $L_{ThM} = L_{ThT} = L_{ThC} = 6.59$ dB SPL, $K_M = K_T = K_C$.

In the next step, p_M was computed as a function of L_M and L_T according to the similarity model (Eq. 7-8). Masker weight is predicted to be small for an intense masker combined with a low-intensity standard. Fig. 57 (right panel) displays the resulting difference limens according to Eq. 7-49. In computation of p_M, the best-fitting parameters p_{Max} and σ_{Sim} as calculated for the mean data from Experiment 2 (Chapter 7.1.8.3) were used. Performance was simulated for a 55-dB SPL standard. As discussed above, the level dependence of the jnd in quiet is ignored in the present analyses; sensation variance σ^2 was selected to correspond to the difference limen at 30 dB SPL, just as in the left panel of Fig. 57. As the jnd elevation depends only on p_M, the deterioration in performance due to the masker is now predicted to be maximal at $L_M = L_S$, which is certainly incompatible with the data.

7.2.4.1.2 Equal-weight case: Staggered-threshold model

As demonstrated in the preceding chapter, the model based on the Durlach and Braida concept fails to account for the empirically observed jnd's under non-simultaneous masking. One characteristic of the model that is responsible for this failure is that given a masker weight p_M, the predicted jnd elevation is *independent* of masker level. The reason for this behavior is the assumed constant standard deviation of the decision variable. It seems more reasonable that, given a value of p_M, an intense masker influences performance more strongly than a weak masker. Such a pattern is predicted by the staggered-threshold model, where the sensation variance increases with level. Consequently, the jnd-elevation depends not only on p_M, but also on masker level (Fig. 58, upper left panel).

For a given p_M, the predicted jnd depends only on the variance contributed by the masker. Therefore, the jnd will be larger than the jnd in quiet and the jnd elevation increases with masker level (Fig. 58, upper right panel).

If now masker weight as predicted by the similarity model is used, the weight assigned to masker intensity will be maximal if $L_M = L_S$. As masker level increases above standard level, however, this will also increase the variance of the decision variable, resulting in a drop in performance. Consequently, the predicted jnd elevation is small for $L_M < L_S$ (small masker weight, small masker variance), larger for $L_M = L_S$ (maximal masker weight, moderate masker variance) and for $L_M > L_S$ (small masker weight, large masker variance), and small again for $L_M >> L_S$ (very small masker weight, large masker variance). At small values of the similarity parameter σ_{Sim}, the similarity effect dominates and the jnd has a sharp maximum at $L_M = L_S$, while with increasing σ_{Sim}, masker variance has a stronger effect (Fig. 58, lower panel).

The predictions are still incompatible with the data: even for σ_{Sim} as large as 10, ΔL_{DL} is maximum for equal masker and standard level. Recall, that σ_{Sim} estimated for the loudness-matching data fitted in Chapter 7.1.8 lay in the range between 0.5 and 2.6.

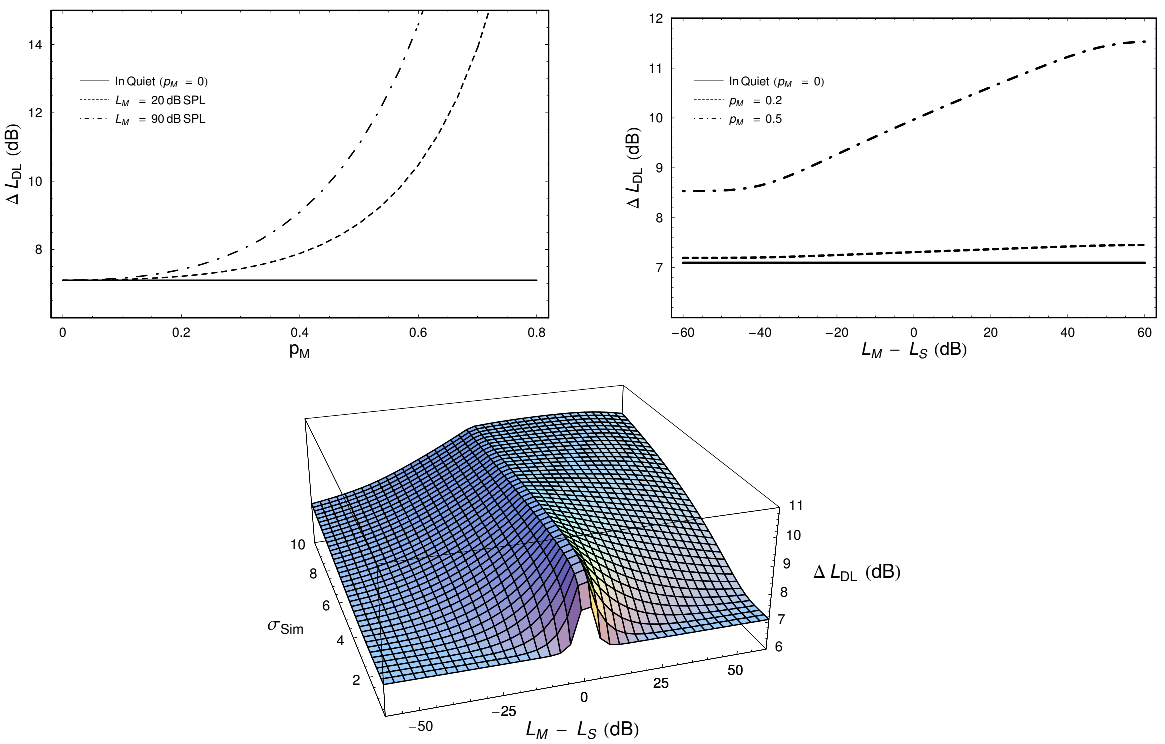

Fig. 58: Equal-weight case, staggered-threshold model.
Upper left panel: The jnd for a 55-dB SPL, 30-ms standard is plotted in quiet (solid line) and as a function of masker weight. Dashed line: 20-dB SPL masker. Dot-dashed line: 90-dB SPL masker. Unbiased choice (C = 0). Upper right panel: Difference limens for a 55-dB SPL, 30-ms standard as a function of the masker-standard level difference, plotted for two different masker weights. The solid line (p_M = 0) represents intensity discrimination in quiet. Unbiased choice (C = 0).
Lower panel: predicted ΔL_{DL} as a function of the masker-standard level difference and σ_{Sim}, with masker weight computed according to the similarity model. Parameters: L_s = 55 dB SPL, C = 0, p_{Max} = 0.5, θ = 0.3, $L_{ThM} = L_{ThT} = L_{ThC}$ = 6.59 dB SPL, $K_M = K_T = K_C$.

7.2.4.2 Masker weight fixed quantity, different weights effective in the two intervals

In this case, masker weights are still modeled as fixed quantities rather than random variables. It is now assumed that masker weight p_{M1} effective in interval 1 differs from masker weight p_{M2} effective in interval 2,

$$p_{M2} = m\, p_{M1}. \qquad (7\text{-}50)$$

The predicted means and variances of the observations made in interval 1 and interval 2 are displayed in Table 5.

		Observation	
State of the world		Interval 1	Interval 2
<S + I, S>	Mean	$p_{M1}\, \mu_M + (1 - p_{M1})\, \mu_{S+I}$	$m \cdot p_{M1}\, \mu_M + (1 - m \cdot p_{M1})\, \mu_S$
	Variance	$p_{M1}^2\, \sigma_M^2 + (1 - p_{M1})^2\, \sigma_{S+I}^2$	$(m \cdot p_{M1})^2\, \sigma_M^2 + (1 - m \cdot p_{M1})^2\, \sigma_S^2$
<S, S + I>	Mean	$p_{M1}\, \mu_M + (1 - p_{M1})\, \mu_S$	$m \cdot p_{M1}\, \mu_M + (1 - m \cdot p_{M1})\, \mu_{S+I}$
	Variance	$p_{M1}^2\, \sigma_M^2 + (1 - p_{M1})^2\, \sigma_S^2$	$(m \cdot p_{M1})^2\, \sigma_M^2 + (1 - m \cdot p_{M1})^2\, \sigma_{S+I}^2$

Table 5: Different-weights case ($p_{M2} = m \cdot p_{M1}$): Means and variances of the observations made in interval 1 and interval 2.

Using Eq. 7-25 and Table 5, P_{Merged}(Corr) can be computed.

To illustrate the effects of differing masker weights, we discuss the extreme case of zero weight assigned to masker intensity in the second interval (m = 0). The idea that the "veridical impression" of the target presented in interval 2 is compared to a somewhat altered or corrupted representation of the target presented in interval 1 is also promoted in the Durlach and Braida (1969) model, where the trace of the observation made in interval 1 is assumed to be corrupted by memory noise and to be compared to the 'direct' observation in interval 2, for which no memory noise is effective. Essentially the same is assumed in the quantitative theory of TOE by Michels and Helson (1954).

As to the decision criterion C, one might adopt the rather extreme view that the listener uses the same cutoff as in quiet, i.e., $C = 0$. In this case, as masker loudness has an effect in interval 1 only, the decision will be biased towards responding "Interval 1 louder" or "Interval 2 louder", depending on the masker-standard level difference. This idea was discussed in Chapter 6.3.4 as a potential explanation for the differences between ΔL_{DL} and jnd_{Match}.

A more reasonable alternative follows from the assumption that the listener adopts a cutoff point C so that there is no bias towards responding "Interval 1 louder" or "Interval 2 louder". I.e., as the listener knows that the increment is presented with equal probability in interval 1 or interval 2, he or she adjusts the criterion C so that the psychometric function passes through the 0.5 point at a 0-dB level difference between the target tone presented in interval 1 and the target tone presented in interval 2 ($L_{S+I} = L_S$). It can be shown that this optimum cutoff is just

$$C_{Unbiased} = (1 - m)\, p_{M1}\, (\mu_M - \mu_S), \qquad (7\text{-}51)$$

where masker weight p_{M1} is the masker weight effective for the standard eliciting a mean value of μ_S on the internal continuum combined with the masker eliciting a mean value of μ_M on the internal continuum. For $m = 0$ (masker receives no weight in interval 2), we have

$$C_{Unbiased} = p_{M1}\, (\mu_M - \mu_S). \qquad (7\text{-}52)$$

7.2.4.2.1 Different weights: Durlach and Braida model

What does the Durlach and Braida (1969) model predict for the case of $m = 0$?

Recall that in line with the preliminary theory, we assume the mean of the random variable representing the intensity of a tone to be

$$\mu = k \log_{10} I, \qquad (7\text{-}53)$$

and its standard deviation to be intensity independent.

If the same decision criterion as in quiet is assumed (C = 0), it follows from Eq. 7-25 and Table 5 that

$$P_{Merged}(\mathrm{Corr}|C = 0, m = 0) =$$
$$\frac{1}{4}\cdot\left(2 + \mathrm{Erf}\left(\frac{k\,[\Delta L - p_{M1}\,(L_M - L_S)]}{20\sqrt{1 + p_{M1}\,(p_{M1} - 1)}\,\sigma}\right) - \mathrm{Erf}\left(\frac{k\,[\Delta L\,(p_{M1} - 1) - p_{M1}\,(L_M - L_S)]}{20\sqrt{1 + p_{M1}\,(p_{M1} - 1)}\,\sigma}\right)\right). \qquad (7\text{-}54)$$

The predicted percentage of correct responses is still independent of standard level, but varies with p_{M1} and the masker-standard level difference. The effect of $L_M - L_S$ becomes clearer if we let k = 10 and $\sigma = 1/(20\sqrt{1 + p_{M1}\,(p_{M1} - 1)}$, so that the right hand side of Eq. 7-54 simplifies to

$$\frac{1}{4}\cdot\left(2 + \mathrm{Erf}(\Delta L - p_{M1}\,(L_M - L_S)) - \mathrm{Erf}(\Delta L\,(p_{M1} - 1) - p_{M1}\,(L_M - L_S))\right). \qquad (7\text{-}55)$$

The first Erf()-term represents trials with the increment in the second interval, the second Erf()-term represents the remaining trials. For constant p_{M1} and ΔL, Eq. 7-55 is maximum for

$$(L_M - L_S) = \Delta L/2,\ \Delta L > 0 \qquad (7\text{-}56)$$

For this reason, if p_{M1} is constant, performance is not predicted to be maximum at a masker-standard level difference of 0 dB but for masker level slightly exceeding standard level (Fig. 59, dot-dashed line).

If masker weight p_{M1} is computed according to Eq. 7-8 (similarity model), the predicted behavior is compatible with the observed data in that there is a jnd elevation at intermediate positive masker-standard level differences (Fig. 59, solid line). The substantial increase in the difference limen for intermediate *negative* masker-standard level differences is incompatible with the psychophysical data from Experiment 2, however, that demonstrate the effects of the masker to be small in this situation. We note that masker weight according to Eq. 7-8 decreases with the deviation of $L_M - L_S$ from 0 dB (Fig. 39). The point of maximum performance is therefore predicted to be slightly above $L_M - L_S = \Delta L/2$ because of the decrease of p_{M1}. If we now compare predicted performance for a masker-standard level difference of, e.g., 5 dB *above* and 5 dB *below* the point of maximum performance, it is obvious that p_{M1} will be much smaller in the former case. To give an example, if performance were maximum at $L_M - L_S$ = 5 dB, the maximum value of p_{M1} would be effective 5 dB below that point ($L_M - L_S$ = 0 dB), while p_{M1} would be rather small 5 dB above the point of maximum performance. For this reason, the model predicts the largest jnd elevation to occur at *negative* masker-standard level differences (Fig. 59, solid line), which is incompatible with the data.

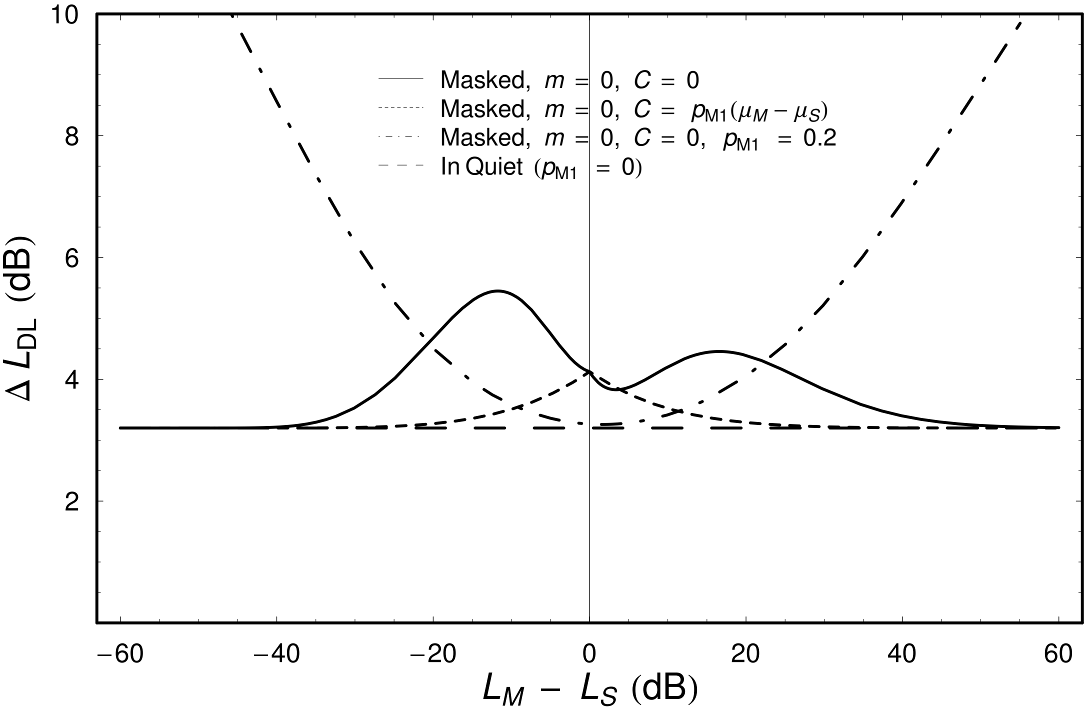

Fig. 59: Different-weights case, Durlach and Braida model: ΔL_{DL} for a 55-dB SPL standard plotted as a function of the masker-standard level difference, with masker weight computed according to the similarity model. Parameters: k = 10, σ = 4.15 dB, p_{Max} = 0.606, σ_{Sim} = 0.708, θ = 0.3, $L_{ThM} = L_{ThT} = L_{ThC}$ = 6.59 dB SPL, $K_M = K_T = K_C$. Solid line: m = 0, C = 0. Short-dashed line: m = 0, C = p_{M1} ($\mu_M - \mu_S$). For comparison, the dot-dashed line shows predicted ΔL_{DL} for the m = 0, C = 0 case, with fixed p_{M1} = 0.2. Long-dashed line: in quiet.

The left panel of Fig. 60 shows the resulting ΔL_{DL} in the C = 0 case as a function of the masker-standard level difference and the similarity parameter σ_{Sim}. At the value of σ_{Sim} computed for the data from Exp. 1 (0.708), ΔL_{DL} would be maximal for $L_M - L_T$ near 0.

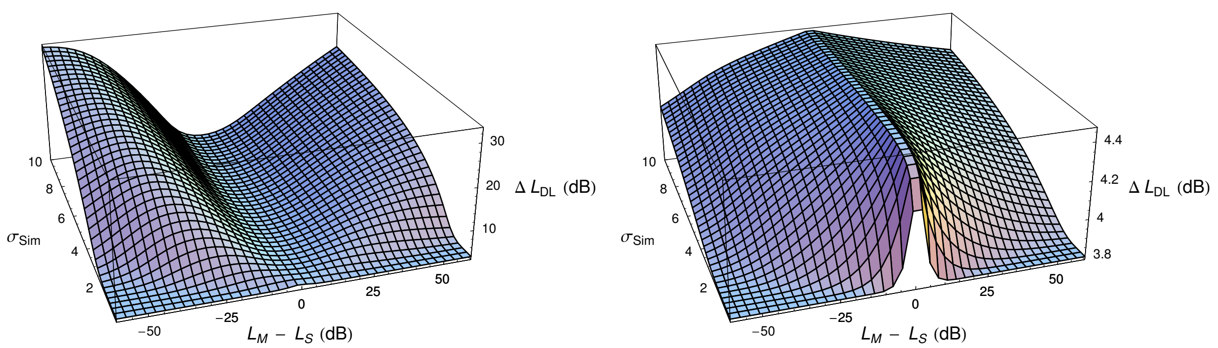

Fig. 60: Different-weights case, Durlach and Braida model. Predicted jnd (ΔL_{DL}) for a 55-dB SPL standard. Masker weight computed according to the similarity model. Left panel: m = 0, C = 0. Right panel: m = 0, C = p_{M1} ($\mu_M - \mu_S$). Fixed parameters: k = 10, σ = 4.15 dB, p_{Max} = 0.606, θ = 0.3, $L_{ThM} = L_{ThT} = L_{ThC}$ = 6.59 dB SPL, $K_M = K_T = K_C$.

If the 'unbiased' criterion $C = p_{M1}$ ($\mu_M - \mu_S$) is used, P_{Merged}(Corr| m = 0, $C = p_{M1}$ ($\mu_M - \mu_S$)) does not depend on L_M, but only on p_{M1}, so that performance is worst for maximum masker weight (right panel of Fig. 60). In other words, if p_{M1} is computed according to the similarity model, the strongest decrease in performance is again predicted to occur at $L_M = L_S$. Note that performance is predicted to be better than for the equal-weight case, as masker variance from interval 2 does not contribute to total variance if m = 0.

7.2.4.2.2 Different weights: Staggered-threshold model

As seen in Fig. 61, if the decision variable is modeled according to the staggered-threshold model, the effect of mergence on intensity discrimination performance in the different-weights case is quite similar to performance predicted using the Durlach and Braida model.

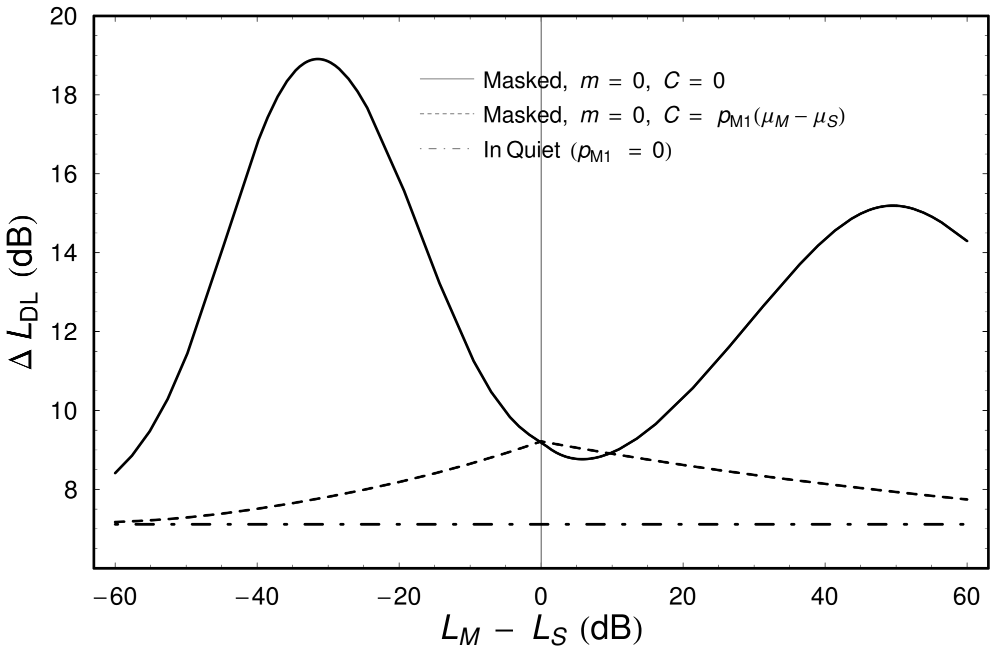

Fig. 61: Different-weights case, staggered-threshold model: ΔL_{DL} for a 55-dB SPL standard plotted as a function of the masker-standard level difference. Masker weight computed according to the similarity model. Parameters: p_{Max} = 0.606, σ_{Sim} = 2.0, θ = 0.3, $L_{ThM} = L_{ThT} = L_{ThC}$ = 6.59 dB SPL, $K_M = K_T = K_C$. Solid line: m = 0, C = 0. Dashed line: m = 0, $C = p_{M1}$ ($\mu_M - \mu_S$). Dot-dashed line: in quiet.

For C = 0, the smallest difference limen is expected for masker level slightly exceeding standard level and the effect of the masker is again greater for negative than for positive masker-standard level differences (solid line in Fig. 61).

Note that the value of σ_{Sim} used in the simulation was 2.0, which is greater than the best-fitting value found for the data from Experiment 2 (0.708). As Fig. 62 (left panel) shows, ΔL_{DL} would be maximal for $L_M - L_T$ near 0 at such a small value of σ_{Sim}.

For $C = p_{M1}$ ($\mu_M - \mu_T$), the difference limen (dashed line in Fig. 61) depends on $L_M - L_S$ in the same way as for the equal-weight case, with the maximum jnd elevation expected for equal masker and target level, especially at small values of σ_{Sim} (Fig. 62, right panel). Performance is predicted to be better as in the equal-weight case, however, as masker variance from interval 2 does not contribute to total variance if $m = 0$.

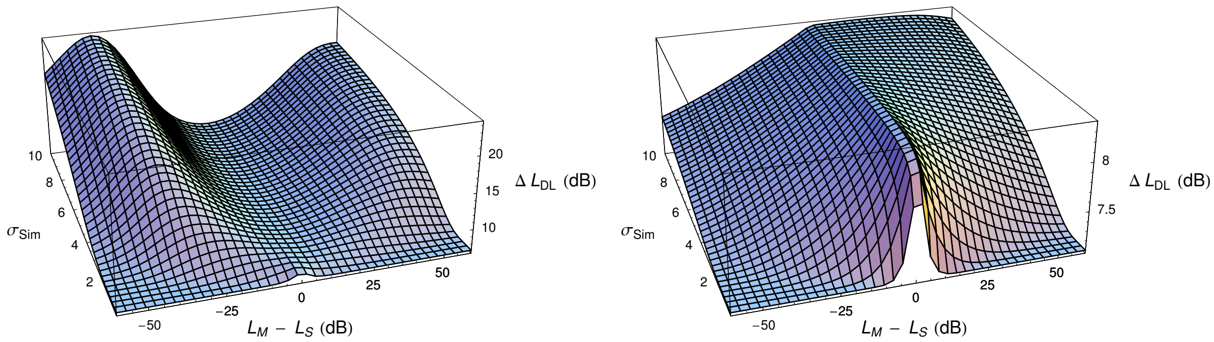

Fig. 62: Different-weights case, staggered-threshold model, with masker weight computed according to the similarity model: predicted jnd (ΔL_{DL}) for a 55-dB SPL standard. Left panel: $m = 0$, $C = 0$. Right panel: $= 0$, $C = p_{M1}(\mu_M - \mu_S)$. Fixed parameters: $p_{Max} = 0.606$, $\theta = 0.3$, $L_{ThM} = L_{ThT} = L_{ThC} =$ 6.59 dB SPL, $K_M = K_T = K_C$.

To summarize, the predictions are not compatible with the psychophysical data demonstrating mid-difference humps but also only very small effects of non-simultaneous masking for masker levels below standard level.

7.2.4.3 Masker weight random variable

In this chapter, the weight assigned to masker loudness is modeled as a random variable. In the preceding analyses, masker weight was assumed to be constant in all trials presenting the same combination of masker and target tone. It seems more reasonable that from trial to trial, a listener will be more or less successful in ignoring the masker, so that *mean* weight assigned to masker loudness is still a function of target-masker similarity, but the value of p_M effective in a given trial will be subject to random fluctuations. The variance associated with masker weight can also be used to model an increase in variance caused by the very process of mergence. In the preceding analyses, the computation of a weighted average was assumed to be lossless, i.e., the arithmetic operations introduced no additional noise. For such operations performed by neural systems, it seems reasonable to assume the process of merging to be noisy. The resulting variance can be included into the model in terms of the variance associated with masker weight.

For the sake of simplicity, it is assumed that the mean of the random variable representing masker weight is the same in both intervals ($p_{M1} = p_{M2} = p_M$, cf. Chapter 7.2.4.1). As mean masker weight is identical for both intervals, the optimum (unbiased) cutoff of $C = 0$ is assumed. Above that, the standard deviation of the random variable representing masker weight is modeled to be constant, i.e., independent of the mean.

From Table 3 and Eqs. 7-25 and 7-28, it follows that

$$P_{Merged}(\mathrm{Corr}|C = 0) = \tfrac{1}{2}\left(1 + \mathrm{Erf}\left(\frac{(\mu_{S+I} - \mu_S)\,(1 - p_M)}{\sqrt{2\,\mathrm{Var}_{Merged}(\text{random masker weight})}}\right)\right), \qquad (7\text{-}57)$$

where p_M denotes the mean of masker weight and Var_{Merged}(random masker weight) is the variance of the decision variable,

$$\mathrm{Var}_{Merged}(\text{random masker weight}) = \sigma_{pM}^2(\mu_S^2 + \mu_{S+I}^2 + 2\mu_M^2) + \sigma_S^2\,[(p_M - 1)^2 + \sigma_{pM}^2] + \sigma_{S+I}^2[(p_M - 1)^2 + \sigma_{pM}^2] + 2\,\sigma_M^2\,(p_M^2 + \sigma_{pM}^2), \qquad (7\text{-}58)$$

where σ_{pM} denotes the standard deviation of masker weight. If one compares the above equation to $P_{Merged}(\mathrm{Corr}|C = 0)$ for the case of masker weight being a fixed quantity and identical in both intervals (Eq. 7-45), it is obvious that the numerator remains the same, but that the variance of masker weight contributes to the total variance. Equally important, variance now depends not only on the variability of the observations (σ_M, σ_I, σ_{S+I}), but also on their expected values (first term on the right hand side of Eq. 7-58).

7.2.4.3.1 Random masker weight: Durlach and Braida model

As the square of the expected value of the random variable representing the sensory effect of stimulus intensity contributes to variance according to Eq. 7-58, the increase in variance effected by a masker corresponding to, e.g., $\mu_M = -20$ is equal to the increase in variance effected by a masker corresponding to $\mu_M = +20$. If now the relation between sound pressure level and the expected value on the internal continuum was now again assumed to be $\mu = L$ (as in the above analyses based on the preliminary theory of intensity resolution), the model would predict a large jnd increase for *negative* masker sound pressure levels, which is of course unreasonable. As the unit of the scale used for the internal continuum (Eq. 7-30) is arbitrary, this problem can be resolved by writing

$$\mu = k \log_{10}(b \cdot I/I_0), \qquad (7\text{-}59)$$

where sound intensity $I = 10^{L/10}\, I_0$. The factor b/I_0 appearing in the argument to log() can be used to make sure that μ is positive for the smallest sound pressure level used in the simulation. In Fig. 63, predicted performance for a 55-dB SPL standard is plotted. As minimum masker level was 55 dB SPL − 60 dB SPL = − 5 dB SPL, b was set to $\sqrt{10}$, so that the value of μ corresponding to a sound pressure level of −5 dB SPL was 0.

With random masker weight, mergence is predicted to have an effect on the difference limen even for mean masker weight of zero (Fig. 63, long-dashed line). This behavior is due to the contribution of masker-weight variance, masker variability and the sensory effect of the masker to the variance of the decision variable $x_1 - x_2$ even for $p_M = 0$,

$$\mathrm{Var}_{Merged}(\text{random masker weight}|\, p_M = 0) = \sigma_{pM}^2(\mu_S^2 + \mu_{S+I}^2 + 2\mu_M^2) + \sigma_S^2\,[1 + \sigma_{pM}^2] + \sigma_{S+I}^2[1 + \sigma_{pM}^2] + 2\,\sigma_M^2\,\sigma_{pM}. \qquad (7\text{-}60)$$

If mean masker weight p_M is computed according to the similarity model, the jnd elevation is again predicted to be maximum at $L_M = L_T$ (Fig. 63, short-dashed line).

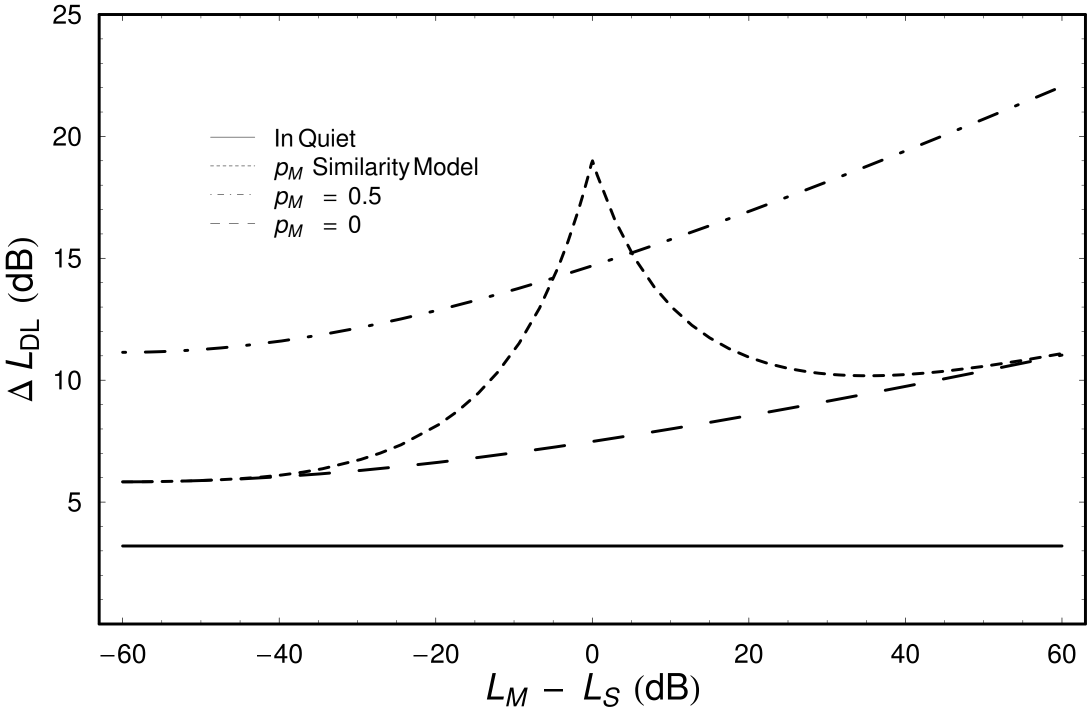

Fig. 63: Durlach and Braida model, random masker weight: ΔL_{DL} for a 55-dB SPL, 30-ms standard plotted as a function of the masker-standard level difference. Parameters: $C = 0$, $\sigma_{pM} = 0.1$, $b = \sqrt{10}$, $\sigma = 4.15$. Solid line: in quiet. Short-dashed line: p_M according to the similarity model (Eq. 7-8), $p_{Max} = 0.606$, $\sigma_{Sim} = 2.0$, $\theta = 0.3$, $L_{ThM} = L_{ThT} = L_{ThC} = 6.59$ dB SPL, $K_M = K_T = K_C$. Dot-dashed line: $p_M = 0.5$. Long-dashed line: $p_M = 0$.

7.2.4.3.2 Random masker weight: Staggered-threshold model

If the decision variable is modeled according to the staggered-threshold model, essentially the same relation between the masker-target level difference and ΔL_{DL} as for the Durlach and Braida model discussed in the preceding chapter is predicted (Fig. 64).

Taken together, it has to be concluded that modeling masker weight as a random variable does not result in the mergence model correctly predicting the small effects at masker levels smaller than standard level and the mid-difference hump in intensity discrimination.

Now it could be argued that this failure is due to the assumption of constant masker weight variance σ^2_{pM}. If σ_{pM} was assumed to increase with mean masker weight p_M, however, the peak of ΔL_{DL} predicted at $L_M = L_S$ would be even more pronounced, as p_M is maximal for identical masker and target loudness according to the similarity model.

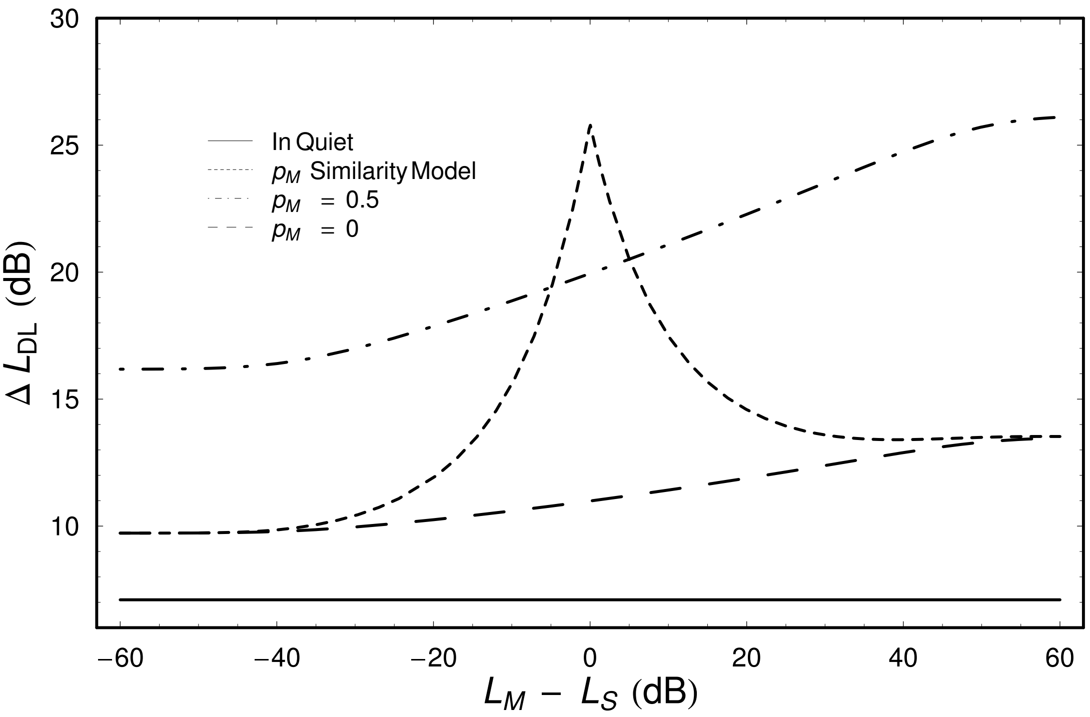

Fig. 64: Staggered-threshold model, random masker weight: ΔL_{DL} for a 55-dB SPL, 30-ms standard plotted as a function of the masker-standard level difference. Parameters: $C = 0$, $\sigma_{pM} = 0.1$. Solid line: in quiet. Short-dashed line: p_M according to Eq. 7-8, $p_{Max} = 0.606$, $\sigma_{Sim} = 2.0$, $\theta = 0.3$, $L_{ThM} = L_{ThT} = L_{ThC} = 6.59$ dB SPL, $K_M = K_T = K_C$. Dot-dashed line: $p_M = 0.5$. Long-dashed line: $p_M = 0$.

7.2.5 Discussion

As discussed in Chapter 4.1.3, the effects of the alteration of loudness caused by a non-simultaneous masker on the intensity jnd can not be accounted for by existing models of the relation between intensity resolution and loudness.

In this chapter, the idea proposed in the similarity model was pursued that the inclusion of masker intensity information is the common mechanism underlying both loudness enhancement and the jnd elevations observed und non-simultaneous masking. A formal signal-detection model was used to predict the increase in ΔL_{DL} caused by mergence, i.e., the inclusion of masker information in the decision in an intensity discrimination experiment. The loudness enhancement hypothesis (Carlyon and Beveridge, 1993) promotes the seemingly simple idea that loudness enhancement increases loudness variability and therefore the jnd elevation is monotonically related to the amount of loudness enhancement. Yet, this relation could not be derived from the model. Neither the contribution of masker sensation noise to the variance of the decision variable, nor the explicit inclusion of noise introduced by the process of mergence resulted in the predictions of the model being at least qualitatively compatible with the data. If mean and variance of the random variables representing tone intensity were modeled according to the staggered-threshold model by Viemeister (1988), the SDT model predicted an increase of ΔL_{DL} with masker intensity for constant weight p_M assigned to masker intensity. The empirically observed mid-difference hump, i.e., the decrease in ΔL_{DL} at large masker-standard level differences, was not predicted, however. For loudness enhancement, this effect was correctly predicted by the similarity model, where p_M is a function of the similarity in loudness between masker and target. Yet, using p_M computed according to the similarity model (Eq. 7-8) resulted in ΔL_{DL} predicted to be maximum at masker levels close to target level.

The only condition in which the SDT model predicted mid-difference humps was the case of a bias introduced by the inclusion of masker information. This pattern resulted for different masker weights effective in the two observation intervals (e.g., $p_{M2} = 0$) combined with a cutoff parameter not compensating for the masker-induced systematic difference between the observations made in interval 1 and interval 2 in this situation. Aside from the fact that a strong response bias seems unlikely in a 2I, 2AFC task with trial-by-trial feedback (Green and Swets, 1966), the model also predicted the mid-difference jnd elevation to be larger at *negative* than at positive masker-standard level differences. This pattern is incompatible with the psychophysical data (Fig. 24), which show jnd's close to the values in quiet for masker levels smaller than standard level.

It has to be concluded that is surprisingly difficult to formulate a formal model accounting for the simple idea of mergence causing a drop in performance.

8 Interval-produced accent: Model and experiments

In Chapter 4.3, it was concluded that some theoretical considerations speak against loudness enhancement as the cause of the accent. The 'interrupted processing' model proposed by Povel and Okkerman (1981), on the other hand, seems capable to account for the effects of temporal structure on the interval-produced accent reported by the same authors (Chapter 4.3.4).

Due to the seriously limited data basis, the empirical evidence for or against the two models is weak, however.

Specifically, the temporal limit of the accent proposed by Povel and Okkerman (1981) is incompatible with the loudness enhancement data, but the estimate is likely to be biased due to a problem in Povel and Okkerman's experimental design. In this chapter, the effect of sequence structure is studied in a preliminary experiment and Experiment 4, which are designed to provide a more reliable estimate of the time constant effective for the interval-produced accent. A quantitative model based on the 'interrupted processing' hypothesis is used as a mathematical tool for estimating the time constant. This model also makes predictions concerning an important sequence parameter that has not been discussed up to now, namely the starting condition or initial position.

Experiment 5 studies the effects of presentation level. A midlevel hump in accent perception would present strong evidence for a relation between the interval-produced accent and loudness enhancement, while presentation level should have no effect according to the interrupted processing model. To simplify the comparison between the results from loudness enhancement experiments and the accent perception data, a modern loudness matching task is used.

8.1 Quantitative model based on the interrupted-processing hypothesis

As discussed in Chapter 4.3.4, the 'interrupted processing' model proposed by Povel and Okkerman (1981) seems to be able to qualitatively account for the effects of temporal structure on the interval-produced accent reported by the same authors. It is also sufficiently specific to allow the formulation of a quantitative model of the relation between temporal sequence structure and the loudness difference between first and second tones short-long equitone sequence. As it will be demonstrated below, the predicted relation between temporal structure and the observed accent is also compatible with loudness enhancement as the cause of the accent. The model is used as a mathematical tool for the estimation of the time constant effective for the accent.

8.1.1 Time course of processing

The interrupted processing model assumes growth of loudness to be a function of processing in auditory sensory memory (Cowan, 1987; cf. Chapter 4.3.4). What is known about the time course of this processing? In Massaro's (1970) experiments on backward recognition masking, identification performance was demonstrated to be a monotonically increasing function of the intertone interval. The gradient of this function also decreased with intertone interval and performance reached an asymptotic value at approximately 300 ms.

Massaro (1975) proposed that the dependence of identification performance on processing time t can be modeled as

$$d'(t) = \alpha(1 - e^{-\theta t}), \tag{8-1}$$

where d' is the SDT sensitivity index, α is the asymptotic value of d' effective for t approaching infinity, and θ is a parameter determining the rate of growth. Regarding information extraction, which is the relevant variable in identification tasks, Massaro and Loftus (1996) pointed out that such function represents a process "[...] that resolves some fixed proportion of the potential information that remains to be resolved per unit of time" (p. 75). With respect to growth of loudness, Cowan (1987) presented evidence for the idea that loudness N also depends on the degree of processing in auditory sensory memory, so that the same relation between processing time and loudness should be effective.

Should processing time be expressed as the offset-onset interval (ISI) or as the inter-onset interval? It seems unlikely that loudness of a very long tone (e.g., 10 s) should in any way depend on the temporal position of a following tone, so that the silent interval between tone offset and onset of the following stimulus can not be the variable relevant for the interrupted processing model. Similarly, Massaro and Loftus (1996) argued that the presence of an inverse duration effect in auditory persistence (the duration of shorter sounds was overestimated relative to the duration of longer sounds; Efron, 1970) can not be explained by assuming that "[...] auditory sensory store is simply a fixed appendage tacked onto the end of an auditory stimulus." (p. 78). The authors concluded that processing time should be defined as presentation time plus the silent interval, i.e., as the inter-onset interval (IOI = sound duration plus ISI). Data by Massaro (1974) supported this view. Identification of a test vowel was measured in an ABRM paradigm as a function of the interval between vowel onset and mask onset. In one condition, the vowel was only 26 ms long, while it was continuous in the other condition. The time course of discriminability was similar in the two conditions.

For growth of loudness, however, it seems questionable whether sound duration should be assumed to have no effect in the sense that presentation of the 'interrupting sound' (masker) after a given IOI results in the same reduction in loudness independent of tone duration. It seems likely that loudness grows faster during stimulus presentation, where acoustic energy is continually input into the auditory system (Zwislocki, 1969), than after stimulus termination. Indirect evidence for this hypothesis is the fact that in experiments measuring temporal loudness summation, loudness of a tone burst reaches an asymptote at a tone duration of 100 ms (e.g., Port, 1963; see Zwislocki, 1969, for a review), which represents a considerably shorter time constant than the 250 ms assumed by Povel and Okkerman (1981) for the interval-produced accent. A simple way to include the effects of both tone duration d and the inter-onset-interval into the model is to express loudness of a sound x as

$$N(\mathrm{ISI}_x, d) = N_{Max}(d)\,[1 - r_0(d)\, e^{-s\, ISIx}],\; 0 \le r_0(d) < 1, \tag{8-2}$$

where $\mathrm{ISI}_x = \mathrm{IOI}_x - d$ denotes the silent interval between the sound and the masker, N_{Max} denotes maximum perceived intensity that will be reached for ISI_x approaching infinity, and s is the decay parameter.

The parameter $r_0(d)$ can be given a simple interpretation: it denotes the reduction in loudness relative to the maximum value N_{Max} that is obtained if ISI = 0, i.e., if the second sound immediately follows the first sound.

If we denote by N_0 loudness at ISI = 0, it follows that

$$r_0(d) = 1 - \frac{N_0(d)}{N_{Max}(d)},\ 0 \le N(d) \le N_{Max}. \qquad (8\text{-}3)$$

The parameter N_0 depends on tone duration d. Note that N_{Max} also varies with tone duration due to temporal loudness summation.

If it is assumed that growth of sensation during stimulus presentation follows exactly the same time course as after stimulus termination, an equivalent formulation can be derived using the function from Eq. 8-1 and expressing loudness of a sound x as

$$\mathrm{N}^*(\mathrm{IOI}_x) = N_{Max}\,[1 - e^{-s\,IOIx}]. \qquad (8\text{-}4)$$

Now, Eq. 8-2 predicts the same growth of sensation as a function of the inter-onset interval as Eq. 8-4 if the same value of N_{Max} is used and r_0 is set to $1 - N^*(d)/N_{Max}$.

Eq. 8-2 is equivalent to the function used by Harris and Dallos (1979) to model the recovery of auditory nerve responses. Median probe response magnitude R (relative to the value obtained without masker) was expressed as

$$\mathrm{R}(\Delta T) = 1 - r_d\, e^{-\Delta T/\gamma} \qquad (8\text{-}5)$$

where ΔT is the silent interval between forward masker and probe stimulus, r_d is the relative reduction in probe response magnitude at $\Delta T = 0$, and γ is the time constant of recovery.

Zwislocki's (1969) model of temporal summation predicts essentially the same functional relation between loudness N of the sound burst and sound duration t. Zwislocki assumes that a stimulus causes a peripheral neural response which is largest at stimulus onset and equilibrates at approximately 100 ms after onset. The neural activity is integrated across time; recent activity receives more weight, and loudness is directly proportional to integrated spike count. For burst durations larger than 5 ms, the function fitting data from several loudness summation experiments (Eq. 13 in Zwislocki, 1969) can be approximated by

$$N(t) = N_{Max}\,(1 - 0.614\, e^{-25\,t}), \qquad (8\text{-}6)$$

with sound duration t expressed in seconds.

Generally, the class of exponential growth functions proved useful for fitting data in a wide variety of cases. Examples include amplitude growth of neuromagnetic evoked responses as a function of ISI in isochronous sequences (Sams et al., 1993), amplitude of MEG responses to the first sound of a tone pair as a function of ISI (McEvoy et al., 1997), and performance curves in visual digit recall experiments (Loftus and Ruthruff, 1994). In the present context, it is especially important to note that the decrease in forward loudness enhancement with the conditioner-target interval (Zwislocki and Sokolich, 1974) can also be modeled by a decaying exponential at least if one ignores the non-monotonicity reported for some subjects at ISIs smaller than 20 ms.

Therefore, if we express loudness of a tone in a short-long sequence as a function of the silent interval *preceding* it, we can write

$$N_{\mathrm{LEH}}(ISI) = N_{\infty} + (N_{\mathrm{Max}} - N_{\infty})\, e^{-s\, ISI}, \qquad (8\text{-}7)$$

where N_{∞} is loudness 'in quiet' (ISI = ∞) and N_{Max} is maximum loudness effective at ISI = 0, where loudness enhancement is assumed to be maximum.

In the short-long sequences, the accent is determined by the difference in loudness of between the second and the first sounds ($N_2 - N_1$). For the interrupted-processing approach, Eq. 8-2 predicts this difference to be

$$N_2 - N_1 = N(ISI_L, d) - N(ISI_S, d) = N_{\mathrm{Max}}\, r_0\, (e^{-s\, ISI_S} - e^{-s\, ISI_L}). \qquad (8\text{-}8)$$

For the loudness enhancement approach (Eq. 8-10), we find

$$N_2 - N_1 = N_{\mathrm{LEH}}(ISI_S) - N_{\mathrm{LEH}}(ISI_L) = (N_{\mathrm{Max}} - N_{\infty})\, (e^{-s\, ISI_S} - e^{-s\, ISI_L}). \qquad (8\text{-}9)$$

Both approaches predict the accent to be proportional to the difference ($e^{-s\, ISI_S} - e^{-s\, ISI_L}$). Therefore, fitting the assumed function to the data obtained in experiments measuring the interval-produced accent provides a mathematical tool for estimation of the time constant of the process underlying the interval-produced accent, regardless of whether interrupted processing or loudness enhancement is assumed to be the cause of the accent. In the following, Eq. 8-2 will be used, so that --for the sake of simplicity-- the quantitative model is termed the "interrupted processing model".

Again, it must be emphasized that for short-long equitone sequences, it is not possible to infer from an accent reported on the second tone of a pair, whether the second tone reduced loudness of the first tone (interrupted processing), or the first tone caused loudness of the second tone to be enhanced. To decide between the models, it is crucial whether the estimated time course is compatible with loudness enhancement, or interrupted processing in the short auditory store, or both.

Concerning the latter question, Cowan (1984, 1987) concluded from his review of experiments indexing auditory sensory memory that the short auditory store decays within about 300 ms. Is this value compatible with the results reported for the interval-produced accent? In Povel and Okkerman's (1981) experiments, long interval duration was 340 ms in one condition. The proportion of accents reported on the second sounds dropped to 50% at short interval duration of roughly 250 ms (Fig. 8). Thus, one might conclude that the results could indeed be explained by interrupted processing during storage in *short* auditory memory. The data do not appropriately test the range of intervals within which accent perception due to interrupted processing occurs, however. The difference between long and short interval duration linearly *decreased* with short interval duration in the sequences used. For this reason, there are two alternative explanations for the absence of the relevant accents when short intervals were longer than 250 ms. It might have indeed been due to the fact that all intervals exceeded the duration of auditory sensory memory, such that all sounds were processed completely. The alternative explanation, however, is that at a short interval duration of 250 ms, the difference between long and short interval duration was too small to generate the effect. Even if processing had still been incomplete for both first and second sounds, the difference in degree of processing between the first and second tones would have been negligible due to the very small interval differences, resulting in almost identical loudness for the two sounds.

The second condition in Povel and Okkerman's study where short interval duration was fixed to 50 ms is easier to interpret. The proportion of accents reported on the second sounds reached its maximum value at a long interval duration of roughly 250 ms. In terms of the model, this can be taken as evidence for processing of the second sounds being complete after 250 ms, so that any further increase in the long ISI did not produce any change in the perceived intensity of the second sounds. Possibly, though, the estimate of maximum ISI duration at which an accent is perceived might be underestimated due to ceiling effects, i.e., the interval differences might have been too large in this condition. A superior test of the temporal limit would be to vary both the intra-pair and the inter-pair interval over a wider range while maintaining moderate differences.

If one assumes the accent to be a phenomenon emerging in rhythmic sequences only, the temporal limits of rhythm perception should be effective for the accent. From the review of Fraisse (1982), it can be concluded that sequences are observed as rhythmic rather than as a succession of single events only if the longest IOI is smaller than about 1600 ms.

Based on the assumed "growth-of-sensation", a quantitative model relating the temporal parameters of the short-long sequences to loudness of the first and second sounds can be formulated. For fitting the experimental data, a functional relation between magnitude of sensation (Eq. 8-2) and the dependent variable used in the respective experiment will be introduced. In the next chapter, the effects of an additional sequence parameter will be discussed and incorporated into the model.

8.1.2 Effects of the initial position

At first sight one might assume that the temporal structure of an equitone sequence that is constructed as an alternating succession of short and long IOIs is unambiguously defined by the parameters IOI_S, IOI_L, and sound duration. It is important to realize, however, that this is true only for the stationary case, where the sequence is presented continually. Because each sequence must start at some moment in time, an additional parameter is necessary to completely specify the sequence. This parameter is the starting condition or *initial position*.

Initial position has frequently been reported to have a strong influence on the perceived organization of rhythms (Fraisse, 1982; Longuet-Higgins and Lee, 1982; Preusser, Garner, and Gottwald, 1970). Nevertheless, in most studies initial position was treated as a nuisance parameter, an effect which researchers tried to minimize by starting sequences at random temporal points (Parncutt, 1994), at a subliminal intensity level which was slowly increased (Vos, 1977), or at an initially high presentation rate which was continually decreased (e.g., Boker and Kubovy, 1998). The present study recognizes initial position as an integral aspect of each rhythm and therefore systematically assesses its influence.

What can be expected regarding the influence of initial position on intensity perception? This question is related to a potential explanation for the accents perceived on the first tones of the pairs in the experiments by Povel and Okkerman (1981). At small absolute IOI differences (20 ms and 30 ms), subjects reported such accents in 70 - 90% of all trials. Certainly, neither interrupted processing, nor loudness enhancement of the second tone by the first tone can account for perception of the *first* sounds as more intense. For example, even if the effect of interrupted processing was negligible (as can be assumed at very small IOI differences), this would not explain why the relative frequency of accents perceived on first sounds was considerably above 50%. To account for the effect, Povel and Okkerman assumed that there is an orienting response to the first tone of a sequence, which causes the perception of an accent on this sound "[...] and, by induction, on all other first sounds of the periodically recurring groups" (p. 568). Because *all* sequences used in experiments 1, 3, and 4 of

Povel and Okkerman's study began with a complete pair, this assumption could indeed explain the perception of the first sounds as louder than the second sounds.

Results from Experiment 2 of Povel and Okkerman's study demonstrated a clear effect of initial position. An additional 500-ms tone, presented before the first tone pair of the sequence and with an ISI equal to the long ISI of the sequence, caused an increase in the proportion of accents perceived on the second tones. In the condition with the long ISI fixed to 340 ms, the function relating short interval duration to the proportion of accents reported on the second sounds appeared to be shifted to the right (Fig. 65) for all subjects. In the condition with the short ISI fixed to 50 ms, subject 1 produced very high proportions of accents reported on the second sounds (80-100 %) at all long-interval durations (70-220 ms). The remaining two subjects showed nearly no effect of initial position. Due to subject 1, there was an effect in the average data, however.

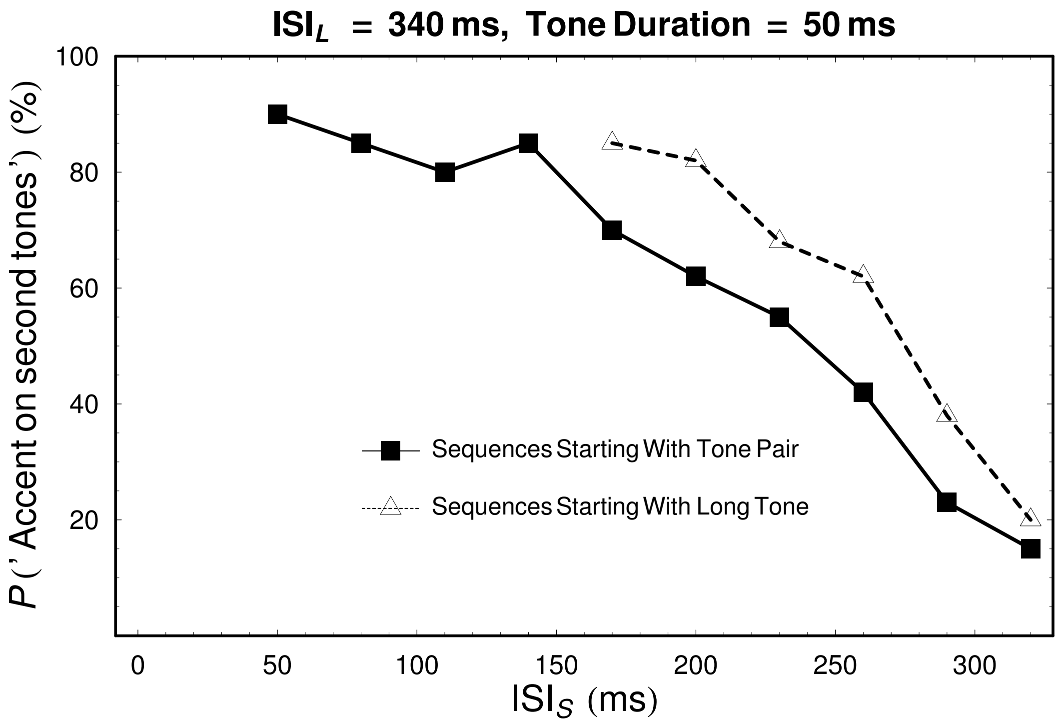

Fig. 65: Effect of the starting condition. Mean proportions of trials within which the three listeners reported an accent on the second tones of the tone pairs in the experiments of Povel and Okkerman (1981) plotted as a function of the short ISI. Long ISI was fixed to 340 ms. Squares: Sequence starting with tone pair. Triangles: Sequence starting with a 500-ms tone, followed by an ISI of 340 ms and then the first tone pair.

In the quantitative model proposed here, it is assumed that initial position introduces a bias in favor of sounds belonging to the same category as the very first sound of the sequence (e.g., first sounds of the pairs). The model predicts that for a given pair of intervals, IOI_S and IOI_L, first sounds will be perceived as more intense if the sequence began with a first sound than if the sequence began with a second sound.

In the preliminary experiment and in Experiment 4, initial position was varied systematically. The first sound was either the first or second sound of a pair (a sound initiating a short IOI or a sound initiating a long IOI, respectively).

8.1.3 Model Specification

It is now possible to specify the functional relation between the sequence parameters (IOI_S, IOI_L, sound duration, and initial position) and the perceived intensities of first and second sounds. The following assumptions are made:

1. There are exactly two categories of sounds (first sounds and second sounds, respectively).
2. Two variables, X_1 and X_2, represent loudness of the first sounds and the second sounds, respectively. Recurring to the concept underlying Thurstone's law of comparative judgment (cf. Luce and Galanter, 1963, pp. 214ff), the perceived intensities of both categories of sounds (X_1 and X_2) elicited by a specific stimulus are assumed to be random variables rather than fixed quantities. In the model, the expected values of the variables X_1 and X_2 depend on the time available for processing the respective sound category, on the initial position of the sequence, and on sound duration.
3. The random variables X_1 and X_2 are normally distributed with means μ_1 and μ_2, identical standard deviations $\sigma_1 = \sigma_2 = \sigma$, and covariance $\mathrm{cov}(X_1, X_2) = 0$ (Thurstone's Case *V*, cf. Luce and Galanter, 1963, p.54).
4. If $X_2 > X_1 + C$ (loudness of the second sounds exceeds loudness of the first sounds by at least an amount C), the listener will respond that the second sounds were louder than the first sounds. Otherwise, he or she will respond that the first sounds were more intense.

According to these assumptions, the probability for responding "Second sounds accented" follows Thurstone's equation of comparative judgment that relates the response probability to the distribution parameters by

$$p(\text{“Second Sounds Louder”}) = p(X_2 - X_1 > C) = \int_{x=C}^{\infty} \mathrm{PDF}(N_{\mu_2 - \mu_1, \sqrt{2}\sigma}, x)\, dx, \qquad (8\text{-}10)$$

where $\mathrm{PDF}(N_{\mu,\sigma}, x)$ denotes the probability density function of the normal distribution with mean μ and standard deviation σ evaluated at the point x. The cutoff (parameter C) represents response bias (Luce and Galanter, 1963, p. 225*f*). If C is zero, choice is unbiased.

Technically speaking, the integral in Eq. 8-10 is an *S*-shaped function mapping the real numbers (loudness differences) onto the interval [0, 1] (response probabilities).

For a given stimulus waveform and presentation level, the means of the two distributions are assumed to depend on processing time, initial position and sound duration only. The effect of initial position is additive in the model. If the sequence begins with a complete pair, loudness of the first sounds is assumed to be elevated by a positive amount $N_{Max} \cdot N_{InitPos}$. If the sequence begins with a sound initiating a long IOI (i.e., the first sound of the first pair is omitted), loudness of the second sounds is elevated by the same amount.

The means of the variables representing loudness of the first sounds and second sounds, respectively, are given by

$$\mu_1 = \begin{cases} N(\mathrm{ISI}_S, d) + N_{Max}\, N_{InitPos} & \text{if } \textit{Initial Position} = \text{First Sound} \\ N(\mathrm{ISI}_S, d) & \text{else,} \end{cases} \quad (8\text{-}11)$$

and

$$\mu_2 = \begin{cases} N(\mathrm{ISI}_L, d) + N_{Max}\, N_{InitPos} & \text{if } \textit{Initial Position} = \text{Second Sound} \\ N(\mathrm{ISI}_L, d) & \text{else,} \end{cases} \quad (8\text{-}12)$$

where $N(\mathrm{ISI}, d)$ denotes loudness according to Eq. 8-2.

8.1.4 Sequence parameters

The temporal structure of the short-long sequences is uniquely defined by the values of the short and the long IOI (IOI_S and IOI_L), sound duration d, and the category of the first and last sound of the sequence.

From the time course of growth of sensation assumed in the model (monotonically increasing function with decreasing gradient) it follows that given a constant *absolute* difference between long and short IOIs, perception of the second sounds as more intense should vary with the duration of the short IOI. For a short interval of , e.g., 200 ms, a long interval of 400 ms should cause a larger difference between the perceived intensities of first sounds and second sounds than if the interval pair were IOI_S = 2000 ms and IOI_L = 2200 ms, respectively. For this reason, in the following experiments, *relative* IOI difference

$$\Delta_{IOI} = \frac{IOI_L - IOI_S}{IOI_S} \quad (8\text{-}13)$$

will be varied independently of the period

$$T = \mathrm{IOI}_S + \mathrm{IOI}_L = \mathrm{IOI}_S\,(2 + \Delta_{IOI}), \quad (8\text{-}14)$$

which is inversely related to the tempo of the sequence. If we use, e.g., $\Delta_{IOI} = 0.6$, the long IOI is constructed by making the short interval 60% longer.

Fig. 66 displays a schematic representation of the stimuli. Each combination of Δ_{IOI} and T uniquely determines a pair of short and long IOIs.

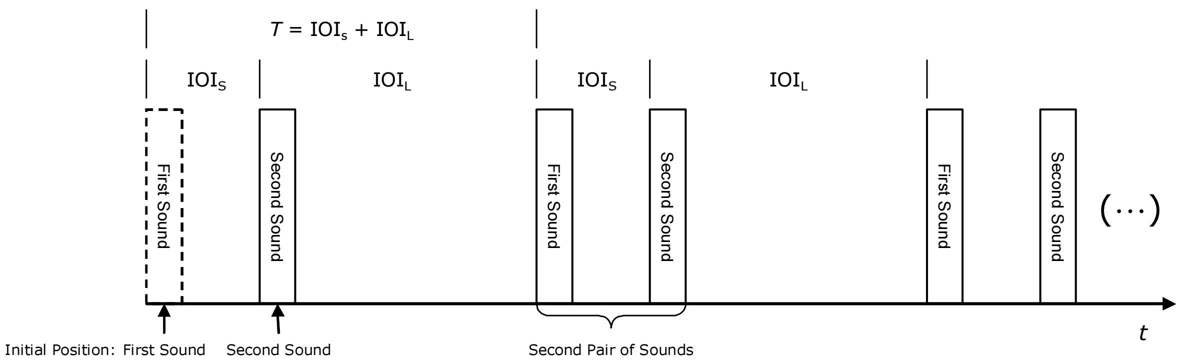

Fig. 66: The temporal sequence parameters T (period) and Δ_{IOI} (relative difference between IOI_S and IOI_L) were varied. Each sequence started either with the first or with the second sound of a pair (Initial Position).

The use of relative IOI difference as the relevant parameter is also supported by the fact that some of the sequences constructed as an alternation of short and long IOIs are similar to rhythmic structures frequently found in western music (e.g., duple rhythm). Specifying the stimulus parameters in terms of *relative* IOI difference and tempo rather than in terms of short-interval duration and *absolute* IOI difference seems more appropriate against this background, because the former is the representation used in western musical notation. Vos (1977), in accordance with these ideas, concluded that “[...] the tempo of the sequence [...] is a codeterminant of the perceptual organization” (p. 190).

8.2 Preliminary experiment

In Povel and Okkerman's (1981) experiments, listeners were required to respond whether they perceived the first or the second element of perceptual groups of two sounds as more accented. Although the authors did not report any problems with this procedure, listeners can be expected to have had difficulties differentiating between the first and second elements at fast tempi and small differences between long and short IOIs, as the smallest relative IOI differences were about 10%. This value is close to the just noticeable difference of intervals for the interval durations used (Schulze, 1989; Hirsh, Monahan, Grant, and Singh, 1990; ten Hoopen et al., 1994).

To avoid such problems, a preliminary experiment was conducted to test the appropriateness of an alternative procedure. In the experiment, subjects marked the position of the louder (accented) sounds by tapping in synchrony with them. The dependent variable was the proportion of taps marking second sounds as louder within a trial,

$$P_{Tap}(\text{Second Sounds}) = \frac{\text{Number of taps marking second sounds}}{\text{Total number of taps}} . \qquad (8\text{-}15)$$

In the analysis, this variable is used as an estimate of p("Second Sounds Louder"), the probability of the second sounds being perceived as more intense than the first sounds, which is modeled by Eqs. 8-10 to 8-12.

8.2.1 Hypotheses

Loudness of a sound is expected to monotonically increase with available processing time (8-2), or, in terms of loudness enhancement, to be inversely proportional to the inter-pair interval (Eq. 8-7).

In terms of interrupted processing, one possible definition of the time constant of growth of loudness is the ISI between a tone and the next tone at which loudness has grown to 90% of maximum loudness N_{Max}. According to Eq. 8-2, the value of s corresponding to such time constant depends on r_0, however, which is unknown and also depends on sound duration. As argued above, loudness can be expected to grow faster during than after stimulus presentation. Therefore, assuming the same time course of growth of sensation during and after stimulus presentation results in a somewhat underestimated growth function, but Eq. 8-4 can be used in this case, which does not depend on tone duration. For illustrating the effects of different values of s on growth of sensation, we define the time constant as the IOI between the test tone and the next tone at which loudness has grown to 90% of maximum loudness N_{Max}. If this definition is used, the value of the growth parameter s corresponding to loudness reaching 90% of its maximum value after an IOI of 250 ms + 50 ms, as it is assumed by Povel and Okkerman, 1981, is 7.68.

The model also predicts a systematic effect of initial position.

For data analysis in terms of an analysis of variance, the following qualitative hypotheses can be derived:

If the relative IOI difference Δ_{IOI} is greater than 0 it follows from Eqs. 8-2, 8-11, 8-12, and 8-13 that

$$IOI_S < IOI_L \Rightarrow \mu_1 < \mu_2. \qquad (8\text{-}16)$$

The difference ($\mu_2 - \mu_1$) between the means of the variables representing loudness of the second and the first tones, respectively, determines p(„Second Sounds Louder") (Eq. 8-10). For $IOI_S < IOI_L$, the model predicts the second sounds to be perceived as louder than the first sounds. A listener will indicate this perception by more often tapping in synchrony with the second than with the first sounds. At any given period T, a larger Δ_{IOI} results in a larger absolute IOI difference,

$$\Delta_{IOI1} < \Delta_{IOI2} \Rightarrow IOI_{L|\Delta IOI1} - IOI_{S|\Delta IOI1} < IOI_{L|\Delta IOI2} - IOI_{S|\Delta IOI2}. \qquad (8\text{-}17)$$

This effect of Δ_{IOI} on μ_1 and μ_2 leads to the prediction of an increase of P_{Tap}(Second Sounds) with Δ_{IOI} (main effect Δ_{IOI}).

Given a constant value of the difference between long and short IOI, the model predicts the difference between loudness of the first and second tones to decrease with period T. In the experiments, relative IOI difference (Δ_{IOI}) was varied, however. The model predicts an interaction between Δ_{IOI} and period: the effect of Δ_{IOI} on the dependent variable P_{Tap}(Second Sounds) is expected to be most pronounced at intermediate periods. The period T resulting in the strongest effect depends on the growth parameter s in Eq. 8-2.

Because depending on initial position, $N_{InitPos} \cdot N_{Max}$ is added to either μ_1 or μ_2, the model predicts a main effect of this stimulus parameter.

8.2.2 Method

8.2.2.1 Stimuli

The stimuli were equitone sequences constructed as an alternation of long and short IOIs. There were six levels of relative IOI difference Δ_{IOI} (0.077, 0.167, 0.400, 0.750, 1.333, and 2.500). These values are more densely spaced at small IOI differences where the growth of sensation function is assumed to be steepest. Three levels of the period T (400 ms, 800 ms, and 1600 ms) and two levels of *Initial Position* (first sound, or second sound, i.e., sound initiating a long IOI) were used. Each parameter combination was presented once to each subject, resulting in a completely crossed and balanced three-way factorial within-subjects design. Each sequence was presented for 60 seconds.

In order to maintain the attention of the listeners, the drum-module produced five different sampled percussive sounds (e.g., rim shot, sticks), with durations between 21 and 24 ms. In each trial, only one sound was presented; sounds were randomized across trials. Presentation level for each sound was adjusted so that loudness was approximately equal for all sounds. As no sound proof chamber was available at the time the experiment was conducted, listeners were allowed to adjust presentation level to a comfortable value in the range between 70 and 90 dBA at the beginning of the session. Ambient noise level was approximately 39.5 dBA.

8.2.2.2 Apparatus

The sounds were generated by a drum-module (Yamaha RY 30), which also served as the input device. A software MIDI-sequencer (Cubase VST 3.6), running on a Pentium-PC, controlled stimulus generation via MIDI and recorded the tapping. Subjects tapped on a rubber pad located on top of the drum module. The tapping produced no electro acoustical signal, but the mechanical sound of the tap was audible. The stimuli were presented via Sennheiser HD 25 closed dynamic headphones directly connected to the drum module.

8.2.2.3 Listeners

There were 20 participants, 14 of whom were undergraduate psychology students of the University of Technology Berlin and received partial course credit. The remaining subjects were volunteers. Participants ranged in age from 19 to 63 (mean age = 29.3 years); all were naïve with respect to the aim of the experiment. With the exception of one subject suffering from a tinnitus (but receiving no medication), all reported normal hearing. In order to avoid problems in the data analysis due to very variable tapping, only three subjects without any practical musical training took part. All other subjects had played their instrument for one to 50 years (mean = 16 years) and had practiced on a regular basis during the last 1.5 years. They had received between one and 20 years of musical instruction. In the analysis, it became evident that two subjects had tapped more complex rhythms instead of tapping in synchrony with accented elements in a substantial number of trials. Their data were discarded, resulting in $N = 18$.

8.2.2.4 Procedure

Subjects sat in front of the drum module positioned on a table. At the beginning of each session, they were instructed to tap with their middle or index finger. In order to acquaint the subjects with the device, the experimenter introduced the tapping device and asked subjects to do some self-paced tapping while the response pad produced a sound which would also be used in the stimuli. Participants were then informed that they were about to listen to simple rhythms which might contain accented elements. The latter were defined as sounds that are perceived as particularly intense. Subjects were instructed to tap in synchrony with the more intense sounds. It was emphasized that they were not required to detect any objective structure, but that their subjective perception was of interest. They were asked to tap only if they perceived louder sounds. They were also allowed to change the position of their tapping during a trial in case they detected a discrepancy between perceived accentuation and their tapping. Vos (1977) reported that accent perception in the type of stimuli used here can be ambiguous; therefore percepts were expected to change during some trials. Additionally, it was anticipated also that participants might fail to tap in synchrony with the louder elements, especially at fast tempi. In such cases they should be able to correct their responses.

At the beginning, practice trials were given. Subsequently, each of the 6 × 3 × 2 (Δ_{IOI} × Period × Initial Position) sequences was presented once, presentation order was randomized. In order to minimize carry-over effects, there was a 3-*s* pause after each sequence, followed by a rhythm consisting of random IOIs. This rhythm lasted 4 seconds, was presented in a markedly different timbre (a bass sound), and was followed by another pause of 3 seconds. Each subject participated in a one-hour session.

8.2.2.5 Data analysis

The dependent variable P_{Tap}(Second Sounds) was the proportion of taps marking sounds initiating long IOIs as accented (Eq. 8-15). A tap was counted as marking a sound if the temporal distance between the tap and the onset of the sound was smaller than $\text{IOI}_S/2$. If a tap fell into such a symmetric temporal interval centered at, e.g., the onset of a first sound of a pair, the tap was counted as marking a first sound. Taps falling neither into a respective interval around a first or a second sound were not counted. For each trial, from the numbers of taps marking first and second sounds, respectively, P_{Tap}(Second Sounds) was computed. If within a trial there were no taps at all, P_{Tap}(Second Sounds) was set to ½, because subjects had been instructed not to tap when perceiving no sounds as louder. To avoid missing values in the data set, P_{Tap}(Second Sounds) was also set to ½ for trials within which none of the taps could be counted as marking a sound. A value of P_{Tap}(Second Sounds) = ½ indicates that the perceived intensities of first and second sounds did not differ. P_{Tap}(Second Sounds) = 0 represents a sequence for which the second sounds were never marked as louder, but only first sounds were marked as more intense. If P_{Tap}(Second Sounds) = 1, only second sounds were marked as louder within that trial.

8.2.3 Results

To test the effects of relative IOI difference, period and initial position on the proportion of taps marking second sounds as accented, P_{Tap}(Second Sounds), a ΔIOI × T × Initial Position repeated-measures ANOVA with Greenhouse-Geisser degrees-of-freedom correction was conducted.

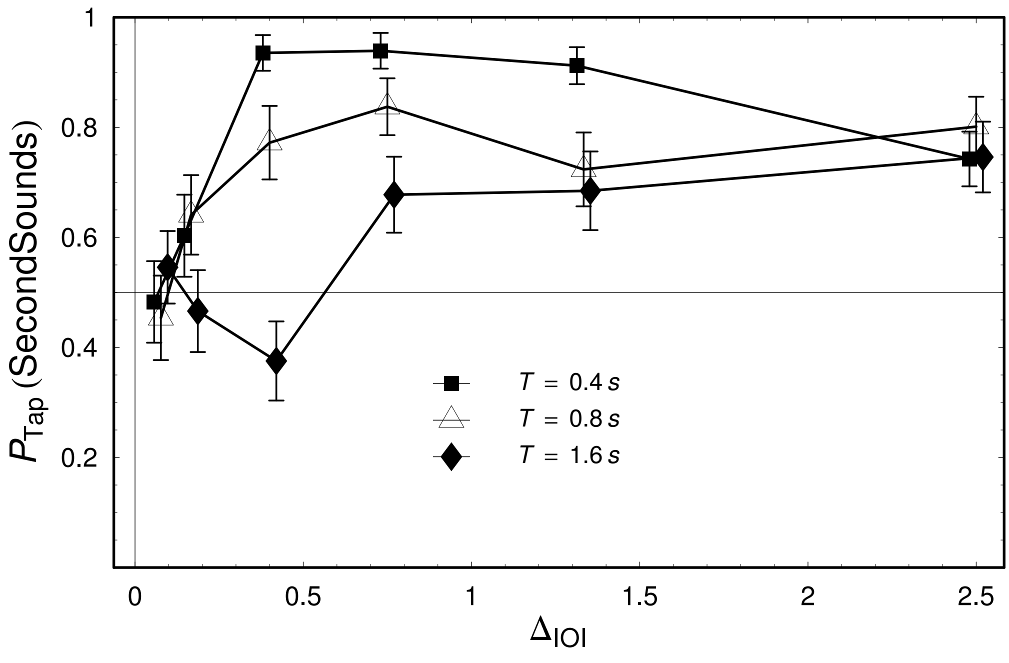

Fig. 67: Proportion of taps marking second sounds as accented as a function of the relative difference between long and short IOI, Δ_{IOI}. Squares: T = 0.4 s. Triangles: T = 0.8 s. Diamonds: T = 1.6 s. To avoid cluttering, the lines are shifted by 0.02 along the x-axis. Error bars show ± 1 SEM.

As expected, P_{Tap}(Second Sounds) was near 0.5 at the smallest value of Δ_{IOI} and increased with relative IOI difference (Fig. 67). This main effect of Δ_{IOI} was significant, $F(5, 85) = 16.29$, Greenhouse-Geisser $\hat{\varepsilon} = 0.725$, $p < 0.001$. The deviations from monotonicity present at each period, as for example the decrease in P_{Tap}(Second Sounds) at the largest Δ_{IOI} at $T = 0.4$ are incompatible with the model.

Mean values of P_{Tap}(Second Sounds) were significantly smaller at slower tempi (T = 0.4 s: M = 0.769, SD = 0.360; T = 0.8 s: M = 0.705, SD = 0.409; T = 1.6 s: M = 0.583, SD = 0.422), $F(2, 34) = 3.06$, $\hat{\varepsilon} = 0.957$, $p < 0.003$.

As Fig. 67, indicates, a significant Δ_{IOI} × period interaction was present in the data, $F(10, 170) = 5.42$, $\hat{\varepsilon} = 0.621$, $p < 0.001$. Conform to the predictions of the model, the effect of Δ_{IOI} was smaller if the short IOI was already rather long (T = 1.6 s).

A separate ΔIOI × Initial Position ANOVA conducted for the data at T = 1.6 s indicated a significant effect of ΔIOI, $F(5, 85) = 5.59$, $\hat{\varepsilon} = 0.812$, $p < 0.002$. This is evidence for the accent being present at intra-pair intervals longer than 270 ms, as all IOIs were greater than 355 ms in this condition.

As expected, there was a significant effect of initial position: if sequences started with a sound initiating a long IOI, a larger value of P_{Tap}(Second Sounds) was observed (M = 0.801, SD = 0.336) than if the sequences started with a sound initiating a short IOI (M = 0.571, SD = 0.442), $F(1, 17) = 70.35$, $p < 0.001$ (Fig. 68). At the fastest tempo (T = 0.4 s), the effect of Initial Position was very large at the smallest values of ΔIOI. The observed pattern resembles the data produced by one listener in Povel and Okkerman (1981), who also reported an accent on the second tones in virtually all trials if the sequence began with an additional tone presented before the first pair. Occasionally, the proportion of second sounds marked as louder was identical or even greater in the latter condition.

The Δ_{IOI} × Initial Position interaction was also significant, $F(5, 85) = 8.74$, $\hat{\varepsilon} = 0.640$, $p < 0.001$.

There was a significant Δ_{IOI} × Period × Initial Position interaction, $F(10, 170) = 3.611$, $\hat{\varepsilon} = 0.567$, $p < 0.003$. Inspection of the data suggests that this effect was present mainly because at the fastest tempo (period = 400 ms), there was a larger difference due to initial position at small relative IOI differences ($\Delta_{IOI} = 0.077$ and $\Delta_{IOI} = 0.167$) than at the slower tempi.

The Period × Initial Position interaction was not significant.

8.2.4 Discussion

In the preliminary experiment, a tapping response mode was used to avoid potential problems with a task similar to the one used by Povel and Okkerman (1981). Results were compatible with the expected effects of relative interval difference, period, and initial position. The effect of ΔIOI at T = 1.6 s is evidence for the accent being present at intra-pair intervals longer than 350 ms.

There were also some unexpected effects. Specifically, regarding the very large effect of initial position, one could argue that subjects might have tried to synchronize as fast as possible with the sequence rather than concentrating on the position of perceived accents. Such strategy would have resulted in preferred synchronization with the category of sounds presented first and might have contributed to the rather strong effect of initial position in some of the sequences.

Moreover, even though most subjects had practice in playing an instrument, they reported difficulty in timing the taps precisely enough to mark the targeted sounds at fast tempi. Such problems might have be the cause of the decrease in P_{Tap}(Second Sounds) observed at the largest value of ΔIOI, i.e., for the shortest intra-pair intervals.

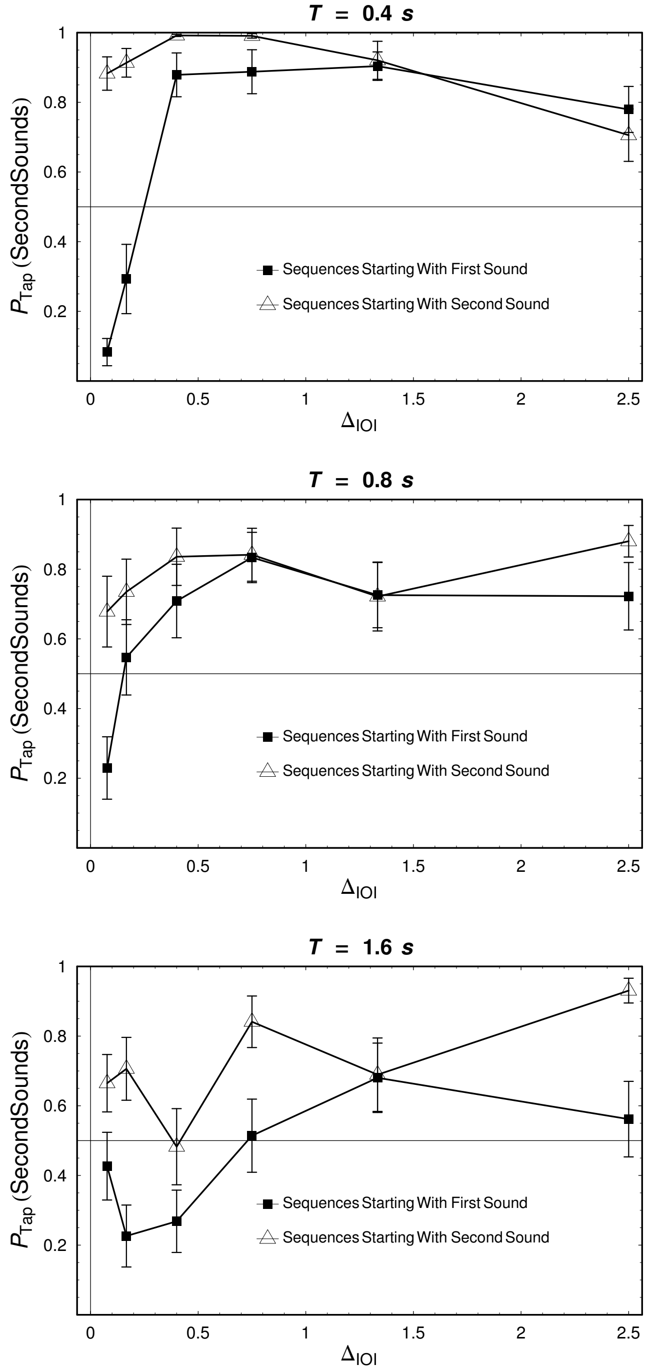

Fig. 68: Proportions of taps marking second sounds, P_{Tap}(Second Sounds), as a function of Δ_{IOI} and period T. Squares: Initial Position = first sound. Triangles: Initial Position = second sound. Error bars show ± 1 SEM.

8.3 Experiment 4: Effects of sequence structure on the interval-produced accent

To avoid the problems with the tapping procedure used in the preliminary experiment, a different response mode was used in Experiment 4.

8.3.1 Method

8.3.1.1 Stimuli

A similar temporal structure as in the preliminary experiment was used. Again, there were three periods *T* (500 ms, 1000 ms, and 2000 ms) and two initial positions (first sound or second sound of a sound pair). Four values of relative IOI difference (Δ_{IOI}) were presented. To avoid structures common in western music (e.g., duple rhythm), values representing complex interval ratios were generated. The four values of Δ_{IOI} were 0.1185, 0.2510, 0.5651, and 1.4495. These values are more densely spaced at small IOI differences.

T		Δ_{IOI}			
		0.1185	0.2510	0.5651	1.4495
500 ms	IOI_S	236.01 ms	222.12 ms	194.92 ms	144.95 ms
	IOI_L	263.98 ms	277.88 ms	305.08 ms	355.05 ms
1000 ms	IOI_S	472.03 ms	444.27 ms	389.85 ms	289.90 ms
	IOI_L	527.97 ms	555.75 ms	610.15 ms	710.10 ms
2000 ms	IOI_S	944.06 ms	888.50 ms	799.70 ms	579.80 ms
	IOI_L	1055.94 ms	1110.51 ms	1220.30 ms	1420.21 ms

Table 6: Durations of the short and long inter-onset-intervals for each sequence (defined by ΔIOI and T) presented in Experiment 4.

Due to the response mode employed in the experiment, there was a fourth parameter (*Last Sound*) which represented the category of the sound ending a sequence (Fig. 69).

There were 4 × 3 × 2 × 2 (Δ_{IOI} × Period × Initial Position × Last sound) parameter combinations. To avoid effects of perceptual multistability, the sequences were considerably shorter than in experiment 1. As it was not clear whether the absolute length of the sequences (in seconds) or the number of repetitions of the short-long IOI pair should be constant, a trade-off between these aspects was attempted by presenting 25 repetitions at *T* = 500 ms (sequence length 12.5 *s*), 15 repetitions at *T* = 1000 ms (length 15.0 *s*), and 9 repetitions at *T* = 2000 ms (length 18.0 *s*).

The stimuli were generated by using a rim-short like sampled sound (duration = 50 ms). Only one sound was used throughout the experiment. Again, as no sound proof chamber was available, listeners were allowed to adjust presentation level to a comfortable value in the range between 70 and 90 dBA at the beginning of the experiment. Ambient noise level was approximately 38.2 dBA.

8.3.1.2 Apparatus

Stimuli were wave-files stored on the hard disk of a PC. These wave-files were played back via a stereo sound card (Terratec Xlerate). Subjects listened to the stimuli via closed Sennheiser HD-25 headphones connected directly to the sound card. A Tcl/Tk program controlled the experiment. Subjects responded by clicking with a computer mouse on response buttons displayed on screen.

8.3.1.3 Listeners

Twenty-five undergraduate psychology students of the University of Technology Berlin participated in the experiment. They received partial course credit. One subject had to be excluded from the analyses because of not participating in all three sessions. The remaining $N = 24$ participants ranged in age from 19 to 35 (mean age = 24.8 years). All reported normal hearing and were naive with respect to the aim of the experiment. None of the subjects had participated in the preliminary experiment.

8.3.1.4 Procedure

Participants sat in front of a computer monitor on which instructions and response buttons were presented. They were informed that in the sequences they were about to hear, there would be a regular alternation of louder and softer sounds. After presentation of each sequence, listeners made a decision regarding the *sound ending the sequence* by indicating whether this sound was one of the louder sounds. They responded by clicking with the computer mouse on one of two response buttons on the monitor (yes-no decision).

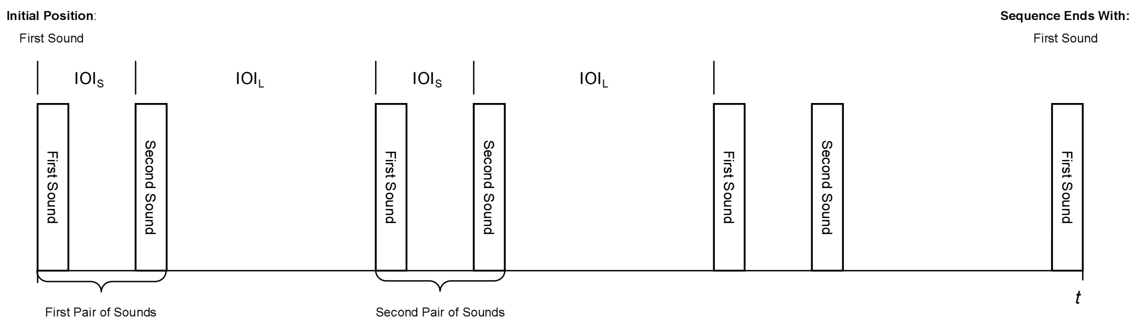

Fig. 69: 'Short-long' equitone sequence as presented in Experiment 4. The example sequence comprises three pairs of sounds, starts with a complete pair (Initial Position = first sound), and ends with a first sound (a sound ending a long IOI). Listeners responded whether the sound ending the sequence was one of the louder sounds.

An experimental session began with practice trials. Then, each of the 4 × 3 × 2 × 2 (Δ_{IOI} × Period × Initial Position × Last sound) sequences was presented three times; order of presentation was randomized.

The experiment was self-paced. After one third and two thirds of all trials were presented, there was a 5 minutes break. A session took approximately 80 minutes. Each subject participated in three sessions of identical structure, resulting in nine presentations of each sequence and a completely crossed and balanced four-way within-subjects design.

8.3.1.5 Data analysis

For each sequence (Δ_{IOI} × Period × Initial Position), and each position of the last sound, a listener responded nine times whether he or she had perceived the last sound of the sequence to be one of the louder sounds. If for example the sequence ended with a second sound (a complete pair) and the listener indicated this sound to be one of the louder sounds, this was taken as evidence for the listener having perceived the second sounds in the sequence as louder than the first sounds. In the following, the dependent variable is termed *P*("Last sound loud").

At this point, it is important to note that the procedure relies on the reasonable assumption that listeners responded on the basis of rhythmic anticipation. This assumption is not necessary for the case of sequences ending with a complete pair, but essential for sequences ending with a 'first sound'. In this condition, the sequence ends with a pair of tones followed by a single sound that is separated from the onset of the second sound of the last pair by a long IOI (Fig. 69). Now, strictly speaking, this single tone is no 'first sound', as it is not followed by another tone presented after a short IOI. As the sound is part of a rhythmic sequence within which accented and non-accented sounds alternate, however, it is assumed that subjects base their responses on the perception of periodically recurring accents. Assume that during presentation of the three tone pairs in Fig. 69, perception of the second sounds as accented emerged. In this case, after presentation of the third pair, the listener would anticipate the next tone to be soft, and base his or her response on this rhythmic anticipation. If one recalls the procedure used in the preliminary experiment, it becomes evident that listeners used exactly the same 'anticipatory responding' in that setting: it is generally agreed that synchronisation tapping can only be explained by an anticipatory mechanism and not by simple reacting (see Vorberg and Wing, 1996, for a review). In other words, a subject in the preliminary experiment *anticipated* the next sound to be accented and therefore timed her or his tap accordingly.

Another argument supporting this view becomes apparent if one assumes for a moment that subjects had indeed based their responses on the perception of only the very last sound of a sequence and not on the structure of periodically recurring accents. As the sound ending the sequence is always followed by a relatively long inter-trial-interval, the interrupted processing model would predict this sound to be always perceived as loud. I.e., subjects should have responded "Last sound loud" in each and every trial.

Nevertheless, a bias towards responding "Last sound loud" resulting from the complete processing of the last sound cannot be ruled out. Such bias should be independent of the temporal structure of the sequence, however. In modeling the data from Experiment 4, a bias parameter will be included.

For data analysis in terms of an ANOVA, across both endings of a sequence, the proportion of trials in which second sounds were perceived as louder than first sounds can be computed as

$$P(\text{'Second Sounds Louder'}) = \tfrac{1}{2}\,[1 + P(\text{"Last sound loud"}|\text{Sequence ending with second sound}) - P(\text{"Last sound loud"}|\text{Sequence ending with first sound})] \qquad (8\text{-}18)$$

where *P*("Last sound loud"|Sequence ending with second sound) is the proportion of trials presenting a sequence ending with a second sound within which the listener responded with "Last sound loud".

If a subject indicated nine times that she or he perceived the last sound of a specific sequence (defined by the temporal parameters) as loud when the sequence ended with a complete pair, and indicated nine times that she or he perceived the last sound of the sequence as "not loud" when the sequence ended with a first sound, the resulting value of *P*(*'Second Sounds Louder'*) is 1. If a subject responded nine times that she or he perceived the last sound of the sequence as loud when this sound belonged to the category of first sounds, and also indicated nine times that she or he perceived the last sound of the sequence as soft when it belonged to the other category of sounds, *P*(*'Second Sounds Louder'*) is 0. If a subject stated equally often to have perceived the last sound as loud, regardless which sound category ended the sequence, the value of *P*(*'Second Sounds Louder'*) is ½.

8.3.2 Model

For each sequence (uniquely defined by Δ_{IOI}, period T, and initial position), two response proportions were obtained in the experiment.

If the sequence ended with a first sound (a sound following a long IOI), the proportion of trials in which the listener indicated that the last sound was accented, denoted as *P*("Last sound loud"|Sequence ending with first sound), can be taken as an estimate of *p*("First Sounds Louder"). The model (Eq. 8-10) predicts

$$p(\text{“First Sounds Louder”}) = p(X_1 - X_2 > C) = \int_{x=-\infty}^{-C} \mathrm{PDF}(N_{\mu_2 - \mu_1, \sqrt{2}\sigma}, x)\, dx. \qquad (8\text{-}19)$$

Similarly, if the sequence ended with a second sound (a sound following a short IOI), the proportion of trials in which the listener responded that he or she perceived the last sound as loud, denoted as *P*("Last sound loud"| Sequence ending with second sound), is an estimate of

$$p(\text{“Second Sounds Louder”}) = p(X_2 - X_1 > C) = \int_{x=C}^{\infty} \mathrm{PDF}(N_{\mu_2 - \mu_1, \sqrt{2}\sigma}, x)\, dx. \qquad (8\text{-}20)$$

8.3.3 Hypotheses

P("Last sound loud") is expected to increase with ΔIOI for sequences ending with a second sound and to decrease with ΔIOI for the remaining sequences.

According to the model, accents perceived on the first tones can only be due to initial position, so that *P*("Last sound loud") should be smaller for sequences ending with a first sound (i.e., a single tone preceded by a long IOI). *P*("Last sound loud") is predicted to be greater if the sound ending the sequence belongs to the same category of sounds (first or second sounds, respectively) as the sound initiating the sequence.

For data analysis in terms of an ANOVA, the hypotheses formulated for the dependent variable P_{Tap}(Second Sounds) of the preliminary experiment (Chapter 8.2.1) also apply to *P*('Secound Sounds Louder) observed in Experiment 4. Specifically, if $C = 0$,

$$\begin{aligned} &P(\text{‘Second Sounds Louder’}) = \\ &\tfrac{1}{2}\,[1 + p(X_2 - X_1 > 0) - p(X_2 - X_1 < 0)] = \\ &\tfrac{1}{2}\,(1 + p(X_2 - X_1 > 0) - [1 - p(X_2 - X_1 > 0)]) = \\ &p(X_2 - X_1 > 0) = \int_{x=0}^{\infty} \mathrm{PDF}(N_{\mu_2 - \mu_1, \sqrt{2}\sigma}, x)\, dx, \end{aligned} \qquad (8\text{-}21)$$

i.e., Eq. 8-21 and Eq. 8-10 are equivalent.

8.3.4 Results and discussion

In the left panel of Fig. 70, proportions of "Last Sound Loud"-responses are displayed as a function of Δ_{IOI} at each of the three periods T. As expected, listeners only infrequently perceived the first sound of a pair as loud (open symbols). The proportion of "Last Sound Loud" responses generally increased with Δ_{IOI} for sequences ending with a second sound, while they decreased for sequences ending with a first sound. At the smallest Δ_{IOI}, the difference in *P*("Second Sounds Loud") between sequences ending with a first and a second tone, respectively, was small.

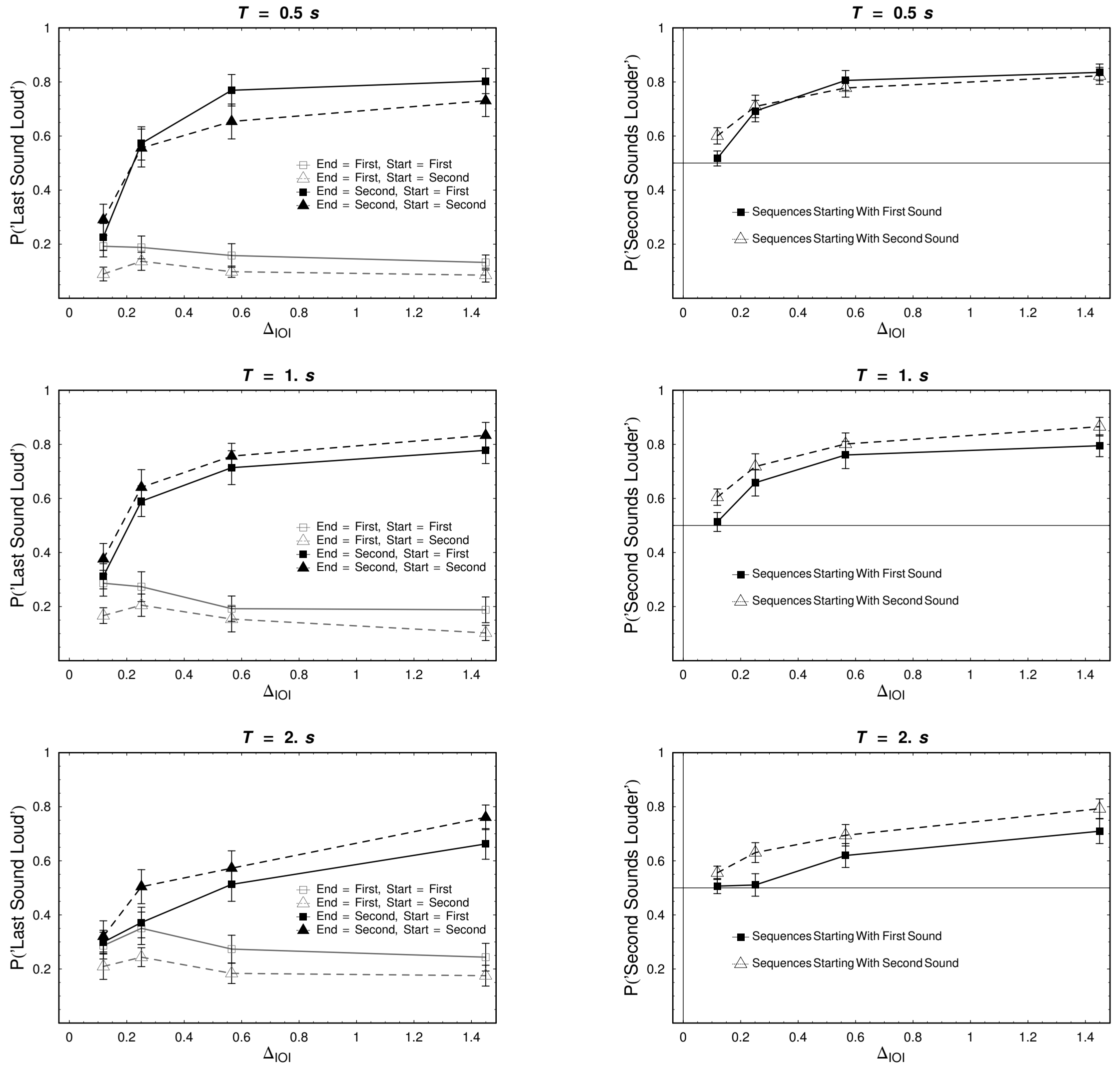

Fig. 70: Left panel: Proportion of "Last Sound Loud" responses for each period T, plotted as a function of relative IOI difference Δ_{IOI}. Open symbols: sequences ending with first sound. Filled symbols: sequences ending with second sound. Squares: sequences starting with first sound. Triangles: sequences starting with second sound.
Right panel: *P*('Second Sounds Louder') *for each period T, plotted as a function of Δ_{IOI}. Squares: Initial Position = first sound. Triangles: Initial Position = second sound. Error bars show ± 1 SEM.*

Except in one condition, initial position had the expected effect. The proportion of "Last Sound Loud" responses was greater if the sequence started with the same sound category (first sound, second sound) that also ended the sequence. The notable exception were fast sequences (T = 0.5 s) ending with a second sound. In this case, the Initial Position caused a change in the opposite direction.

The right panel of Fig. 70 displays the observed values of *P('Second Sounds Louder')* computed according to Eq. 8-16. Repeated measures analyses of variance (with Greenhouse-Geisser *df*-correction) were conducted for this variable.

As seen in Fig. 71, the predicted increase of the proportion of accents perceived on the second sounds with relative IOI difference Δ_{IOI} was present in the data, $F(3, 75) = 66.17$, $\hat{\varepsilon} = 0.690$, $p < 0.001$.

The effects of period T, $F(2, 50) = 7.63$, $\hat{\varepsilon} = 0.770$, $p < 0.004$, and the Δ_{IOI} × Period interaction, $F(6, 150) = 2.79$, $\hat{\varepsilon} = 0.737$, $p < 0.027$, were also significant. Fig. 71 displays the expected larger values of *P('Second Sounds Louder')* at smaller values of T. The difference between the means observed at the two faster tempi was very small.

There was a significant effect of initial position, $F(1, 25) = 8.25$, $p < 0.008$. Contrary to the hypotheses, at T = 0.5 *s*, *P('Second Sounds Louder')* was larger if the sequence started with a first sound. Yet, the Period × Initial Position interaction was not significant, $F(2, 50) = 1.86$, $\hat{\varepsilon} = 0.856$.

Neither the Δ_{IOI} × Initial Position nor the Δ_{IOI} × Period × Initial Position interaction were significant at an alpha level of 0.05.

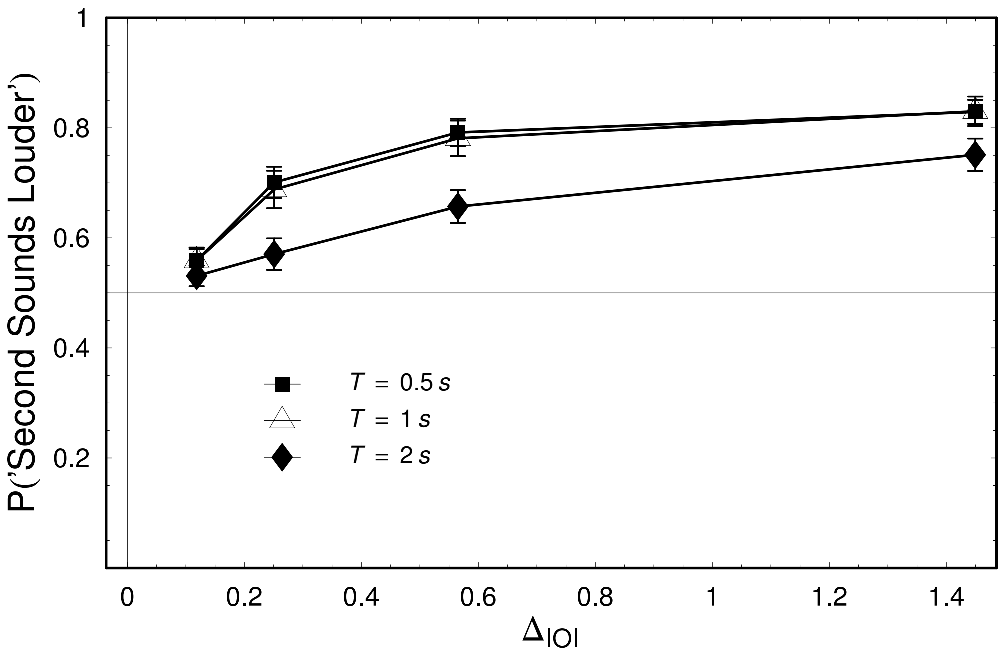

Fig. 71: Proportion of accents reported on the second sounds plotted as a function of Δ_{IOI}. Squares: period T = 0.5 s. Triangles: T = 1 s. Diamonds: T = 2 s. Error bars show ± 1 SEM.

Can the data provide an estimate of the interval range within which the interval-produced accent can be observed?

Fig. 71 suggests that there was an effect of Δ_{IOI} even at T = 2 s, where all IOIs were longer than 580 ms (Table 6). If the maximum intra-pair interval at which an accent on the second tones is perceived was 250 ms --as suggested by Povel and Okkerman (1981)-- there should have been no effect of the interval difference at the slowest tempo, as an ISI of 250 ms corresponds to an IOI of 300 ms. A repeated-measures Δ_{IOI} × Initial Position ANOVA conducted for *P('Second Sounds Louder')* observed at $T = 2$ s indicated a significant effect of Δ_{IOI}, $F(3, 75) = 22.47$, $\hat{\varepsilon} = 0.821$, $p < 0.001$. All but one of the pair wise post-hoc comparisons between the means of *P('Second Sounds Louder')* were significant at an α_{FW} of 0.05, the exception being the pair of the two smallest Δ_{IOI}'s used. The pair wise comparisons were computed with non-pooled error terms --because the value of $\hat{\varepsilon}$ indicated a non-spherical variance-covariance matrix-- and using a Bonferroni-procedure (cf. Maxwell and Delaney, 2004). These results are evidence for the maximum intra-pair interval at which an interval-produced accent can be observed to be significantly greater than 250 ms.

8.3.5 Model fitting

The quantitative model as defined by Eqs. 8-2, 8-11, 8-12, 8-19, and 8-20 includes six parameters. The decay parameter s is most important for the present purposes because it allows estimating the temporal limit of the interval-produced accent. N_{Max} is the asymptotic value of loudness approached at very long IOIs. The parameter r_0 denotes the relative reduction in perceived intensity at ISI = 0, i.e., if there is no silent interval between the target tone and the next tone. The effect of the initial position is represented by $N_{InitPos}$. C is the cut-off (or bias-parameter), and $\sigma = \sigma_1 = \sigma_2$ is the common standard deviation of the random variables X_1 and X_2 representing loudness of first and second sounds, respectively. The standard deviation represents sensitivity, i.e., it determines the effect a given difference $\mu_2 - \mu_1$ has on the response probabilities.

It follows from the model equations that

$$\mu_1 = \begin{cases} N_{Max}(1 - r_0\, e^{-s\,(IOI_S - d)} + N_{InitPos}) & \text{if } \textit{Initial Position} = \text{First Sound} \\ N_{Max}(1 - r_0\, e^{-s\,(IOI_S - d)}) & \text{else,} \end{cases} \qquad (8\text{-}22)$$

and

$$\mu_2 = \begin{cases} N_{Max}(1 - r_0\, e^{-s\,(IOI_L - d)} + N_{InitPos}) & \text{if } \textit{Initial Position} = \text{Second Sound} \\ N_{Max}(1 - r_0\, e^{-s\,(IOI_L - d)}) & \text{else.} \end{cases} \qquad (8\text{-}23)$$

As sound duration was constant, N_{Max} and r_0 are also constant, i.e., the same values are effective for all sequences presented in the experiment. Therefore, it is not possible to obtain independent estimates for all of the parameters.

This can be illustrated by considering the predicted proportion of "Last sound loud" responses for sequences ending with a complete pair (i.e., a second tone), which is given by

$$\text{P(“Last sound loud”|Sequence ending with second sound)} =$$

$$\begin{cases} \frac{1}{2}\,\text{Erfc}\left(\frac{C + N_{Max}\, r_0(e^{-s\,(IOI_L - d)} - e^{-s\,(IOI_S - d)}) + N_{Max}\, N_{InitPos}}{2\sigma}\right) & \text{if Initial Position = First Sound} \\ \frac{1}{2}\,\text{Erfc}\left(\frac{C + N_{Max}\, r_0(e^{-s\,(IOI_L - d)} - e^{-s\,(IOI_S - d)}) - N_{Max}\, N_{InitPos}}{2\sigma}\right) & \text{if Initial Position = Second Sound,} \end{cases} \qquad (8\text{-}24)$$

where $\text{Erfc}(x) = 1 - \text{Erf}(x)$.

First, as N_{Max} and r_0 are constant, the factor $N_{Max} \cdot r_0$ determining the effect of growth of sensation can be replaced by N and we can also write $N'_{InitPos} = N_{Max}\, N_{InitPos}$. Second, as Luce and Galanter (1963, p. 216) have pointed out for the equation of comparative judgment, which is the decision model underlying Eq. 8-24, σ is merely the unit of the scale representing loudness. Therefore, the effects of growth of sensation, initial position, and response bias must be interpreted in relation to σ. Without loss of generality, we divide both the numerator and the denominator of the arguments to Erfc() by σ and write $C^* = C/\sigma$, $N^* = N/\sigma$, and $N^*_{InitPos} = N'_{InitPos}/\sigma$, so that Eq. 8-24 simplifies to

$$P(\text{“Loud”}|\text{Sequence ending with second sound}) = \begin{cases} \frac{1}{2}\,\mathrm{Erfc}\left(\dfrac{C^* + N^*\left(e^{-s\,(IOI_L - d)} - e^{-s\,(IOI_S - d)}\right) + N^*_{InitPos}}{2}\right) & \text{if Initial Position = First Sound} \\ \frac{1}{2}\,\mathrm{Erfc}\left(\dfrac{C^* + N^*\left(e^{-s\,(IOI_L - d)} - e^{-s\,(IOI_S - d)}\right) - N^*_{InitPos}}{2}\right) & \text{if Initial Position = Second Sound,} \end{cases} \quad (8\text{-}25)$$

which contains four free parameters only. In the following, we use Eq. 8-25 and simply drop the superscript “*”. The same substitutions are made for *P*(“Loud”|Sequence ending with first sound).

Estimates of the four parameters were obtained by fitting the model to the mean values of *P*(“Loud”| Sequence ending with first sound) and *P*(“Loud”| Sequence ending with first sound) observed for each sequence (defined by Δ_{IOI}, period, and initial position). The data provided 4 × 3 × 2 × 2 = 48 equations. The computer algebra system Mathematica 5.0.1 was used for computing least-squares estimates of the parameters.

Fig. 72, left panel, displays observed and predicted proportions of “Last Sound Loud” responses. The coefficient of determination R^2 (Eq. 7-19; Kvålseth, 1985) was 0.872, which represents a reasonable fit. The fit was closest for the data observed at the slowest tempo, $T = 2$ s. The increase in the observed proportions as Δ_{IOI} was increased from 0.1185 to 0.2510 was larger than predicted, especially at the faster tempi. The predicted proportion of accents reported on the second sounds (aggregated over the two sequence endings, Eq. 8-18) is shown in the right panel of Fig. 72.

The best fitting value of the growth parameter s was 2.77 (95% confidence interval: 2.17, 3.37).

N was estimated to be 6.45.

The best fitting value of $N_{InitPos}$ was 0.11.

The fit indicated a bias towards responding that the sound ending the sequence was soft, as *C* was estimated to be 0.57.

The fact that *N* was estimated to be considerably greater than $N_{InitPos}$ and *C* indicates that the effect of growth of sensation was stronger than the effects of initial position and response bias.

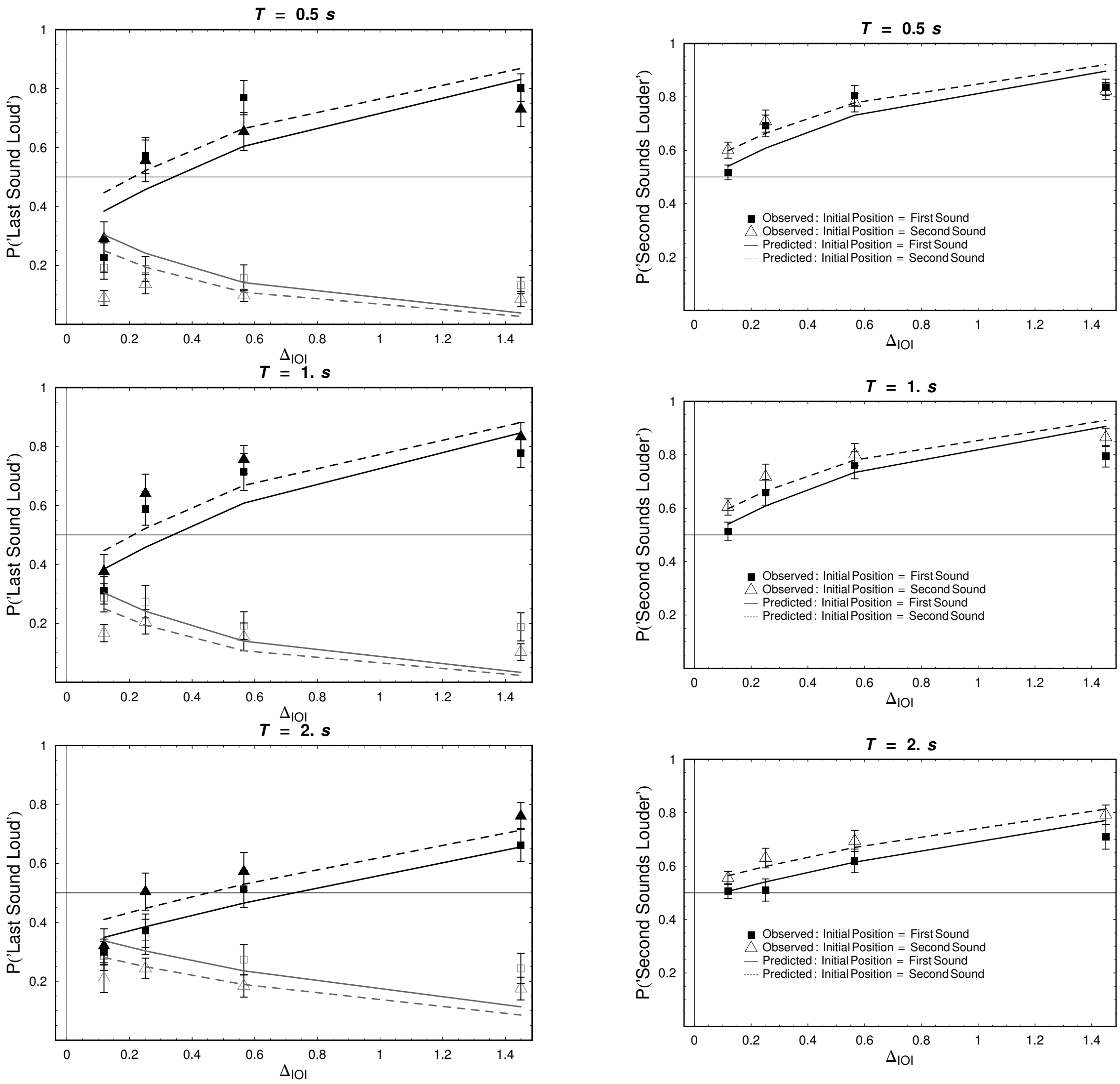

Fig. 72: Left panel: Observed (symbols) and predicted (lines) proportions of "Last Sound Loud" responses for each period T, plotted as a function of Δ_{IOI}. Open symbols/gray lines: sequences ending with first sound. Filled symbols/black lines: sequences ending with second sound. Squares/solid lines: sequences starting with first sound. Triangles/dashed lines: sequences starting with second sound. Best-fitting parameters: $s = 2.77$, $N = 6.45$, $N_{InitPos} = 0.11$, $C = 0.57$. $R^2 = 0.872$.
Right panel: Observed (symbols) and predicted (lines) P('Second Sounds Louder') for each period T, plotted as a function of Δ_{IOI}. Squares/solid lines: Initial Position = first sound. Triangles/dashed lines: Initial Position = second sound. Error bars show ± 1 SEM.

Note that the observed proportion of accents reported on the second tones seemed to asymptote at a value of 0.8 rather than 1.0 at the fastest tempo ($T = 0.5$ s). The problem of a restricted range of observed psychometric functions is normally attributed to "lapses" that result in, for example, the percentage of correct responses to be smaller than 100% even for large signal intensities in a detection task (cf. Klein, 2001). For this reason, it might be suspected that the value of s estimated above underestimates the 'true' growth parameter as the empirical asymptote is reached at a smaller ISI than the theoretical asymptote (1.0). The effect of 'lapses' can be included in the model by modifying Eq. 8-19 so that the minimum proportion of 'Last Sound Louder' responses is predicted to be λ rather than 0 and the maximum proportion of 'Last Sound Louder' responses is predicted to be (1 − λ) rather than 1 (e.g., Wichmann and Hill, 2001),

$$p(\text{“First Sounds Louder”}) = \lambda + (1 - 2\lambda)\, p(X_1 - X_2 > C) =$$
$$\lambda + (1 - 2\lambda) \int_{x=-\infty}^{-C} \mathrm{PDF}(N_{\mu_2 - \mu_1, \sqrt{2}\sigma}, x)\, dx. \qquad (8\text{-}26)$$

An analogous modification applies to Eq. 8-20.

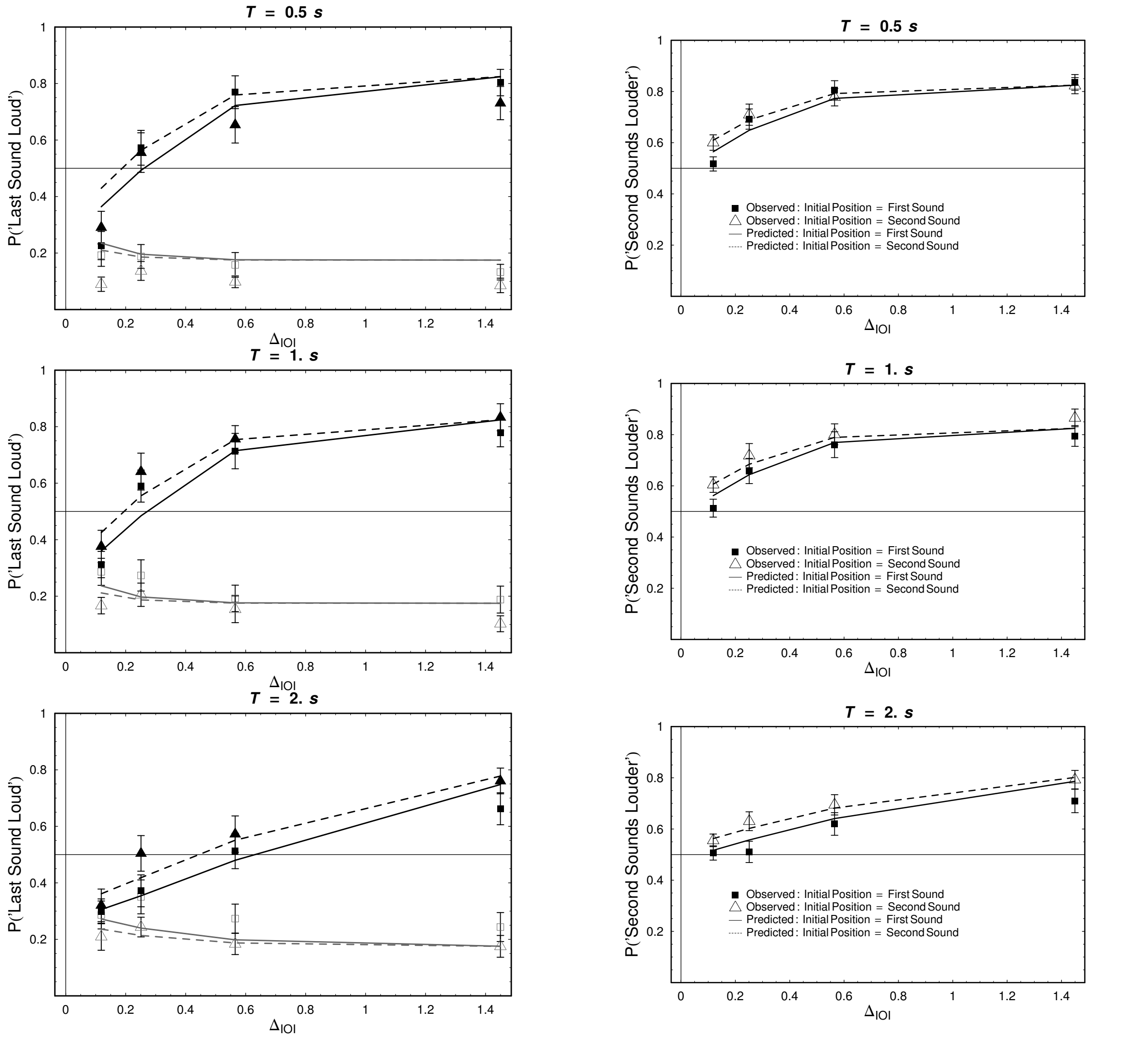

Fig. 73: Fits of the model version accounting for the restricted range of proportions of "Last sound loud" responses (lapse parameter λ). Left panel: Observed (symbols) and predicted (lines) proportions of "Last Sound Loud" responses for each period T, plotted as a function of Δ_{IOI}. Open symbols/gray lines: sequences ending with first sound. Filled symbols/black lines: sequences ending with second sound. Squares/solid lines: sequences starting with first sound. Triangles/dashed lines: sequences starting with second sound. Best-fitting parameters: $s = 2.91$, $N = 17.58$, $\lambda = 0.175$, $N_{InitPos} = 0.19$, $C = 1.33$. $R^2 = 0.929$.
Right panel: Observed (symbols) and predicted (lines) P('Second Sounds Louder') *for each period T, plotted as a function of Δ_{IOI}. Squares/solid lines: Initial Position = first sound. Triangles/dashed lines: Initial Position = second sound. Error bars show ± 1 SEM.*

As Fig. 73 illustrates, inclusion of the additional 'lapse parameter' λ improved the fit (R^2 = 0.929). The estimated value of the growth parameter s was 2.91, which is only 5% greater than the value estimated using the original model, however. It can thus be concluded that the rather slow growth of sensation indicated by the estimate of s can not be attributed to the restricted range of observed "Last sound louder" responses.

The model can also be used to compute parameter estimates for the data obtained by Povel and Okkerman (1981). The dependent variable in their experiments was the proportion of trials in which listeners reported accents on the second tones. Two tone durations were presented. The model version including the lapse parameter λ was used for fitting, according to which

$$p(\text{“Second Sounds Louder”}) = \lambda + (1-2\lambda)\, p(X_2 - X_1 > C) = \lambda + (1-2\lambda) \int_{x=C}^{\infty} \mathrm{PDF}(N_{\mu_2-\mu_1,\sqrt{2}\sigma}, x)\, dx. \tag{8-27}$$

As discussed above, both N_{Max} and N_0 can be assumed to vary with tone duration,

$$p(\text{“Second Sounds Louder”}) = \lambda + (1-2\lambda) \begin{cases} \frac{1}{2}\mathrm{Erfc}\left(\frac{C + N_{Max}(d)\, r_0(d)\left(e^{-s\, ISI_L} - e^{-s\, ISI_S}\right) + N_{Max}(d)\, N_{InitPos}}{2\sigma}\right) & \text{if } \textit{Initial Position} = \text{First Sound} \\ \frac{1}{2}\mathrm{Erfc}\left(\frac{C + N_{Max}(d)\, r_0(d)\left(e^{-s\, ISI_L} - e^{-s\, ISI_S}\right) - N_{Max}(d)\, N_{InitPos}}{2\sigma}\right) & \text{if } \textit{Initial Position} = \text{Second Sound.} \end{cases} \tag{8-28}$$

In order to avoid fitting a model with a large number of free parameters, however, two separate fits were computed for the 50-ms and the 150-ms tone duration so that the same number of free parameters as for the data from Experiment 4 was estimated. The same replacements as above were made, resulting in

$$\text{p(“Second Sounds Louder”)} = \lambda + (1-2\lambda) \cdot \begin{cases} \frac{1}{2}\mathrm{Erfc}\left(\frac{C^* + N^*\left(e^{-s\, ISI_L} - e^{-s\, ISI_S}\right) + N^*_{InitPos}}{2}\right) & \text{if Initial Position = First Sound} \\ \frac{1}{2}\mathrm{Erfc}\left(\frac{C^* + N^*\left(e^{-s\, ISI_L} - e^{-s\, ISI_S}\right) - N^*_{InitPos}}{2}\right) & \text{if Initial Position = Second Sound} \end{cases} \tag{8-29}$$

for each sound duration. In the following, the superscript "*" is again dropped.

For the 50-ms duration, inter-individual variability was large in the condition presenting a 500-ms tone before the first pair of tones and a short ISI of 50 ms, as one listener reported accents on the second tones in nearly 100% of all trials except for the smallest interval difference (20 ms). For the remaining two listeners, the proportion of accents perceived on the second tones increased with interval difference and was nearly identical to the data from the condition without the additional 500-ms tone at the smaller long interval durations. Data from this condition were excluded from the analysis. Aside from the proportion of accents reported on the second sounds appearing to asymptote at a value smaller than 100% the model provided an excellent fit (R^2 = 0.94; Fig. 74, upper row). Surprisingly, the value estimated for the growth parameter s was 1.00 only, which represents an even slower growth of sensation than the value of 2.77 estimated for the data from Experiment 4 of the present study. A fit of the model with the additional lapse parameter λ resulted in an improved fit but an even smaller estimate of s = 0.82. Recall, that Povel and Okkerman interpreted their data as to demonstrate that loudness reaches an asymptote at an inter-tone interval of approximately 250 ms. As discussed above, this estimate is likely to underestimate the time constant of growth of sensation due to the negative correlation between ISI_S and the difference between the long and short interval for the sequences using a long interval of 340 ms and a potential ceiling effect for the sequences using a short interval of 50 ms. Indeed, the estimate of s provided by the quantitative model indicates that the data are compatible with a significantly slower growth of sensation. Note also the decision bias indicated by a value of C = 1.38, which accounts for the large proportion of "First tones accented" responses that was observed for $ISI_S = ISI_L$ = 340 ms, independent of initial position (Fig. 74, upper row, right panel).

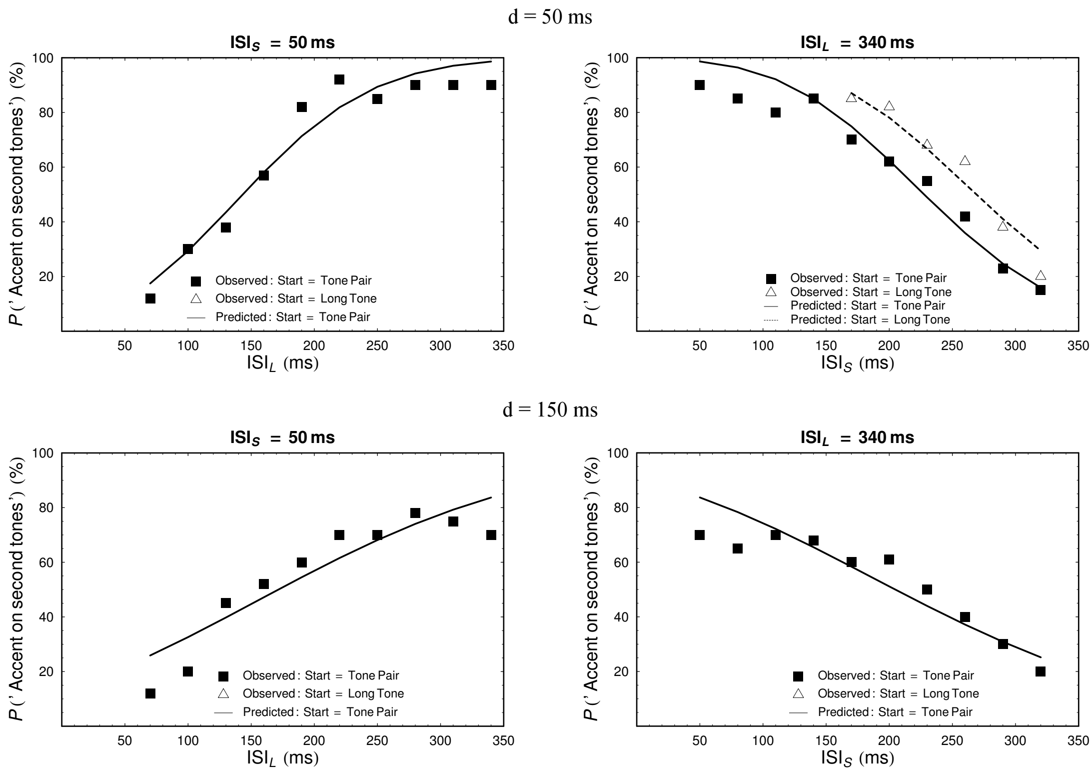

Fig. 74: Observed (symbols) and predicted (lines) proportions of "Second Tones Loud" responses as obtained by Povel and Okkerman (1981, Exp. 1-3).
Upper row: tone duration 50 ms. Squares/solid lines: Sequences starting with tone pair. Triangles/dashed line: Sequences starting with long tone. Best-fitting parameters: s = 1.00, N = 20.08, $N_{InitPos}$ *= 0.32, C = 1.38.* R^2 *= 0.94.*
Lower row: tone duration 150 ms. Best-fitting parameters: s = 0.67, N = 14.64, $N_{InitPos}$ *= 0 (fixed), C = 1.10.* R^2 *= 0.85.*

For the 150-ms tone duration, initial position was not varied. In this case, the parameters C and $N_{InitPos}$ have exactly the same effect in the model, so that it is impossible to compute independent estimates of both parameters. Therefore, $N_{InitPos}$ was set to zero. The best-fitting value of s was 0.67, which is even smaller than for the 50-ms tone duration data. Again, one could suspect that this very low estimate was due to the observed asymptotic value of P('Accent on second tones') being only about 0.7 instead of 1.0. The model including the lapse parameter λ again produced a better fit (R^2 = 0.91). The estimated value of s was now only 0.39, however.

To complete the picture, a fit to the data from the preliminary experiment was also computed, although the patterns observed in this experiment were quite irregular, variability was large, and only a small number of data points were collected per listener. Not surprisingly, the fit was rather weak (R^2 = 0.72, Fig. 75). The best-fitting value of s was 4.30, which is larger than for the data from Experiment 4 and from Povel and Okkerman (1981), but still below s = 7.68 corresponding to a time constant of 300 ms. The remaining estimates were N = 39.50, λ = 0.21, C = 0, and $N_{InitPos}$ = 3.04. As discussed above, at period = 400 ms the influence of initial position under the smallest levels of Δ_{IOI} was much larger than predicted by the model, presumably due to a procedural bias.

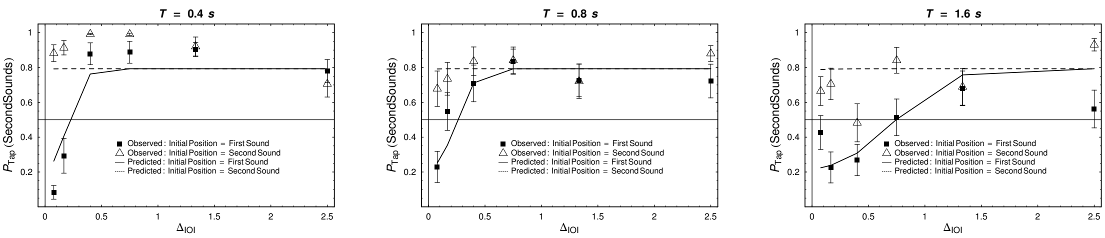

Fig. 75: Fit of the quantitative model to data from the preliminary experiment. Observed (symbols) and predicted (lines) proportions of taps marking the second sounds as louder, plotted as a function of Δ_{IOI}. Squares/solid lines: Initial Position = first sound. Triangles/dashed lines: Initial Position = second sound. Best-fitting parameters: s = 4.41, N = 39.50, λ = 0.21, $N_{InitPos}$ = 3.04, C = 0. R^2 = 0.72. Error bars show ± 1 SEM.

To illustrate growth of sensation corresponding to the different estimates of the parameter s obtained above, we use Eq. 8-4 that assumes loudness to follow the same time course during and after stimulus presentation (cf. the discussion in Chapter 8.2.1). Fig. 71 plots predicted loudness as a function of the inter-onset-interval between test tone and the next sound. The values of the growth parameter s estimated from the data of Experiment 4 and from the data reported by Povel and Okkerman (1981), respectively, differ considerably. This inconsistency can only in part be attributed to the difficulty of estimating the time constant of an exponential growth function from behavioral data. Nevertheless, in both cases loudness grows considerably slower than it would be the case for a time constant of 300 ms. Recalling that the parameter s also governs the decay of forward loudness enhancement with the silent interval preceding the target tone (Eq. 8-7), loudness enhancement of the second by the first tone could explain the accent only if it is assumed that there is an effect at 'conditioner-target' intervals larger than the 500 ms reported by Zwislocki and Sokolich (1974).

For both interrupted processing and loudness enhancement, a potential explanation of the longer time constant observed in the accent perception experiments is that the short-long sequences represent a 'multiple looks' situation. To give an example, while listeners compare the loudness of only one pair of tones (target and conditioner) in each trial of a three-tone matching experiment, up to 25 pairs were presented in Experiment 4. Thus, the probability that a listener 'detected' a small loudness difference between the first and the second tones was considerably greater due to the 'multiple looks' situation (cf. Chapter 9.5 in Green and Swets, 1966). This would explain why the proportion of "Second tones louder" still differed from 50% at rather long intra-pair intervals in Experiment 4. From this point of view, the proportion of accents 'detected' on the second tones is not the optimum dependent variable, as it does not provide a direct measure of the loudness difference between second and first tones.

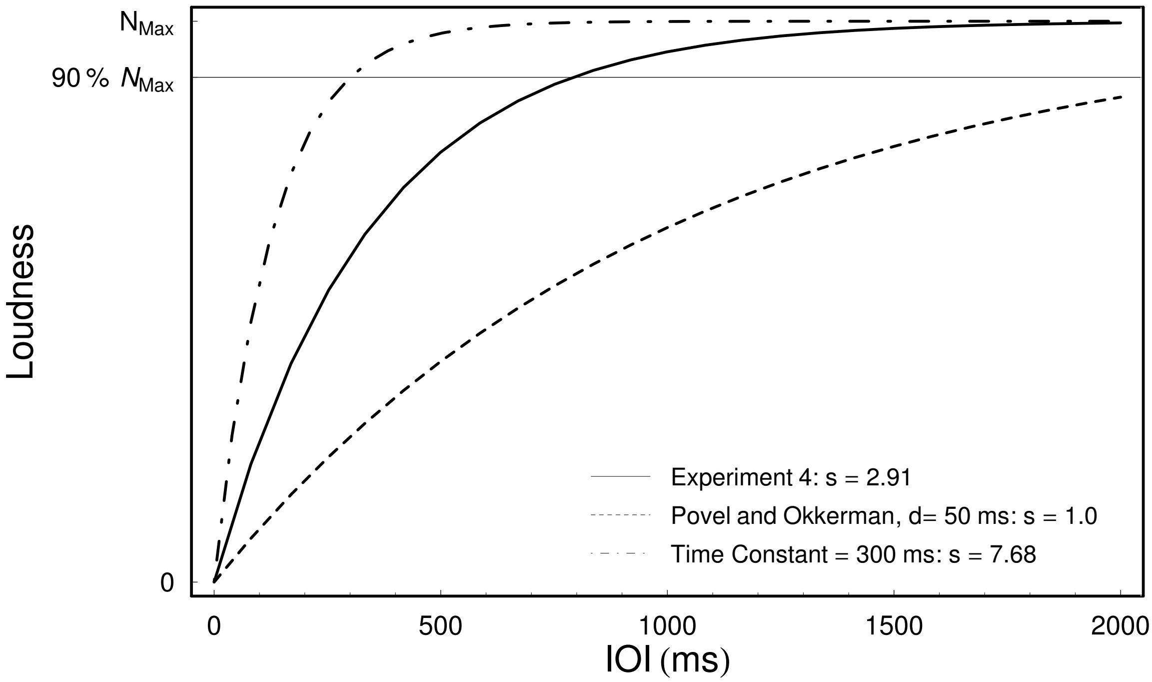

Fig. 76: Predicted growth of sensation as a function of the silent interval between test tone onset and the onset of the next sound. Solid line: growth parameter s as estimated from the data of Experiment 4. Dashed line: s estimated from the data by Povel and Okkerman (1981), tone duration 50 ms. Dot-dashed line: growth of sensation corresponding to a time constant of 300 ms as proposed by Povel and Okkerman.

8.3.6 Discussion

Experiment 4 provided important empirical evidence regarding the effect of temporal sequence structure on the interval-produced accent. As relative IOI difference was varied independent of intra-pair interval duration, the data allowed deriving an unbiased estimate of the temporal limit of accent perception. An ANOVA confirmed that a significant accent on the second tones was present for intra-pair intervals up to 800 ms.

The estimates of the growth parameter *s* computed for the data from Experiment 4 confirmed the results of the ANOVA reported in the preceding chapter, as they also indicate the temporal limit of the interval-produced accent to be considerably larger than the 300 ms proposed by Povel and Okkerman.

Even though the model based on the concept of interrupted processing was capable of quantitatively accounting for not only the data from Experiment 4, but also for the data reported by Povel and Okkerman (1981), the estimated time course is not completely compatible with the interrupted processing hypothesis, or at least not with the assumption that processing in the *short* auditory store is the mechanism relevant for the accent. The duration of the short store was estimated to be 250-300 ms (Cowan, 1984; Massaro and Loftus, 1996), which is not compatible with the growth-of-loudness functions estimated for the accent data (Fig. 76). This conclusion must be viewed in relation to the difficulties of estimating time constants from behavioural data and the ‘multiple looks’ argument discussed above, however.

The same qualification applies to the fact that the estimate of the time constant derived from the data by Povel and Okkerman indicates a temporal limit that is not compatible with the time course of forward loudness enhancement, as Zwislocki and Sokolich (1974) found enhancement to disappear at an ISI of 500 ms.

To conclude, Experiment 4 provided important empirical evidence regarding the time course of the interval-produced accent. The quantitative model could account for the effects of temporal sequence structure and initial condition. It was also useful in deriving a more reliable estimate of the temporal limit of accent perception from the data by Povel and Okkerman (1981).

8.4 Experiment 5: Effects of presentation level on the interval-produced accent

As the most interesting effects in loudness enhancement are associated with stimulus intensity, Experiment 5 studied the influence of presentation level on the interval-produced accent.

The interrupted processing hypothesis makes no assumptions concerning presentation level and thus predicts no effect of this parameter.

The presence of a 'midlevel hump in interval-produced accent' would certainly represent an important link between the accent and loudness enhancement. On the other hand, the midlevel hump is due to the effect of masker-target similarity according to the similarity model. In equitone sequences, the similarity between first and second tones is by definition constant and therefore independent of level. Above that, the mergence hypothesis predicts loudness enhancement to be absent in equitone pairs.

Experiment 4 and most previous experiments used as dependent variable the proportion of trials in which listeners reported the second tones of the tone pairs to be louder than the first tones.

To facilitate comparisons with loudness enhancement experiments, an adaptive procedure was used in Experiment 5 to obtain loudness matches between first and second sounds. The dependent variable was the level difference between first and second sounds required to make the two sound categories sound equally loud. Povel and Okkerman (1981) reported data from a loudness matching task, which indicated that depending on temporal structure, the level difference between the first and second tones ($L_1 - L_2$) required for the loudness match was −4 to +4 dB. Positive values correspond to the second tones being perceived as louder than the first tones of the same presentation level.

In Experiment 5, the level of either the first or the second tones was constant within each track and set to 25, 55, or 85 dB SPL. Each sequence comprised six tone pairs. The listener responded whether the first or the second tones had been louder.

8.4.1 Method

8.4.1.1 Stimuli

The tone sequences consisted of 1-kHz pure tones, with a steady-state duration of 20 ms and gated on and off with 5-ms cosine-squared ramps.

In each trial, six tone pairs were presented. A sequence always started with a complete pair, i.e., initial position was not varied. The long ISI (offset-onset interval) was always 600 ms, the short ISI was varied between 100 and 600 ms (Fig. 77).

8.4.1.2 Apparatus

A MATLAB program was used for stimulus generation and control of the experiment. Stimuli were 24-bit resolution, 44.1-kHz sampling-frequency wave files. They were played back via an *M*-Audio Delta 44 PCI audio-card. Only one channel was used. Level differences between first and second tones were produced digitally. The signal was fed into a custom-made programmable attenuator, amplified in a headphone amplifier, and fed into the right channel of Sennheiser HDA 200 headphones.

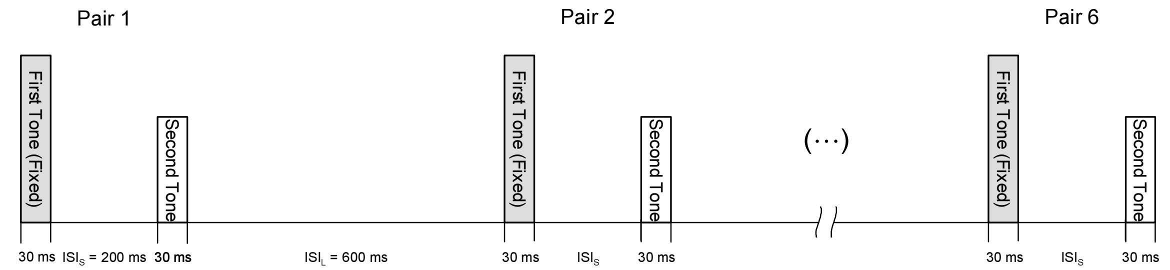

Fig. 77: Trial configuration used in Experiment 5. Four randomly interleaved tracks were used. The example configuration depicts a track with the level of the first tones fixed and the level of the second tones adjusted by the adaptive procedure (lower and upper track interleaved). Six pairs of tones were presented in each trial. The long interval ISI_L *was always 600 ms, the short interval* ISI_S *was 100, 200, 400, or 600 ms.*

For calibration, a Brüel and Kjær Artificial Ear 4153 (IEC 318 specification), a Brüel and Kjær Microphone 4134**,** a Rhode and Schwartz calibrator, and a Norwegian Electronics Sound Meter 108 were used.

The experiment was conducted in a sound-proof chamber (“Camera Silenta” of the Institute of Technical Acoustics at the University of Technology Berlin). Listeners were tested individually.

8.4.1.3 Listeners

Four listeners participated in the experiment. One of them (DO) was the author; the others were paid an hourly wage. Listeners DO and YS had already participated in Experiments 1 and 2. For the ear used, all listeners had hearing levels (as measured by pure-tone audiometry according to ISO 8253) better than 10 dB in the frequency range between 125 Hz and 8000 Hz, with exception of DO (HL = 10.9 dB at 1500 Hz).

Participants were fully informed about the course of the experiment. All except the author were naïve with respect to the aim of the experiment. Listeners received at least 2 hours of practice. If necessary, it was allowed for further practice until performance stabilized.

8.4.1.4 Procedure

In each track, the level of either the first or the second tones was constant and set to 25, 55, or 85 dB SPL. Presentation level of the fixed tones will be denoted as target level in the following.

Four randomly interleaved tracks were presented in each run. In two of the tracks, the level of the first tones was fixed and the adaptive procedure adjusted the level of the second tones. As in Experiment 2, an upper and a lower track were used. The upper track converged on the 79.4% “Second tones louder” point on the psychometric function (3-down, 1-up rule), while the lower track converged on the 20.6% “Second tones louder” point (1-down, 3-up rule).

In the remaining two tracks, the level of the second tones was fixed and level of the first tones was adjusted by the adaptive procedure to track either the 79.4% “First tones louder” point or the 20.6% “First tones louder” point on the psychometric function.

If only two interleaved tracks are used with, e.g., the level of the first tones fixed throughout the experiment, a listener could adopt a strategy of judging only the loudness of the second tones, effectively transforming the task into an absolute identification experiment. The procedure using four interleaved tracks was chosen to encourage listeners to actually compare the loudness of first and second tones in each trial.

Note that the task is equivalent to the loudness matching task in quiet used in Experiment 2, but with multiple observations. Therefore, to familiarize listeners with the rhythmic task, they received several runs with only **one** pair of tones at the beginning of the experiment. Observation interval 1 contained the fixed-level standard, observation interval 2 contained the comparison tone. Listeners selected the interval containing the louder tone by pressing one of two response buttons. The two tones were marked visually by two LEDs. LED 1 was switched on 50 ms before the onset of the tone 1 and switched off 50 ms after its offset. LED 2 marked tone 2 presented in the second interval in exactly the same way. As in Experiment 2, two randomly interleaved tracks were presented (3-down, 1-up and 3-up, 1-down rule, respectively). After this introductory training, listeners were informed that in each trial, they would now hear exactly the same pair of two tones six times and that either the first or the second tones would be louder. They were encouraged to optimize performance by basing their decision as to whether tone 1 or tone 2 had been louder on **all** of the six presentations of the pair. The LEDs marked tone 1 and tone 2 of each pair. Several runs of this task were presented for training.

No feedback was provided.

The upper tracks started with a sound-pressure-level of the adjusted tones 11 dB above target level (the level of the fixed tones), while in the lower track, the level of the adjusted tones was 11 dB below target level initially. There were eight reversals per track. Step size was 5 dB until the fourth reversal, and 2 dB for the remaining six reversals. If in one track had eight reversals had already occurred before termination of the other track, it was nevertheless presented with an a priori probability of 0.15.

In each run, the arithmetic mean of the difference between the level of the first tones (L_1) and the level of the second tones (L_2) at the last four reversals was computed for each track. The arithmetic mean of these four values (first or second tones adjusted, lower or upper track) was taken as the loudness match $L_1 - L_2$. A pair of tracks (lower and upper) was discarded if the standard deviation was greater than 6 dB in either the upper or the lower track.

As in Experiment 3,

$$\text{jnd}_{Match} = \frac{x_{0.794} - x_{0.206}}{2}, \qquad (8\text{-}30)$$

was computed as a measure of loudness variability.

In each block, only one standard level and one value if ISI_S were presented. Listeners received the conditions in pseudo-random order.

An experimental session lasted approximately one hour. Listeners took one ore two short breaks in each session.

8.4.2 Results

Fig. 78 (left panel) displays loudness matches between first and second tones ($L_1 - L_2$). Positive values indicate an accent on the second tones, i.e., the level of the first tones had to be increased above the level of the second tones for loudness of first and second tones to match.

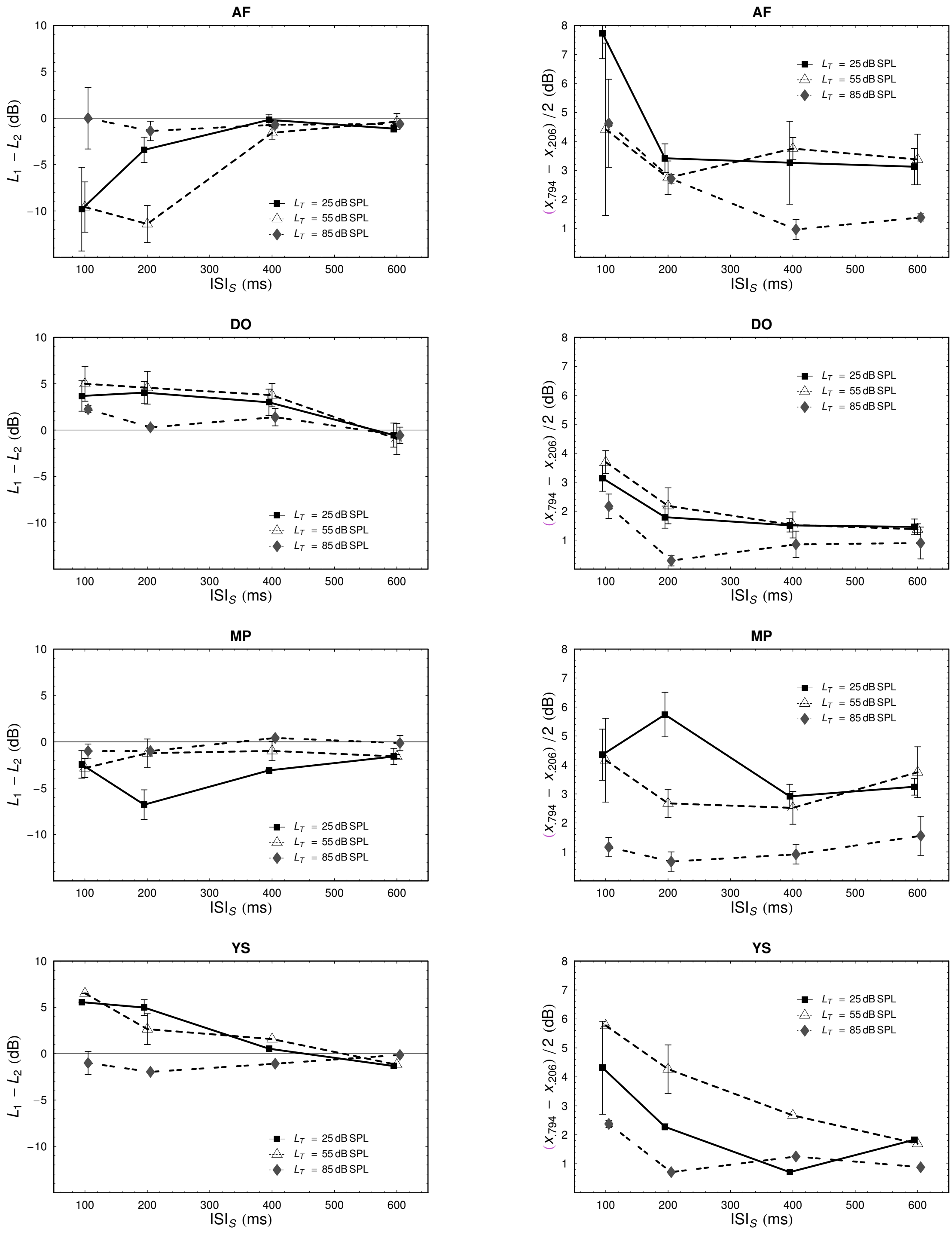

Fig. 78: Left panel: Loudness matches between first and second tones ($L_1 - L_2$) as a function of short interval duration (ISI_S). The long interval ISI_L was 600 ms. Right panel: loudness variability jnd_{Match}. Squares: target level 25 dB SPL. Triangles: target level 55 dB SPL. Diamonds: target level 85 dB SPL. To avoid cluttering, lines are shifted by 5 ms along the x-axis. Error bars show ± 1 SEM.

Generally, the effect of short interval duration was smallest at the highest presentation level (85 dB SPL; Fig. 78, left panel, diamonds). In this condition, the maximum value of $L_1 - L_2$ was 2.25 dB (listener DO, ISI_S = 100 ms). For listener MP, the effect of ISI_S was small at the 55-dB SPL target level also.

At the lower target levels, short interval duration produced the expected effect for two listeners (DO and YS). The level increment required to make the first tones sound equally loud as the second tones was 3.68 to 6.52 dB at the smallest ISI_S and decreased monotonically with short interval duration, reaching a value close to zero in the isochronous condition ($ISI_S = ISI_L = 600$ ms). Loudness matches at the smallest ISI_S were compatible with data by Povel and Okkerman (1981) who reported the loudness match between first and second tones to be 4 dB at the largest interval difference (50 ms versus 340 ms). The patterns observed at the 25-dB and the 55-dB SPL target level did not differ substantially, so that it can be concluded that there is no midlevel hump in accent perception.

Loudness matches of the remaining two listeners followed a totally unexpected pattern. An "inverse interval-produced accent" was observed, i.e., at small values of ISI_S, loudness matches indicate that AF and MP perceived the **first** tones to be louder. The level difference $L_1 - L_2$ was about −10 dB at the two smallest values of ISI_S for listener AF.

Loudness variability was generally larger at the smaller short-interval durations (Fig. 78, right panel) where loudness matches also indicated the strongest loudness differences. For listener YS, jnd_{Match} was larger at the 55-dB than at the 25-dB SPL target level, even though loudness matches did not differ substantially between these conditions.

For the two listeners who reported an 'inverse' accent, intensity discrimination difference limens were obtained in order to detect potential difficulties with the task. As mentioned above, listeners were instructed as if the loudness matching task involving rhythmic sequences were a multiple observation intensity discrimination task. For measurement of the difference limens, the same setting and procedure as in Experiment 1 was used, except that no masker was presented and a 3-down, 1-up rule was applied in the adaptive procedure. Standard level was 25, 55 or 85-dB SPL. In two observation intervals, the standard and the standard-plus-increment were presented. Listener selected the interval containing the louder sound. Visual trial-by-trial feedback was provided. For each listener, a long and a short value of the silent interval between the first and the second tone were used.

As seen in Fig. 79, the near miss to Weber's law was present at both ISIs. Difference limens tended to be larger at the short inter-tone interval. This is at odds with the predictions of the preliminary theory by Durlach and Braida (1969), according to which performance should be better at short ISIs, as the information provided by the sensory trace is assumed to decrease with the pause between the two tones. Therefore, the intensity discrimination data might indicate a procedural problem at small intra-pair intervals.

In fact, AF reported difficulties with the visual markers at the short intra-pair intervals. He was frequently uncertain as to whether the louder tone had been marked by LED 1 or LED 2. Certainly, this presents no problem as long as both tones are clearly audible and short and long interval durations are sufficiently different to cause a salient grouping of the sequence into tone pairs. In such condition, it should be possible to ignore the visual markers and to decide simply whether the first or the second tones of the pairs had been louder. AF reported to have used such a strategy. On the other hand, if the level of one tone is so low that only the other tone is audible, it is necessary to rely on the visual signals. The data produced by AF could be explained by assuming that he had a bias towards responding 'First Tones Louder' in conditions where he ignored the LEDs because the temporal separation between the tones of a pair was too small.

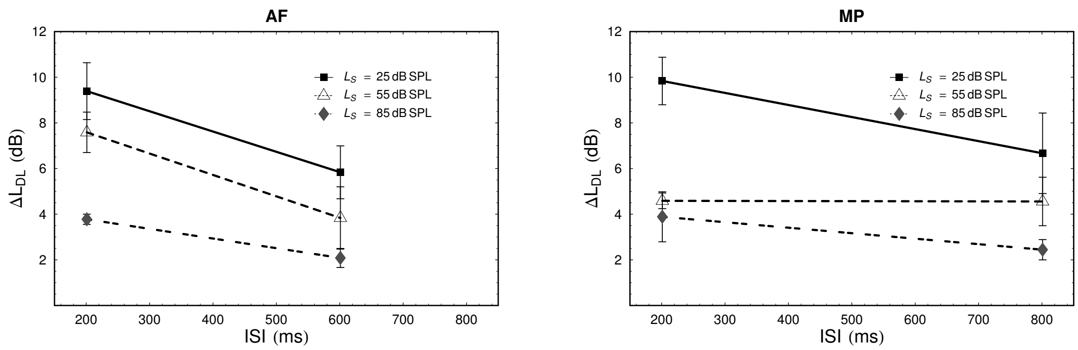

Fig. 79: Jnd's *(ΔL_{DL} = 10 $\log_{10}[1 + \Delta I_{DI}/I]$) as a function of the silent interval between the two tones. Adaptive 2IFC procedure (3-down, 1-up). Left panel: listener AF. Right panel: listener MP. Squares: 25-dB SPL standard. Triangles: 55-dB SPL standard. Diamonds: 85-dB SPL standard. Error bars show ± 1 standard error of the mean (SEM).*

To gain insight into to the problem, AF was tested in four blocks of 100 trials in which the level difference between the first tones and the second tones was either −25 dB or 25 dB. For such extremely large ΔL's, perfect performance can be expected. Six pairs of tones were presented in each trial. The first tones were louder in 50 trials and the second tones were louder in the remaining 50 trials. The level of the louder tones was 25 or 55 dB SPL and the intra-pair interval was either 100 or 200 ms. In each block, only one target level and one value of ISI_S was presented.

Table 7 displays the proportion of responses for each of the conditions. At the 25-dB SPL target level, the level of the soft tones was 0 dB SPL only, which is below detection threshold, so that AF must have used the visual markers in this condition. The listener made a small number of errors at the 100-ms ISI, but there was no evidence for a response bias. This finding indicates that contrary to his own impression, AF had no substantial problems in using the visual information provided by the LEDs for responding.

			Response	
L_T	ISI_S	Condition	"First Tones Louder"	"Second Tones Louder"
25 dB SPL	100 ms	$L_1 > L_2$	94.2%	5.8%
		$L_1 < L_2$	5.9%	94.1%
	200 ms	$L_1 > L_2$	100%	0%
		$L_1 < L_2$	0%	100%
55 dB SPL	100 ms	$L_1 > L_2$	96.3%	3.7%
		$L_1 < L_2$	21.6%	78.4%
	200 ms	$L_1 > L_2$	100%	0%
		$L_1 < L_2$	5.8%	94.2%

Table 7: Proportions of "First Tones Louder" and "Second Tones Louder" responses in the discrimination experiment conducted for listener AF. In each of the four blocks ($L_T \times ISI_S$), the level of the second tones was 25 dB below target level in half of the trials ($L_1 > L_2$). In the remaining trials, L_1 was 25 dB below target level ($L_1 < L_2$).

At the 55-dB SPL target level, where all tones were clearly audible and AF reported to ignore the visual markers at small intra-pair intervals, the substantial proportion of errors in the $L_2 > L_1$ condition is evidence for a bias towards responding 'First Tones Louder'. It remains unclear if the

5.8% error rate observed with the 200-ms intra-pair interval also indicates a response bias or can be attributed to lapses of attention.

Taken together, the data suggest that the loudness matches produced by AF at the 100-ms intra-pair interval are likely to have been distorted by a procedural problem.

Still, this does not explain the negative values of $L_1 - L_2$ at a short interval duration of 200 ms. Moreover, listener MP who produced a comparable pattern of loudness matches, did not report any problems with the procedure.

8.4.3 Discussion

For two of the four listeners, the loudness matches obtained in Experiment 5 indicated the same dependence of the interval-produced accent on temporal structure as in previous experiments. For a constant inter-pair interval of 600 ms, the level increment necessary for loudness of the first tones to match loudness of the second tones decreased with the intra-pair interval. There was a clear effect of presentation level, as the effect of the intra-pair interval was very small at 85-dB SPL, but substantial at the lower target levels. The interrupted-processing hypothesis can not account for any effect of presentation level. There was also no indication of a midlevel hump that would have been evidence for loudness enhancement as the cause of the accent. Note that if the loudness difference between the second and the first tones was independent of level, *smaller* loudness matches $L_1 - L_2$ should have been observed at the lowest level, as the loudness function for pure tones is steeper at low than at higher levels. The level increment required to make the first tones sound equally loud as the second tones was as large as 6.5 dB. A comparably large amount of loudness enhancement was not found in the experiments using the three-tone matching procedure and identical forward masker and target. As pointed out in Chapter 4.3.1, this discrepancy could be due to loudness of the first tone being reduced in the masker-equal-to-target condition, as such effect can not be detected in the three-tone matching task using a forward masker. In Experiment 2 of this study, mean loudness matches between target and comparison tone in the condition of identical masker and target ($\Delta L_{M\text{-}T} = 0$ dB; Fig. 28) were even negative, indicating a *reduction* of tone 2 loudness relative to loudness in quiet.

It is difficult to understand why loudness matches produced by the remaining two listeners indicated an accent on the **first** rather than the second tones at small values of ISI_S. At least for one listener, there seemed to have been a problem with the experimental procedure. The data might also be interpreted to demonstrate that the interval-produced accent is a cognitive phenomenon that is possibly influenced by experience or musical training (cf. Vos, 1977).

8.5 General Discussion

Experiment 4 and Experiment 5 provided important empirical evidence about the effects of temporal structure and presentation level on the interval-produced accent. The rationale behind the experiments was to decide whether the two effects are compatible with either loudness enhancement or interrupted processing as the mechanism underlying the accent.

The data from Experiment 4 indicated the temporal limit of accent perception to be considerably larger than both the duration of the *short* auditory store, which should be relevant for interrupted processing, and the maximum conditioner-target interval at which loudness enhancement is observed. Although many of the problems associated with estimating the time constant of an exponential growth function were avoided by using a mathematical model, it remains unclear whether the presentation of multiple tone pairs in the short-long sequences, which generates a ‘multiple looks’ situation, could be an explanation for the larger temporal limit observed in the accent-perception experiments.

The level dependence found in Experiment 5 can clearly not be accounted for by the interrupted processing model that makes no assumptions regarding the effects of presentation level. Similarly, the mergence hypothesis, which represents the most successful model of loudness enhancement, can neither account for non-zero matches if masker and target are physically identical, nor for the level dependence of the accent observed in Experiment 5.

The ‘inverse interval-produced accent’ observed for two of the four listeners in Experiment 5 is incompatible with either model and also with the data from previous experiments and could only in part be attributed to procedural problems.

Taken together, neither interrupted processing, nor loudness enhancement can account for the complete range of findings in Experiments 4 and 5.

If one recalls the data from Experiment 2, it is striking that loudness perception in pairs of identical tones seems to be difficult to understand. This conclusion applies to the complex pattern of results found for loudness matches between two tones in quiet (the time order error; cf. Hellström, 1985). Additionally, in Experiment 2, loudness matches between target tone and comparision tone were frequently different from matches in quiet if masker level was equal to target level. This finding presents a problem for the mergence hypothesis; inter-individual variability was also rather large. Loudness perception in multiple equitone pairs presented rhythmically again differs from the patterns found for a single pair of sounds.

9 Summary and conclusions

In the study of auditory intensity processing, several interesting and complex effects emerge as one moves from a single sound presented in isolation to two sounds presented in temporal proximity.

If a brief tone is presented together with an intense forward or backward masker, separated by a masker-target interval between 50 and 400 ms, the dependence of the just-noticeable intensity increment on presentation level differs distinctly from the near-miss to Weber's law observed in quiet. Midlevel jnd's are strongly increased relative to the values in quiet, while the masker has only small effects at low and high standard levels (the midlevel hump in intensity discrimination; e.g., Zeng, Turner, and Relkin, 1991).

The non-simultaneous masker also introduces a systematic shift of target loudness towards masker loudness, i.e., loudness of the target is enhanced by a masker higher in level than the target, while at masker levels below target level, loudness is reduced (e.g., Zwislocki and Sokolich, 1974).

In rhythmic equitone sequences constructed as pairs of sounds separated by a longer interval, an accent is perceived on the second sounds of the pairs, i.e., loudness of the second sounds exceeds loudness of the first sounds. This "interval-produced accent" (Povel and Okkerman, 1981) is observed only if the intra-pair interval is shorter than approximately 800 ms and if the difference between the inter-pair and the intra-pair interval is sufficiently large.

The objective of this work was the identification of the mechanisms underlying the three phenomena. More specifically, motivated by the observation of a multitude of similarities between the effects of a non-simultaneous masker on loudness and intensity resolution, as for instance the midlevel hump found in both loudness enhancement and intensity discrimination experiments, the idea was pursued that the phenomena might be caused by a *common mechanism*. In the same line of reasoning, the question was raised whether the interval-produced accent can be interpreted simply as a special case of loudness enhancement, i.e., whether the same mechanism that causes loudness of a target tone to be enhanced by an intense masker also underlies the interval-produced accent.

The systematic review of existing psychophysical data presented in Chapter 3 provided further evidence for the similarity between the phenomena, although differences in the effects of masker duration and forward versus backward maskers were identified.

Three models have been proposed for the effects of non-simultaneous masking on intensity resolution. The review of data provided the basis for a critical evaluation of these models in Chapter 4.

The recovery-rate model by Zeng, Turner, and Relkin (1991), which attributes the midlevel hump to different recovery rates of high spontaneous rate (SR) and low-SR auditory nerve neurons, can elegantly account for the basic effects of a forward masker on intensity resolution as well as for the rather complex interaction between non-simultaneous and simultaneous masking. As it has been pointed in previous studies, however, the model has to be rejected because it is incompatible with the midlevel humps in intensity discrimination observed with contralaterally presented maskers and backward maskers.

Formulated as an alternative to the recovery-rate model, the referential encoding hypothesis (Plack and Viemeister, 1992 a, b; Carlyon and Beveridge, 1993) has no difficulty accounting for the latter observations, because it is based on more central mechanisms. The model assumes that the masker degrades the memory trace of the target tone presented in the first observation interval, so that the listener is forced to use context coding (Durlach and Braida, 1969). This explains the midlevel hump caused by both forward and backward maskers, as referential encoding is assumed to be efficient at low and high standard levels, where the internal coding references detection threshold and discomfort level are available, while at intermediate levels, the perceptual distance to these references is too large. In previous studies, it was concluded that this model is successful in accounting for most empirical observations. In the evaluation of the referential encoding hypothesis it became evident, however, that the model can not account for the reduced midlevel humps observed if a masker differing from the standard in duration or spectral content is presented. Unlike the recovery-rate model, the referential encoding hypothesis also has difficulty explaining the smaller midlevel humps found at long masker-standard ISIs, at least for the case of a forward masker. The most serious objection against the referential encoding hypothesis, however, seems to be the discrepancy between intensity resolution in quiet observed in a small-range, one-interval paradigm on the one hand and the difference limens found under non-simultaneous masking on the other hand. As the model attributes the midlevel hump to context coding, the same level dependence of the jnd can be expected in a one-interval (absolute identification) experiment, where use of the trace mode is precluded. Yet, data from one-interval experiments presenting a small range of intensities provide no evidence for a midlevel hump. Equally important, in the intensity discrimination experiments, masker and standard level were constant within each block. Therefore, standard intensity was located at the edge of the intensity range, where context coding was demonstrated to be relatively efficient (cf. Braida and Durlach, 1988).

The loudness enhancement hypothesis (Carlyon and Beveridge, 1993) is the only model proposed for the midlevel hump in intensity discrimination that can account for the empirical similarities between loudness enhancement and intensity resolution in non-simultaneous masking. The model assumes increased loudness variability due to loudness enhancement to be the cause of the jnd elevation. As the review of the data demonstrated, there are minor differences in the effects some parameters have on the two phenomena. Problems as serious as for the previous two models do not seem to exist, however. It has to be noted, though, that the data basis that can be used for evaluation of the model is smaller, as several parameters have not been studied for loudness enhancement. For instance, it is not apparent why target frequency should have an effect on loudness enhancement. It also became evident that existing models of the relation between intensity resolution and loudness can not account for the effects of the alteration of loudness caused by a non-simultaneous masker on the intensity jnd. This pointed to the need for a formal model of the influence of loudness enhancement on the intensity difference limen.

The loudness enhancement hypothesis also raises the question of a model of loudness enhancement.

Mechanisms located in the auditory periphery, as assumed by the recovery-rate model, are incompatible with loudness of a tone not being reduced, but rather enhanced by an intense forward masker. At more central locations, an enhancement of neural responses has been observed, but there are several inconsistencies between the psychophysical and the neurophysiological data, so that the explanatory value of the physiological findings remains questionable.

Even with additional assumptions, the referential encoding hypothesis can not account for loudness enhancement.

The literature review identified the mergence hypothesis by Elmasian, Galambos, and Bernheim (1980) as a useful model for both loudness enhancement and decrement. It assumes that the listener uses a weighted average between masker and target loudness in a loudness matching experiment. In its original form, this model is incompatible with the midlevel hump in loudness enhancement, however, as it predicts loudness enhancement to be a monotonically increasing function of the masker-target level difference.

Models of assimilation and trace drift originally formulated to explain the time order error, i.e., systematic biases observed in loudness comparisons between two tones in quiet (Hellström, 1985), are virtually equivalent to the mergence hypothesis in the case of non-simultaneous masking.

Based on the reviews of data and models, a new model for loudness and intensity resolution in non-simultaneous masking was proposed in Chapter 5. This "similarity model" uses the concepts formulated in the mergence hypothesis to explain loudness enhancement. As an important extension, it introduces an effect of the perceptual similarity between masker and target on the weight assigned to masker loudness. The assumption that the weighted average between masker and target loudness is less influenced by maskers differing perceptually from the target results in the prediction of a midlevel hump. According to the model, if an intense forward or backward masker is used, masker loudness has a pronounced effect on the weighted average at high and intermediate target levels. As on the loudness dimension, a low level target and the intense masker are very different perceptually, the influence of masker loudness is small in this situation, so that loudness enhancement is predicted to be smaller at low than at intermediate target levels. The similarity model uses the loudness enhancement hypothesis by Carlyon and Beveridge (1993) and predicts the jnd elevation caused by a non-simultaneous masker to be monotonically related to the amount of loudness enhancement. The similarity-related effects of masker duration and masker spectrum that present a problem for the referential encoding hypothesis immediately follow from the similarity model.

Unlike previous models, the similarity model predicts a *mid-difference hump*: for a fixed target level, the largest amount of loudness enhancement, and therefore also the largest jnd elevation, is expected at intermediate masker-target level differences. At large positive or negative level differences, perceptual similarity between masker and target is small, so that only a small effect of the masker is predicted in these conditions. The similarity model was tested in three experiments, in which unlike in previous studies the masker-target level difference was varied independent of target level over a range of 60 dB. Results from these experiments reported in Chapter 6 supported the model, as mid-difference humps were observed for intensity discrimination (Experiment 1) as well as for loudness enhancement (Experiment 2). In Experiment 3, an additional manipulation of the perceptual similarity between masker and target by presenting a masker longer in duration than the target also had the expected effect. The observed mid-difference humps represent further evidence against the referential encoding hypothesis.

According to the similarity model, the common mechanism underlying both loudness enhancement and the jnd elevation is the inclusion of masker intensity information in the decisions made in the respective experiments. Compatible with this view, the predicted correlation between loudness enhancement and the difference limen was significant. On the other hand, some of the individual correlations were weak and the dependence of loudness enhancement and the intensity difference limen ΔL_{DL} on the masker-target level difference was distinctly different for some listeners. The prediction of loudness decrement also resulting in jnd elevations was not supported. Instead, loudness decrement corresponded to small values of ΔL_{DL}.

Surprisingly, the measure of loudness variability jnd_{Match} derived from the data obtained with a double-staircase procedure (Experiment 2) differed distinctly from ΔL_{DL} (Experiment 1). A previous study (Zeng, 1994) reported a close relation between the two measures. The difference between the two measures could not be attributed to response bias caused by the masker.

The loudness enhancement part of the similarity model was formulated in quantitative form in the first part of Chapter 7. The mathematical modeling combines well-established properties of the auditory periphery, represented by the loudness function for pure tones, with the mechanism of the influence of masker loudness on the memory representation of target loudness, which is presumably more centrally located. Fits of the model to data from experiments using a wide variety of stimuli and procedures were reasonable to excellent, even though as few as only two free parameters were used. The quantitative model could also explain why the effects of the masker were least pronounced at the *lowest* target level in Experiment 2, as this observation can be attributed to the shape of the loudness function. The model is motivation for future research because it can be used to derive detailed predictions as to, e.g., loudness enhancement for broadband noise bursts or the effects of varying the masker-target interval. It is important to note, however, that in a recent paper, Scharf, Buus, and Nieder (2002) suggested that the loudness matches obtained in a three-tone matching task must be interpreted in a completely different way. They argued that the loudness matches reflect "loudness recalibration" (that is, a reduction in loudness, cf. Marks, 1994) of the comparison tone rather than loudness enhancement of the target tone. This would explain "loudness enhancement" as defined in the three-tone matching task, i.e., comparison tone level adjusted to higher levels than the target. To explain why the target tone is not subject to loudness reduction, however, the assumption has to be made that "[…] a weak tone presented in close temporal proximity to a stronger tone somehow is protected from ILR [induced loudness reduction, DO]." (Scharf et al., 2002, p. 809). Evidence for this idea was provided by Arieh and Marks (2003), who found virtually no loudness reduction (but also no clear loudness enhancement) for conditioner-target ISIs smaller than 200 ms, but loudness reduction of up to 13 dB at longer ISIs. It remains to be shown whether all of the "loudness enhancement" data can be explained by loudness recalibration. For instance, Nieder et al. (2003) found virtually no effect of conditioner level (80 dB SPL versus 95 dB SPL) on the loudness reduction induced in a 60-dB SPL target tone. In contrast, in Experiment 2 of this study, loudness matches strongly depended on the masker-target level difference. Equally important, the observation of loudness *decrement*, i.e., the loudness of the target being matched by a comparison tone *lower* in level than the target, could be explained by loudness recalibration only if an inducer tone less intense than the test tone produced an *increase* rather than a reduction in test tone loudness. Mapes-Riordan and Yost (1999) found virtually no effect of a 40-dB SPL inducer on the loudness of an 80-dB SPL test tone, but the the level difference between inducer and test tone may have been too large to produce loudness decrement. To resolve these issues, additional experiments are necessary.

In the second part of Chapter 7, a formal signal-detection model was used to predict the increase in the difference limen (ΔL_{DL}) caused by mergence, i.e., by the inclusion of masker information in the decision in an intensity discrimination experiment. The seemingly simple relation assumed by the loudness enhancement hypothesis (Carlyon and Beveridge, 1993), according to which loudness enhancement increases loudness variability and the jnd elevation is therefore monotonically related to the amount of loudness enhancement, could not be derived from the SDT model. Neither the contribution of masker sensation noise to the variance of the decision variable, nor the explicit inclusion of noise introduced by the process of mergence resulted in the predictions of the model being at least qualitatively compatible with the data. Depending on the assumed dependence of the weight assigned to masker loudness on the masker-standard level difference, the model predicted either a monotonic increase of ΔL_{DL} with masker intensity, or a maximum of the jnd elevation located at masker levels close to target level. Mid-difference humps were predicted only if it was assumed that the masker introduces a substantial response bias. In this case, the model predicted jnd elevations to be larger at masker levels *below* than above standard level, however. All of these patterns are incompatible with the psychophysical data. It has to be concluded that is surprisingly difficult to formulate a formal model accounting for the simple idea of mergence causing a drop in performance in an intensity discrimination task.

The review of psychophysical data and potential models provided some arguments against loudness enhancement as an explanation for the interval-produced accent observed in equitone sequences. In loudness matching experiments, the effect of maskers equal in level to the target tone was smaller than the interval-produced accent reported by Povel and Okkerman (1981). It was pointed out, however, that this comparison must be viewed with caution as the two types of experiments do not measure the same thing. A more important argument is the conclusion that “forward enhancement” of the second by the first sound of a pair should partly be compensated for by “backward enhancement” of the first by the second sound. However, the psychophysical data are too limited to decide whether the temporal asymmetry in loudness enhancement is sufficiently large to account for the interval-produced accent. Finally, the mergence hypothesis predicts loudness enhancement to be absent in a pair of identical tones. Data from the only two studies available for the accent seemed to be compatible with a model proposed by Povel and Okkerman (1981), according to which processing of the first sound in auditory sensory memory is interrupted by presentation of the second tone. As the model assumes loudness to increase with processing time, and the intra-pair interval is shorter than the inter-pair interval, interrupted processing results in reduced loudness of the first tone.

As neither the proportion of accents reported on the second sounds, nor the loudness match between first and second sounds can be used directly to decide between the two models, the time course and the level dependence of the accent were studied in two experiments.

Experiment 4 varied the temporal structure of the sequences. The relative difference between inter-pair and the intra-pair interval was varied independent of the duration of the shorter interval. Significant accents on the second tones of the pairs were observed for intra-pair intervals as long as 800 ms, which is considerably larger than the temporal limit of 300 ms proposed by Povel and Okkerman (1981). There was also a systematic effect of the starting condition of the sequences. To avoid many problems associated with the estimation of time constants, a quantitative model was used in Chapter 8 to derive the time constant effective for the interval-produced accent observed in Experiment 4 and in the study by Povel and Okkerman (1981). The estimates indicated the temporal limit of accent perception to be larger than both the duration of the *short* auditory store, which is assumed to be relevant for interrupted processing, and the maximum conditioner-target interval at which loudness enhancement is observed. A potential explanation for the larger temporal limit observed in the accent-perception experiments is the presentation of multiple tone pairs in the short-long sequences, which generates a ‘multiple looks’ situation.

The effect of presentation level on the interval-produced accent was studied in Experiment 5. A modern loudness-matching paradigm was used. The fact that the accent was virtually absent at a presentation level of 85 dB SPL can not be accounted for by the interrupted processing model that makes no assumptions regarding the effects of presentation level. Similarly, the mergence hypothesis, which represents the most successful model of loudness enhancement, can neither account for non-zero matches if masker and target are physically identical, nor for the level dependence of the accent observed in Experiment 5.

For two listeners, the level increment required to make the first tones sound equally loud as the second tones decreased monotonically with increasing short interval duration, reaching a value close to zero in the isochronous condition. These results are equivalent to the pattern found in Experiment 4. Absolutely unexpected, data produced by the remaining two listeners indicated an accent on the *first* tones, i.e., exactly the opposite pattern than the interval-produced accent. This result could only in part be attributed to procedural problems.

Taken together, neither loudness enhancement, nor the interrupted processing model could account for the complete range of findings in Experiments 4 and 5. The results confirm the observation from Experiment 2 that loudness perception in pairs of identical tones is difficult to predict. This conclusion applies to loudness matches between two tones in quiet (the time order error; cf. Hellström, 1985) and loudness matches in the three-tone matching task with identical masker and target. Loudness perception in equitone pairs presented rhythmically again differs from the patterns found for a single pair of sounds.

10 Zusammenfassung

Der Schritt von einem einzelnen, isoliert präsentierten Ton zu zwei Tönen, die zeitlich eng benachbart dargeboten werden, führt zu einer Reihe von interessanten und komplexen Effekten auf die auditorische Intensitätsverarbeitung.

Wird ein kurzer Ton zusammen mit einem Vorwärts- oder Rückwärtsmaskierer hoher Intensität und einem Inter-Stimulus Intervall (ISI) von 50 bis 400 ms dargeboten, so unterscheidet sich die Abhängigkeit der Differenzschwelle (just-noticeable difference, JND) von der Standardintensität deutlich vom „near-miss to Weber's law", das für unmaskierte Töne beobachtet wird. Die Differenzschwellen für Standards mittlerer Intensität sind gegenüber den JNDs „in quiet" deutlich erhöht, während der Maskierer bei niedrigen und hohen Standardpegeln nur einen geringen Effekt hat. Dieser „midlevel hump in intensity discrimination" wurde zuerst von Zeng, Turner und Relkin (1991) beobachtet.

Der nicht-simultane Maskierer bewirkt zudem eine systematische Verschiebung der Lautheit des Zieltons in Richtung der Lautheit des Maskierers. Die Lautheit eines Zieltons wird durch einen Maskierer mit höherem Schalldruckpegel erhöht („loudness enhancement", z. B. Zwislocki und Sokolich, 1974). Durch einen Maskierer, dessen Intensität niedriger als die Zieltonintensität ist, wird die Lautheit des Zieltons herabgesetzt („loudness decrement").

In rhythmischen Sequenzen, die aus paarweise angeordneten physikalisch identischen Klängen bestehen, in denen also das Intra-Paar Intervall kürzer als das Inter-Paar Intervall ist, werden die zweiten Töne der Tonpaare als akzentuiert wahrgenommen. D. h., die Lautheit der zweiten Töne ist höher als die Lautheit der ersten Töne. Dieser „intervallproduzierte Akzent" (Povel und Okkerman, 1981) findet sich nur, falls das Intra-Paar Intervall kürzer als etwa 800 ms und gleichzeitig die Differenz zwischen Inter- und Intra-Paar Intervall hinreichend groß ist.

Ziel dieser Arbeit war die Identifikation der Mechanismen, die den drei eben geschilderten Phänomenen zugrunde liegen. Motiviert durch eine Vielzahl von Ähnlichkeiten zwischen den Effekten eines nicht-simultanen Maskierers auf Lautheit und Intensitätsauflösung, wie beispielsweise die Beobachtung eines midlevel-humps sowohl in loudness enhancement- als auch in Intensitätsdiskriminationsexperimenten, wurde dabei speziell der Idee nachgegangen, beide Phänomene könnten durch einen *gemeinsamen Mechanismus* erklärt werden. Ebenso wurde die Frage aufgeworfen, ob der intervallproduzierte Akzent als ein Spezialfall von loudness enhancement interpretiert werden kann, d. h., ob derselbe Mechanismus, der die Erhöhung der Lautheit eines Zieltons durch einen intensiven Maskierer bewirkt, auch dem intervallproduzierten Akzent zugrunde liegt.

In Kapitel 3 ergab eine systematische Begutachtung der vorhandenen psychoakustischen Daten weitere Evidenz für die Ähnlichkeit der Phänomene, obwohl Unterschiede im Einfluss der Tondauer des Maskierers und der Präsentation von Vorwärts- bzw. Rückwärtsmaskierern festgestellt wurden.

Für den Effekt nicht-simultaner Maskierung auf die Intensitätsauflösung wurden drei Modelle vorgeschlagen. Auf der Basis des Überblicks über die vorhandenen Daten wurden diese Modelle in Kapitel 4 einer kritischen Evaluation unterzogen.

Das „recovery-rate“ Modell von Zeng et al. (1991) führt den midlevel hump auf unterschiedliche Erholungsraten (recovery rates) von Neuronen im Hörnerv mit hoher bzw. niedriger spontaner Entladungsrate dar. Es kann damit in eleganter Weise die grundlegenden Effekte eines Vorwärtsmaskierers auf die Unterschiedsschwelle und außerdem die recht komplexen Interaktionen zwischen simultaner und nicht-simultaner Maskierung erklären. Wie bereits in anderen Studien festgestellt wurde, muss das Modell jedoch verworfen werden, da es inkompatibel mit den unter Rückwärtsmaskierung und mit kontralateral präsentierten Maskierern beobachteten midlevel humps ist.

Die „referential encoding hypothesis“ (Plack und Viemeister, 1992 a, b; Carlyon und Beverdige, 1993) kann problemlos die beiden letztgenannten Effekte erklären, da sie auf zentraleren Verarbeitungsmechanismen basiert. Das Modell nimmt an, dass ein Maskierer die Gedächtnisspur (memory trace) des Zielreizes beeinträchtigt, der im ersten Beobachtungsintervall präsentiert wird. Deshalb muss der Hörer in den „context coding mode“ (Durlach und Braida, 1969) wechseln. Man nimmt an, dass das in diesem Modus angewendetete „referential encoding“ bei niedrigen und hohen Intensitäten effektiv funktioniert, da hier die internen Referenzen Absolutschwelle und Unbehaglichkeitsschwelle zur Verfügung stehen. Dahingegen ist die perzeptuelle Distanz zu diesen Referenzen bei mittleren Standardintensitäten zu hoch, so dass ein midlevel hump resultiert. In früheren Studien wurde der Schluss gezogen, das Modell sei kompatibel mit den meisten empirischen Daten. Bei der Evaluation der referential encoding hypothesis zeigte sich jedoch, dass das Modell nicht erklären kann, weshalb ein in Tondauer oder Frequenzspektrum vom Standard abweichender Maskierer zu einem weniger ausgeprägten midlevel hump führt. Anders als das recovery-rate Modell hat die referential encoding hypothesis außerdem Schwierigkeiten, die Abnahme des humps mit dem Maskierer-Standard ISI zu erklären, zumindest für den Fall eines Vorwärtsmaskierers. Der wichtigste Einwand gegen das Modell scheint jedoch die Diskrepanz zwischen der „in quiet“ beobachteten Intensitätsauflösung in einem one-interval Paradigma und den unter nicht-simultaner Maskierung gemessenen Differenzschwellen zu sein. Da das Modell den midlevel hump auf context coding zurückführt, sollte sich dieselbe Abhängigkeit von der Standardintensität finden, wenn eine one-interval Aufgabe (also eine Identifikationsaufgabe) verwendet wird, da in diesem Fall die Verwendung des context coding Modus zwingend ist. Man beobachtet jedoch in solchen Experimenten keinen midlevel hump, wenn die Spanne der im Experiment präsentierten Intensitäten klein ist. Ein ebenso wichtiger Punkt ist die Tatsache, dass in den relevanten Intensitätsdiskriminationsexperimenten die Pegel von Maskierer und Standard in jedem Block konstant waren. Deshalb befand sich die Standardintensität an einem Extrem der Intensitätsspanne, wo context coding relativ effizient funktioniert (vgl. Braida und Durlach, 1988).

Die loudness enhancement Hypothese (Carlyon und Beveridge, 1993) ist das einzige Modell, das auch die empirischen Ähnlichkeiten zwischen loudness enhancement und Intensitätsauflösung unter nicht-simultaner Maskierung erklären kann. Im Rahmen des Modells wird angenommen, dass eine durch loudness enhancement versachte Lautheitsvariabilität die Ursache der erhöhten Differenzschwellen ist. Wie die Übersicht über die psychoakustischen Daten zeigte, gibt es kleinere Unterschiede im Effekt einzelner Parameter auf die beiden Phänomene. Ähnlich schwerwiegende Probleme wie für die beiden anderen Modelle scheinen jedoch nicht zu existieren. Einschränkend ist jedoch anzumerken, dass der Effekt verschiedenen Parameter auf loudness enhancement nicht untersucht wurde, so dass für die Evaluation des Modells weniger Daten zu Verfügung stehen. Beispielsweise ist nicht offensichtlich, weshalb die Frequenz des Zieltons einen Einfluss auf loudness enhancement haben sollte. Es wurde außerdem klar, dass existierende Modelle für den Zusammenhang zwischen Lautheit und Intensitätsauflösung die Effekte der durch einen nicht-simultanen Maskierer hervorgerufen Lautheitsveränderung auf die Differenzschwelle nicht erklären können. Dies deutete auf die Notwendigkeit hin, ein formales Modell für den Einfluss von loudness enhancement auf die Intensitätsauflösung zu finden.

Die loudness enhancement Hypothese wirft zudem die Frage nach einem Modell für loudness enhancement auf.

Mechanismen in der auditorischen Peripherie sind inkompatibel mit der Tatsache, dass ein intensiver Vorwärtsmaskierer die Lautheit eines Tons nicht reduziert, sondern erhöht. Auf zentralerer Ebene wurde Enhancement neuronaler Reaktionen beobachtet. Es gibt jedoch verschiedene Inkonsistenzen zwischen den psychophysischen und den neurophysiologischen Daten, so dass der Erklärungswert der physiologischen Daten unklar bleibt.

Die referential encoding hypothesis kann auch mit zusätzlichen Annahmen loudness enhancement nicht erklären.

In der Diskussion von Daten und Modellen wurde „mergence hypothesis“ (Elmasian, Galambos und Bernheim, 1980) als sinnvolles Modell sowohl für loudness enhancement, als auch für loudness decrement identifiziert. Die „mergence hypothesis“ nimmt an, dass die Hörer in einem loudness matching Experiment einen gewichteten Mittelwert zwischen der Lautheit des Maskierers und der Lautheit des Zieltons verwenden. In seiner ursprünglichen Form ist das Modell jedoch inkompatibel mit dem für loudness enhancement beobachteten midlevel hump, da es einen montonen Anstieg der Lautheitserhöhung mit der Maskierer-Zielton Pegeldifferenz vorhersagt.

Modelle für Assimilation und trace drift, die ursprünglich formuliert wurden, um den Zeitfehler, also systematische Urteilstendenzen in Lautheitsvergleichen zwischen zwei Tönen „in quiet“, zu erklären (Hellström, 1985), sind im Fall von nicht-simultaner Maskierung nahezu äquivalent zur Mergence Hypothese.

Auf der Basis der Übersicht über Daten und Modelle wurde in Kapitel 5 ein neues Modell für Intensitätsauflösung und Lautheit unter nicht-simultaner Maskierung vorgeschlagen. Dieses „Ähnlichkeitsmodell“ verwendet das Konzept der Mergence Hypothese, um loudness enhancement zu erklären. Als wichtige Erweiterung wurde ein Effekt der perzeptuellen Ähnlichkeit zwischen Maskierer und Zielton in das Modell eingeführt. Die Annahme, dass der gewichtete Mittelwert zwischen Maskierer- und Zieltonlautheit nur in geringem Maße durch einen Maskierer beeinflusst wird, der sich perzeptuell vom Zielton unterscheidet, führt zur Vorhersage eines midlevel hump. Laut Modell hat die Maskiererlautheit einen ausgeprägten Effekt auf den gewichteten Mittelwert, wenn ein intensiver Vorwärts- oder Rückwärtsmaskierer zusammen mit einem Zielton mittleren oder hohen Pegels präsentiert wird. Da ein niedrigpegeliger Zielton und ein intensiver Maskierer auf der Lautheitsdimension sehr unterschiedlich sind, ist der Einfluss der Maskiererlautheit in dieser Situation gering, so dass weniger loudness enhancement als bei mittleren Zieltonpegeln vorhergesagt wird. Das Ähnlichkeitsmodell verwendet die loudness enhancement Hypothese von Carlyon und Beveridge (1993) und sagt vorher, dass die JND-Erhöhung durch einen nicht-simultanen Maskierer eine monoton steigende Funktion von loudness enhancement ist. Die mit perzeptueller Ähnlichkeit in Verbindung stehenden Effekte von Maskierertondauer und -spektrum, die ein Problem für die referential encoding hypothesis darstellen, können unmittelbar aus dem Ähnlichkeitsmodell abgeleitet werden.

Anders als bisherige Modelle sagt das Ähnlichkeitsmodell einen mid-difference hump vorher. Bei gegebenem Zieltonpegel wird maximales loudness enhancement, und somit auch die größte Erhöhung der Unterschiedsschwelle, für mittlere Maskierer-Zielton Pegelunterschiede vorhergesagt.

Bei großen positiven oder negativen Pegelunterschieden ist die perzeptuelle Ähnlichkeit gering, so dass in diesem Fall nur ein kleiner Effekt des Maskierers erwartet wird. Das Modell wurde in drei Experimenten getestet, in denen anders als in bisherigen Untersuchungen die Pegeldifferenz zwischen Maskierer und Zielton/Standard *unabhängig* vom Zieltonpegel über einen Bereich von 60 dB variiert wurde. Die in Kapitel 6 vorgestellten Ergebnisse stützen das Modell, da mid-difference humps sowohl für Intensitätsdiskrimination (Experiment 1), als auch für loudness enhancement (Experiment 2) gefunden wurden. In Experiment 3 hatte die zusätzliche Variation der perzeptuellen Ähnlichkeit zwischen Maskierer und Zielton durch die Verwendung einer längeren Maskierertondauer ebenfalls den erwarteten Effekt. Die beobachteten mid-difference humps sind weitere Evidenz gegen die referential encoding Hypothese.

Laut Ähnlichkeitsmodell ist der sowohl loudness enhancement als auch der Differenzschwellenerhöhung zugrunde liegende gemeinsame Mechanismus die Einbeziehung von Maskiererinformation in die Entscheidungen, die in den jeweiligen Experimenten getroffen werden. Kompatibel mit dieser Sichtweise war die vorhergesagte Korrelation zwischen loudness enhancement und der Differenzschwelle signifikant. Andererseits waren einige der individuellen Korrelationen eher gering und der Zusammenhang zwischen loudness enhancement bzw. der Differenzschwelle ΔL_{DL} und der Pegeldifferenz zwischen Maskierer und Zielreiz war für einige Hörer deutlich verschieden. Darüber hinaus fand sich kein Beleg für die vorhergesagte Differenzschwellenerhöhung bei loudness decrement. Vielmehr war ΔL_{DL} im Fall von loudness decrement klein.

Überraschenderweise unterschied sich das Maß für Lautheitsvariabilität, das aus den mittels eines double-staircase Verfahrens erhobenen Daten (Experiment 2) abgeleitet wurde, stark von den in Experiment 1 gemessenen Differenzschwellen. In einer Studie von Zeng (1994) wurde eine enge Beziehung zwischen den beiden Maßen gefunden. Der Unterschied zwischen den Maßen konnte nicht auf einen durch den Maskierer hervorgerufenen Bias zurückgeführt werden.

Der für die Erklärung von loudness enhancement zuständige Teil des Ähnlichkeitsmodells wurde im ersten Abschnitt von Kapitel 7 in quantitativer Form formuliert. Die mathematische Modellierung kombiniert bekannte Eigenschaften der auditorischen Peripherie, repräsentiert durch die Lautheitsfunktion für Sinustöne, mit dem vermutlich auf zentralerer Ebene angesiedelten Mechanismus des Einflusses der Maskiererlautheit auf die Gedächtnisrepräsentation der Zieltonlautheit. Bei der Anpassung des Modells an experimentelle Daten, die für verschiedene Stimuli und mit unterschiedlichen experimentellen Verfahren erhoben wurden, war die Anpassungsgüte zufriedenstellend bis exzellent, obwohl teilweise nur zwei freie Parameter verwendet wurden. Das quantitative Modell konnte außerdem erklären, weshalb die Effekte des Maskierers in Experiment 2 bei geringen Zieltonpegeln am kleinsten waren. Diese Beobachtung kann auf die Form der Lautheitsfunktion zurückgeführt werden. Das Modell ist Motivation für zukünftige Untersuchungen, da aus ihm detaillierte Vorhersagen, z. B. bezüglich loudness enhancement für kurze Rauschsignale oder den Effekt des Maskierer-Zielton Intervalls, abgeleitet werden können. Allerdings muss an dieser Stelle darauf hingewiesen werden, dass Scharf, Buus und Nieder (2002) kürzlich vorgeschlagen haben, dass die in einer 3-Ton Matching Aufgabe erhobenen loudness matches anders als bisher interpretiert werden müssen. Sie argumentierten, dass die loudness matches "loudness recalibration" (d.h., eine Reduktion der Lautheit, siehe Marks, 1994) des Vergleichstons zeigen, und nicht –wie auch in dieser Arbeit angenommen– loudness enhancement des Zieltons. Dieser Ansatz kann "loudness enhancement", so wie es in der 3-Ton Matching Aufgabe definiert ist, erklären (Vergleichston wird auf höhere Pegel als der Zielton eingestellt). Um zu erklären, warum der Zielton selbst nicht von loudness recalibration betroffen ist, muss allerdings angenommen werden, dass “[…] a weak tone presented in close temporal proximity to a stronger tone somehow is protected from ILR [loudness recalibration, DO].” (Scharf et al., 2002, S. 809). Daten von Arieh und Marks (2003) unterstützen diese Hypothese. Sie fanden nahezu keine Lautheitsreduktion (aber auch kein eindeutiges loudness enhancement) bei Maskierer-Zielton Inetrvallen kleiner als 200 ms, aber bis zu 13 dB Lautheitsreduktion bei längeren Interstimulusintervallen. Es wird noch zu zeigen sein, ob alle der bisherigen “loudness enhancement” Daten durch loudness recalibration erklärt werden können. Beispielsweise fanden Nieder et al. (2003) nahezu keinen Effekt des Konditioniererpegels (80 dB SPL versus 95 dB SPL) auf die in einem 60 dB SPL Zielton induzierte Lautheitsreduktion. Dahingegen hingen die in Experiment 2 der vorliegenden Studie gemessenen loudness matches stark von der Masker-Zielton Pegeldifferenz ab. Ähnlich bedeutsam ist die Tatsache, dass die Beobachtung von loudness *decrement* (die Lautheit des Zieltons entspricht der Lautheit eines Vergleichstons mit *niedrigerem* Pegel) nur dann durch loudness recalibration erklärt werden könnte, wenn ein Konditionierer mit niedrigerem Pegel als der Testton eine *Anhebung* (und nicht eine Reduktion) der Lautheit des Testtons zur Folge hätte. Mapes-Riordan und Yost (1999) fanden nahezu keinen Effekt eines 40 dB SPL Konditionierers auf die Lautheit eines 80 dB SPL Testtons. Allerdings war dort eventuell die Pegeldifferenz zwischen Konditionierer und Testton zu groß, um loudness decrement hervorzurufen. Zur Klärung dieser Punkte sind weitere Experimente notwendig.

Im zweiten Abschnitt von Kapitel 7 wurde ein formales Signalentdeckungsmodell verwendet, um die Erhöhung der Differenzschwelle durch Mergence, also durch die Einbeziehung von Maskiererinformation in die Entscheidung in einem Diskriminationsexperiment, zu erklären. Die loudness enhancement Hypothese (Carlyon und Beveridge, 1993) basiert auf der Annahme, dass loudness enhancement die Lautheitsvariabilität erhöht und die Erhöhung der Differenzschwelle deshalb mit dem Ausmaß von loudness enhancement steigt. Ein dieser auf den ersten Blick einfachen Idee entsprechender Zusammenhang konnte jedoch aus dem Signalentdeckungsmodell nicht abgeleitet werden. Weder der Beitrag von mit dem Maskierer assoziiertem „sensation noise“ (Durlach and Braida, 1969), noch die explizite Berücksichtigung von durch die Mittelung zwischen Maskierer- und Zieltonlautheit hervorgerufener Variabilität führte zu Vorhersagen des Modells, die zumindest qualitativ mit den psychophysischen Daten kompatibel waren. Je nach angenommenem Zusammenhang zwischen der Gewichtung der Maskiererlautheit und der Pegeldifferenz zwischen Maskierer und Standard sagte das Modell entweder einen monotonen Anstieg der Differenzschwelle mit der Maskiererintensität oder aber ein Maximum der JND-Erhöhung für Maskiererpegel nahe dem Standardpegel vorher. Mid-difference humps wurden nur vorhergesagt, wenn ein starker, durch den Maskierer hervorgerufener Bias angenommen wurde. In diesem Fall folgt laut Modell jedoch eine stärkere JND-Erhöhung bei Maskiererintensitäten *unterhalb* der Standardintensität als bei denjenigen oberhalb der Standardintensität. Alle vorhergesagten Muster sind somit nicht mit den Daten vereinbar. Es muss festgestellt werden, dass es überraschend schwierig ist, ein formal-mathematisches Modell für die einfache Idee zu formulieren, Mergence führe zu einer geringeren Diskriminationsleistung.

Die Übersicht über psychophysische Daten und mögliche Modelle identifizierte einige Argumente gegen loudness enhancement als Ursache für den intervallproduzierten Akzent. In loudness matching Experimenten war der Effekt von Maskierern, die mit dem Zielton identisch waren, geringer als der von Povel und Okkerman (1981) berichtete Akzent. Es wurde allerdings darauf hingewiesen, dass dieser Vergleich mit Vorsicht interpretiert werden muss, da die beiden Typen von

Experimenten nicht exakt dasselbe messen. Ein wichtigeres Argument ist, dass „forward enhancement“ des zweiten Tons durch den ersten Ton eines Tonpaars teilweise durch „backward enhancement“ des ersten Tons durch den zweiten Ton kompensiert werden sollte. Allerdings liegen keine hinreichend präzisen psychophysischen Daten vor, um entscheiden zu können, ob die zeitliche Asymmetrie von loudness enhancement groß genug ist, um den Akzent erklären zu können. Schließlich wurde darauf hingewiesen, dass die Mergence für den Fall von identischem Maskierer und Zielreiz keine Lautheitsveränderung vorhersagt. Die Daten aus den beiden einzigen bisher verfügbaren Studien zum intervallproduzierten Akzent schienen kompatibel mit einem vom Povel und Okkerman (1981) vorgeschlagenem Modell zu sein, nach dem die Unterbrechung der Verarbeitung des ersten Tons im kurzen auditorischen Speicher durch die Präsentation des zweiten Tons unterbrochen wird. Da das Modell annimmt, die Lautheit wachse mit der zur Verfügung stehenden Verarbeitungszeit, führt die Verarbeitungsunterbrechung zu reduzierter Lautheit des ersten Tons.

Weder die Häufigkeit eines auf den zweiten Tönen wahrgenommen Akzents, noch der loudness match zwischen ersten und zweiten Tönen kann verwendet werden, um direkt zwischen den beiden alternativen Modellen zu entscheiden. Deshalb wurden der zeitliche Verlauf und die Pegelabhängigkeit des Akzents in zwei Experimenten untersucht.

In Experiment 4 wurde die zeitliche Struktur der kurz-lang Sequenzen variiert. Der relative Unterschied zwischen Inter- und Intra-Paar Intervall wurde unabhängig von der Dauer des kürzeren Intervalls variiert. Signifikante Akzente auf den zweiten Tönen wurden für Intra-Paar Intervalle von bis zu 800 ms beobachtet. Dieser Wert ist bedeutend länger als der von Povel und Okkerman (1981) vorgeschlagene zeitliche Grenzwert von 300 ms. Es gab zudem einen systematischen Effekt der Startposition der Sequenzen. Um viele der mit der Schätzung von Zeitkonstanten verbundenen Probleme zu vermeiden, wurde ein quantitatives Modell verwendet, um die für den in Experiment 4 und in der Studie von Povel und Okkerman (1981) gefundenen Akzent relevanten Zeitkonstanten abzuleiten. Die geschätzten Werte deuten darauf hin, dass der zeitliche Grenzwert des Akzents deutlich länger ist als die Dauer des *kurzen* auditorischen Speichers (der für die Verarbeitung laut Modell relevant ist) und das maximale Maskierer-Zielton Intervall, für das loudness enhancement gefunden wurde. Eine mögliche Erklärung für in Experimenten zum intervallproduzierten Akzent gefundenen längeren zeitlichen Grenzwert ist die Darbietung mehrerer Tonpaare in den kurz-lang Sequenzen, wodurch eine ‚multiple looks’ Situation erzeugt wird.

Der Effekt des Schalldruckpegels auf den intervallproduzierten Akzent wurde in Experiment 5 mit einem modernen loudness matching Verfahren untersucht. Die Tatsache, dass der Akzent bei einem Pegel von 85 dB SPL verschwindend klein war, stellt ein Problem für das Modell der Verarbeitungsunterbrechung dar, das keine Annahmen über den Einfluss des Pegels macht. Ebenso kann die Mergence Hypothese, die das momentan beste Modell für loudness enhancement darstellt, weder die von Null verschiedenen loudness matches zwischen identischen Tönen, noch die in Experiment 5 beobachtete Pegelabhängigkeit des Akzents erklären.

Für zwei Versuchsteilnehmer nahm das Pegelinkrement, das notwendig war, damit die ersten Töne als ebenso laut wie die zweiten Töne wahrgenommen wurden, mit steigendem Intra-Paar Intervall ab. In der isochronen Bedingung war das notwendige Inkrement praktisch Null. Diese Ergebnisse entsprechen dem in Experiment 4 gefundenen Zusammenhang. Absolut unerwartet zeigten die Daten der beiden anderen Versuchspersonen einen Akzent auf den *ersten* Tönen, also ein dem intervallproduzierten Akzent exakt konträres Muster. Dieser Befund konnte nur teilweise auf prozedurale Probleme zurückgeführt werden.

Abschließend muss festgestellt werden, dass weder loudness enhancement noch das auf dem Konzept unterbrochener Verarbeitung basierende Modell die gesamte Bandbreite der in den Experimenten 4 und 5 erhobenen Daten erklären kann. Dieser Befund bestätigt die in Eperiment 2 gemachte Beobachtung, dass Lautheitswahrnehmung in Paaren identischer Töne sich einfachen Erklärungen entzieht. Diese Feststellung bezieht sich auf loudness matches zwischen zwei Tönen „in quiet“ (also Experimenten zum Zeitfehler; vgl. Hellström, 1985) und loudness matches im Fall von identischem Maskierer und Zielton. Lautheitswahrnehmung in rhythmisch präsentierten Paaren physikalisch identischer Töne unterscheidet sich wiederum von den für einzelne Tonpaare gefundenen Mustern.

References

Abbas, Paul J. (**1979**). "Effects of stimulus frequency on adaptation in auditory-nerve fibers," J. Acoust. Soc. Am. **65**, 162-165.

Ananthanarayan, A. K., and Gerken, G. M. (**1983**). "Post-stimulatory effects on the auditory brain stem response: Partial-masking and enhancement," Electroencephalogr. Clin. Neurophysiol. **55,** 223-226.

Arieh, Y., and Marks, L. E. (**2003**). "Time course of loudness recalibration: Implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.

Baddeley, A. D. (**1997**). *Human Memory: Theory and Practice* (Psychology Press: Howe).

Bauer, J. W., Elmasian, R. O., and Galambos, R. (**1975**). "Loudness enhancement in man. I. Brainstem-evoked response correlates," J. Acoust. Soc. Am. **57**, 165-171.

Berg, B. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am., **86**, 149-158.

Berliner, J. E., and Durlach, N. I. (**1973**). "Intensity perception. IV. Resolution in roving-level discrimination," J. Acoust. Soc. Am. **53**, 1270–2287.

Berliner, J. E., Durlach, N. I., and Braida, L. D. (**1977**). "Intensity perception. VII. Further data on roving-level discrimination and the resolution and bias edge effects," J. Acoust. Soc. Am. **61**, 1577–1585.

Berliner, J. E., Durlach, N. I., and Braida, L. D. (**1978**). "Intensity perception. IX. Effect of a fixed standard on resolution in identification," J. Acoust. Soc. Am. **64,** 687-689.

Boker, S. M., and Kubovy, M. (**1998**). "Predicting the grouping of rhythmic sequences using local estimators of information content," in *Computational Auditory Scene Analysis*, edited by D. F. Rosenthal and H. G. Okuno (Lawrence Erlbaum Associates, Mahwah, NJ), 309-319.

Braida, L. D., and Durlach, N. I. (**1972**). "Intensity perception. II. Resolution in one-interval paradigms," J. Acoust. Soc. Am. **51**, 483-502.

Braida, L. D., and Durlach, N. I. (**1988**). "Peripheral and central factors in intensity perception," in *Auditory function*, edited by G. M. Edelman, W.E. Gall, and W. M. Cowan (Wiley, New York).

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context coding," J. Acoust. Soc. Am. **76**, 722-731.

Bregman, Albert S. (**1990**). *Auditory scene analysis: the perceptual organization of sound* (MIT Press, Cambridge, Mass.).

Brosch, M., Schulz, A., and Scheich, H. (**1998**). "Neuronal mechanisms of auditory backward recognition masking in macaque auditory cortex," Neuroreport **9**, 2551-2555.

Brugge, J. F., Merzenich, M. M. (**1973**). "Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation," J. Neurophysiol. **36**, 1138-1158.

Budd, T. W., and Michie, P. T. (**1994**). "Facilitation of the N1 peak of the auditory ERP at short stimulus intervals," Neuroreport **5**, 2513–2516.

Button, D., and Davids, K. (**2004**). “Acoustic information for timing,” in *Time-to-Contact*, edited by H. Hecht and G. J. P. Savelsbergh (Elsevier, Amsterdam), 355-369.

Buus, S., and Florentine, M. (**1991**). “Psychometric functions for level discrimination,“ J. Acoust. Soc. Am. **90**, 1371–1380.

Buus, S., and Florentine, M. (**2001**). “Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered,” JARO **3**, 120–139.

Buus, S., Florentine, M., and Poulsen, T. (**1997**). “Temporal integration of loudness, loudness discrimination, and the form of the loudness function,“ J. Acoust. Soc. Am. **101**, 669–680.

Buus, S., Florentine, M., and Zwicker, T. (**1995**). “Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking,“ J. Acoust. Soc. Am. **98**, 853–861.

Carlyon, R. P., and Beveridge, H. A. (**1993**). “Effects of forward masking on intensity discrimination, frequency discrimination, and the detection of tones in noise”, J. Acoust. Soc. Am. **93**, 2886-2895.

Carlyon, R. P., and Moore, B. C. J. (**1984**). „Intensity discrimination: A severe departure from Weber’s law,“ J. Acoust. Soc. Am. **76**, 1369-1376.

Carlyon, R. P., and Moore, B. C. J. (**1986**a). “Continuous versus gated pedestals and the ‘severe departure’ from Weber’s law,” J. Acoust. Soc. Am. **79**, 453-460.

Carlyon, R. P., and Moore, B. C. J. (**1986**b). “Detection of tones in noise and the ‘severe departure’ from Weber’s law,” J. Acoust. Soc. Am. **79**, 461-464.

Chase, S., Bugnacki, P., Braida, L. D., and Durlach, N. I. (**1983**). “Intensity perception. XII. Effect of presentation probability on absolute identification,” J. Acoust. Soc. Am. **73**, 279-284.

Chatterjee, M., and Zwislocki, J. J. (**1998**). “Cochlear mechanisms of frequency and intensity coding. II. Dynamic range and the code for loudness,” Hear. Res. **124**, 170-181.

Cooper, G. W., and Meyer, L. B. (**1960**). *The Rhythmic Structure of Music* (University of Chicago Press, Chicago, Illinois).

Costalupes, J. A., Young, E. D., and Gibson, D. J. (**1984**). „Effects of continuous noise backgrounds on rate responses of auditory nerve fibers in cat,“ J. Neurophysiol. **51**, 1326–1344.

Cowan, N. (**1984**). “On short and long auditory stores,” Psychological Bulletin **96**, 341-370.

Cowan, N. (**1987**). “Auditory sensory storage in relation to the growth of sensation and acoustic information extraction,” J. Exp. Psychol. Hum. Percept. Perform. **13,** 204-215.

Cowan, N. (**1988**). “Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system,” Psychological Bulletin **104**, 163-191.

Crowder, R. G. (**1976**). *Principles of Learning and Memory* (Erlbaum, Hillsdale/NY).

Crowder, R. G., and Morton, J. (**1969**). “Precategorical acoustical storage (PAS),” Percept. Psychophys. **5**, 365-373.

Darwin, C. J., Turvey, M. T., and Crowder, R. G. (**1972**). “An auditory analog of the Sperling partial report procedure: Evidence for brief auditory storage,” Cognitive Psychology **3**, 255-267.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). “A quantitative model of the effective signal processing in the auditory system. I. Model structure,“ J. Acoust. Soc. Am. **99**, 3615–3622.

Delgutte, B. (**1987**). “Peripheral auditory processing of speech information: implications from a physiological study of intensity discrimination,” in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht), 333-353.

Durlach, N. I., and Braida, L. D. (**1969**). “Intensity perception. I. Preliminary theory of intensity resolution,” J. Acoust. Soc. Am. **46**, 372-383.

Eddins, D. A., and Green, D. M. (**1995**). “Temporal integration and temporal resolution,” in *Hearing*, edited by B. C. J. Moore (Academic Press, San Diego), 207-242.

Efron, R. (**1970**). “The relationship between the duration of a stimulus and the duration of a perception,” Neuropsychologia **8**, 37-55.

Elmasian, R., and Galambos, R. (**1975**). “Loudness enhancement: monaural, binaural, and dichotic,” J. Acoust. Soc. Am. **58**, 229-234.

Elmasian, R., Galambos, R., and Bernheim, A. (**1980**). “Loudness enhancement and decrement in four paradigms,” J. Acoust. Soc. Am. **67**, 601-607.

Evans, E. F, and Palmer, A. R. (**1980**). “Relationship between the dynamic range of cochlear nerve fibers and their spontaneous activity,” Exp. Brain Res. **40**, 115-118.

Falmagne, J. C. (**1986**). “Psychophysical measurement and theory,” in *Handbook of perception and human performance (Vol. 1: Sensory processes and perception)*, edited by K. R. Boff, L. Kaufmann, and J. P. Thomas (Wiley, New York), *1-1 - 1-65.*

Fechner, G. T. (**1860**). *Elemente der Psychophysik* (Breitkopf und Härtel, Leipzig).

Fletcher, H., Munson, W. A. (**1933**). “Loudness, its definition, measurement and calculation,” J. Acoust. Soc. Am. **5**, 82–108.

Florentine, M. (**1983**). “Intensity discrimination as a function of level and frequency and its relation to high-frequency hearing,” J. Acoust. Soc. Am. **74**, 1375-1379.

Florentine, M., and Buus, S. (**1981**). “An excitation pattern model of intensity discrimination,” J. Acoust. Soc. Am. **70**, 1646-1654.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). “Temporal integration of loudness as a function of level,“ J. Acoust. Soc. Am. **99**, 1633–1644.

Florentine, M., Epstein, M., and Buus, S. (**2001**). “Loudness functions for short and long tones,” in *Fechner Day 2001: Proceedings of the 17th Annual Meeting of the International Society for Psychophysics*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (ISP, Leipzig).

Florentine, M., Fastl, H., Buus, S. (**1088**). “Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking,” J. Acoust. Soc. Am. **84**, 195-203.

Florentine, M., Reed, C. M., Rabinowitz, W. M., Braida, L. D., Durlach, N. I., and Buus, S. (**1993**). “Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss,” J. Acoust. Soc. Am. **94**, 2575-2586.

Fraisse, P. (**1982**). “Rhythm and tempo,” in *The psychology of music*, edited by D. Deutsch (Academic Press, New York), 149-180.

Galambos, R., Bauer, J., Picton, T., Squires, K, and Squires, N. (**1972**). “Loudness enhancement following contralateral stimulation,” J. Acoust. Soc. Am. **52**, 1127-1130.

Gibson, D. J., Young, E. D., and Costalupes, J. A. (**1985**). „Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei,“ J. Neurophysiol. **53**, 940–958.

Goodman, L. A. (**1960**). “On the exact variance of products,” Journal of the American Statistical Association **55**, 708-713.

Green, D. M. (**1993**). “Auditory intensity discrimination,” in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), 13-55.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Harris, D. M., and Dallos, P. (**1979**). “Forward-masking of auditory nerve fiber responses,” J. Neurophysiol. **42**, 1083-1107.

Harris, J. D. (**1963**). “Loudness discrimination,“ J. Speech Hear. Disord. Suppl. **11**, 1-63.

Hartmann, W. M. (**1995**). “The physical description of signals,” in *Hearing*, edited by B. C. J. Moore (Academic Press, San Diego), 3-40.

Hellman, R. P. (**1991**). “Loudness scaling by magnitude scaling: Implications for intensity coding,” in *Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens*, edited by G. A. Gescheider and S. J. Bolanowski (Erlbaum, Hillsdale NJ), 215-228.

Hellman, R. P., and Meiselman, C. H. (**1990**). “Loudness relations for individual and groups in normal and impaired hearing,“ J. Acoust. Soc. Am. **88**, 2596–2606.

Hellman, R. P., Scharf, B., Teghtsoonian, M., and Teghtsoonian, R. (**1987**). “On the relation between the growth of loudness and the discrimination of intensity for pure tones,” J. Acoust. Soc. Am. **82**, 448-453.

Hellman, R. P., and Zwislocki, J. J. (**1963**). “Monaural loudness function at 1000 cps and interaural summation,” J. Acoust. Soc. Am. **35**, 856-865.

Hellman, R. P., and Zwislocki, J. J. (**1964**). “Loudness function of a 1000 cps tone in the presence of a masking noise,” J. Acoust. Soc. Am. **36**, 1618-1627.

Hellman, W., and Hellman, R. (**1990**). “Intensity discrimination as the driving force for loudness. Application to pure tones in quiet,” J. Acoust. Soc. Am. **87**, 1255-1271.

Hellström, Å. (**1977**). “Time errors are perceptual. An experimental investigation of duration and a quantitative successive-comparison model,” Psychological Research **39**, 345-388.

Hellström, Å. (**1978**). “Factors producing and factors not producing time errors: An experiment with loudness comparisons,” Percept. Psychophys. **23**, 433-444.

Hellström, Å. (**1979**). “Time errors and differential sensation weighting,” J. Exp. Psychol. Hum. Percept. Perform. **5**, 460-477.

Hellström, Å. (**1985**). “The time-order error and its relatives: Mirrors of cognitive processes in comparing,” Psychological Bulletin **97**, 35-61.

Hellström, Å. (**2000**). “Sensation weighting in comparison and discrimination of heaviness,” J. Exp. Psychol. Hum. Percept. Perform. **26**, 6-17.

Helmholtz, H. v. (**1863**). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Vieweg, Braunschweig).

Hirsh, I. J., Monahan, C. B., Grant, K. W., and Singh, P. G. (**1990**). "Studies in auditory timing. 1. Simple patterns," Percept. Psychophys. **47**, 215-226.

Houtsma, A. J. M., Durlach, N. I., and Braida, L. D. (**1980**). "Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching," J. Acoust. Soc. Am. **68**, 807-813.

Huotilainen, M., Tiitinen, H., Sinkkonen, J., Ilmoniemi, R. J., Pekkonen, E., Parkkonen, L., and Näätänen, R. (**1995**). "MEG reveals intensity-specific neurons in the human auditory cortex," in *Proceedings of the 1st International Conference on Brain Mapping* (Paris).

Irino, T., and Patterson, T. (**2001**). "A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. 109, 2008-2022.

Irwin, R. J., and Zwislocki, J. J. (**1971**). "Loudness effects in pairs of tone bursts," Percept. Psychophys. **10**, 189-192.

Javel, E., and Viemeister, N.F. (**2000**). "Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and applications to intensity discrimination," J. Acoust. Soc. Am. **107**, 908-921.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85-88.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). „Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.

Jesteadt, W., and Bilger, R. C (**1974**). "Intensity and frequency discrimination in one- and two-interval paradigms," J. Acoust. Soc. Am. **55**, 1266-1276.

Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 169-177.

Johnson, J. H., Turner, C. W., Zwislocki, J. J., and Margolis, R. H. (**1993**). "Just noticeable differences for intensity and their relation to loudness," J. Acoust. Soc. Am. **93**, 983-991.

Kallman, H. J., and Massaro, D. W. (**1979**). "Similarity Effects in Backward Recognition Masking," J. Exp. Psychol. Hum. Percept. Perform. **5**, 110-128.

Kiang, N. Y.-S., Wantanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). "Discharge patterns of single fibers in the cat's auditory nerve," Res. Monogr. No. 35, MIT, Cambridge, MA.

Kinchla, R. A., and Smyzer, F. (**1967**). "A diffusion model of perceptual memory," Percept. Psychophys. **2**, 219–229.

Klein, S. (**2001**). "Measuring, estimating, and understanding the psychometric function: A commentary," Percept. Psychophys. **63**, 1421-1455.

Kramer, S. J., and Teas, D. C. (**1982**). "Forward masking of auditory nerve (N1) and brainstem (wave V) responses in humans, " J. Acoust. Soc. Am. **72**, 795-803.

Kvålseth, T. O. (**1985**). "Cautionary Note about R^2," The American Statistician **39**, 279-285.

Lachs, G., Al-Shaikh, R., Bi, Q., Saia, R. A., and Teich, M. C. (**1984**). "A neural-counting model based on physiological characteristics of the peripheral auditory system. V. Application to loudness estimation and intensity discrimination," IEEE Trans. Systems, Man Cybernet. **SMC-14**, 819–836.

Lachs, G., and Teich, M. C. (**1981**). "A neural-counting model incorporating refractoriness and spread of excitation. II. Application to loudness estimation," J. Acoust. Soc. Am. **69**, 774-782.

Laming, D. (**1986**). *Sensory Analysis* (Academic Press, London).

Levitt, H. (**1971**). "Transformed up-down procedures in psychoacoustics,", J. Acoust. Soc. Am. **49**, 467-477.

Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–455.

Liegeois-Chauvel, C., Laguitton, V., Badier, J. M., Schwartz, D., and Chauvel, P. (**1995**). "Cortical mechanisms of auditive perception in man: contribution of cerebral potentials and evoked magnetic fields by auditive stimulations," Rev. Neurol. **151**, 495-504.

Lim, J. S., Rabinowitz, W. M., Braida, L. D., and Durlach, N. I. (**1977**). "Intensity perception. VIII. Loudness comparisons between different types of stimuli,“ J. Acoust. Soc. Am. **62**, 1256–1267.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (**1996**). *SAS System for Mixed Models* (SAS Inst. Inc., Cary, NC).

Loftus, G. R., and Ruthruff, E. (**1994**). "A theory of visual information acquisition and visual memory with special application to intensity-duration tradeoffs," J. Exp. Psychol. Hum. Percept. Perform. **20***,* 33-50.

Longuet-Higgins, H. C., and Lee, C. S. (**1982**). "The perception of musical rhythms," Perception **11**, 115-156.

Loveless, N., and Hari, R. (**1993**). "Auditory evoked fields covary with perceptual grouping," Biol. Psychol. **35**: 1–15.

Loveless, N., Hari, R., Hamalainen, M., and Tiihonen, J. (**1989**). "Evoked responses of human auditory cortex may be enhanced by preceding stimuli,“ Electroencephalogr. Clin. Neurophysiol. **74**, 217-227.

Loveless, N., Levänen, S., Jousmäki, V., Sams, M., and Hari, R. (**1996**). "Temporal integration in auditory sensory memory: neuromagnetic evidence," Electroenceph. Clin. Neurophysiol. **100**, 220–228.

Luce, R. D., and Galanter, E. (**1963**). "Discrimination,“ in *Handbook of Mathematical Psychology, Vol. 1*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York), 191-243.

Luce, R. D., and Green, D. M. (**1974**). "Neural coding and psychophysical discrimination data," J. Acoust. Soc. Am. **56**, 1554-1564.

Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506–3511.

Marks, L. E. (**1994**). "Recalibrating the auditory system: The perception of loudness," J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.

Massaro, D. W. (**1970**). "Preperceptual auditory images," J. Exp. Psychol. **85**, 411-417.

Massaro, D. W. (**1974**). "Perceptual units in speech recognition," J. Exp. Psychol. **102***,* 199-208.

Massaro, D. W. (**1975**). *Experimental Psychology and Information Processing* (Rand McNally, Chicago).

Massaro, D. W., and Loftus, G. R. (**1996**). "Sensory and perceptual storage. Data and theory," in *Memory (Handbook of perception and cognition, 2nd edition)*, edited by E. L. Bjork, and R. Bjork (Academic Press, San Diego), 67-99.

Maxwell, S. E., and Delaney, H. D. (**2004**). *Designing Experiments and Analyzing Data. A Model Comparison Perspective, 2nd Edition* (Lawrence Erlbaum Associates, Mahwah).

McEvoy, L., Levänen, S., and Loveless, N. (**1997**). "Temporal characteristics of auditory sensory memory: neuromagnetic evidence," Psychophysiology **34**, 308-316.

McGill, W. J., and Goldberg, J. P. (**1968**a). "A study of the near miss to Weber's law and pure-tone intensity discrimination," Percept. Psychophys. **4**, 105-109.

McGill, W. J., and Goldberg, J. P. (**1968**b). "Pure-tone intensity discrimination and energy detection," J. Acoust. Soc. Am. **44**, 576-581.

McKenna, T. M., Weinberger, N. M., and Diamond, D. M. (**1989**). "Responses of single auditory cortical neurons to tone sequences," Brain Res. **481**, 142–153.

Michels, W.C., and Helson, H. (**1954**). "A quantitative theory of time-order effects," American Journal of Psychology **67**, 327-334.

Miskiewicz, A., Buus, S., and Florentine, M. (**1994**). "Auditory facilitation: Procedural or sensory effect?," J. Acoust. Soc. Am. **96**, 1429-1434.

Moore, B. C. J. (**1995**). "Frequency analysis and masking," in *Hearing*, edited by B. C. J. Moore (Academic Press, San Diego), 161-205.

Moore, B. C. J., and Glasberg, B. R. (**1982**). "Contralateral and ipsilateral cueing in forward masking," J. Acoust. Soc. Am. **71**, 942-945.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness, and partial loudness," J. Aud. Eng. Soc. **45**, 224-240.

Moore, B. C. J., and Raab, D. H. (**1974**). "Pure-tone intensity discrimination: some experiments relating to the "near-miss" to Weber's law," J. Acoust. Soc. Am. **55**, 1049-1054.

Mori, S., and Ward, L. M. (**1992**). "Intensity and frequency resolution: Masking of absolute identification and fixed and roving discrimination," J. Acoust. Soc. Am. **91**, 246-255.

Murnane, O. D., Prieve, B. A., and Relkin, E. M. (**1998**). "Recovery of the human compound action potential following prior stimulation," Hear. Res. **124**, 182-189.

Needham, J. G. (**1935**). "The effect of the time interval upon the time error at different intensity levels," J. Exp. Psy. **18**, 539-543.

Neff, D. L. (**1985**). "Stimulus parameters governing confusion effects in forward masking," J. Acoust. Soc. Am. **78**, 1966–1976.

Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). "Interactions between test-and inducer-tone durations in induced loudness reduction," J. Acoust. Soc. Am. **114**, 2846–2855.

Nishimura, T., Nakagawa, S., Sakaguchi, T., Hosoi, H., and Tonoike, M. (**2003**). "Effect of a forward masker on the N1m amplitude: varying the signal delay," Neuroreport **14**, 891-893.

Nizami, L., Reimer, J. F., and Jesteadt, W. (**2001**). "The intensity-difference limen for Gaussian-enveloped stimuli as a function of level: tones and broadband noise," J. Acoust. Soc. Am. **110**, 2505-2515.

Nizami, L., Reimer, J. F., and Jesteadt, W. (**2002**). "The mid-level hump at 2 kHz," J. Acoust. Soc. Am. **112**, 642-653.

Oxenham, A. J., and Moore, B. C. J. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921-1934.

Palmer, A. R., and Evans, E. F. (**1980**). "Cochlear fibre rate--intensity functions: no evidence for basilar membrane nonlinearities," Hear. Res **2,** 319-326.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuirer, G., and Elbert, T. (**1995**). "Specific Tonotopic Organizations of Different Areas of the Human Auditory Cortex Revealed by Simultaneous Magnetic and Electric Recordings," Electroencephalogr. Clin. Neurophysiol. **94**, 26-40.

Pantev, C., Hoke, M., Lutkenhöner, B., Lahmertz, K. (**1991**). "Neuromagnetic evidence of an ampliotopic organization of the human auditory cortex," Acta Otolaryngol. **Suppl. 491.** 106-115.

Parncutt, R. (**1994**). "A perceptual model of pulse salience and metrical accent in musical rhythms," Music Perception **11**, 409-464.

Plack, and Carlyon, R. P. (**1995**). "Loudness perception and intensity coding," in *Hearing*, edited by B. C. J. Moore (Academic Press, San Diego), 123-160.

Plack, C. J. (**1996**a). "Loudness enhancement and intensity discrimination under forward and backward masking," J. Acoust. Soc. Am. **100**, 1024-1030.

Plack, C. J. (**1996**b). "Temporal factors in referential intensity coding," J. Acoust. Soc. Am. **100**, 1031-1042.

Plack, C. J. (**1998**). "Beneficial effects of notched noise on intensity discrimination in the region of the 'severe departure'," J. Acoust. Soc. Am. **103**, 2530-2538.

Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (**1995**). "Intensity discrimination under forward and backward masking: Role of referential encoding," J. Acoust. Soc. Am. **97**, 1141-1149.

Plack, C. J., and Viemeister, N. F. (**1992**a). "Intensity discrimination under backward masking," J. Acoust. Soc. Am. **92**, 3097-3101.

Plack, C. J., and Viemeister, N. F. (**1992**b). "The effects of notched noise on intensity discrimination under forward masking," J. Acoust. Soc. Am. **92**, 1902-1910.

Plank, T., and Ellermeier, W. (**2003**). "Discrimination of temporal loudness profiles," in *Fechner Day 2003*, edited by B. Berglund and E. Borg (International Society for Psychophysics: Stockholm), 241-244.

Port, E. (**1963**). "Über die Lautstärke einzelner kurzer Schallimpulse," Acustica **13**, 212–223.

Povel, D.-J., and Okkerman, H. (**1981**). "Accents in equitone sequences," Percept. Psychophys. **30**, 565-572.

Preusser, D., Garner, W. R., and Gottwald, R. L. (**1970**). "The effect of starting pattern on descriptions of perceived temporal patterns," Psychonomic Science **21**, 219-220.

Raab, D. H., and Taub, H. B. (**1969**). “Click-intensity discrimination with and without a background masking noise,” J. Acoust. Soc. Am. **46**, 965-968.

Rankovic, C. M., Viemeister, N. F., Fantini, D. A., Cheesman, M. F., and Uchiyama, C. L. (**1988**). “The relation between loudness and intensity difference limens for tones in quiet and noise backgrounds,“ J. Acoust. Soc. Am. **84**, 150-155.

Relkin, E. M., and Doucet, J. R. (**1991**). “Recovery from prior stimulation. I. Relationship to spontaneous firing rates of primary auditory neurons,“ Hearing Res. **55**, 215–222.

Relkin, E. M., and Doucet, J. R. (**1997**). “Is loudness simply proportional to the auditory nerve spike count?,” J. Acoust. Soc. Am. **101**, 2735-2740.

Relkin, E. M., Doucet, J. R., and Sterns, A. (**1995**). “Recovery of the compound action potential following prior stimulation: evidence for a slow component that reflects recovery of low spontaneous-rate auditory neurons,” Hear. Res. **83**, 183-189.

Richards, V. M., and Zhu, S. (**1994**). “Relative estimation of combination weights, decision criteria, and internal noise based on correlation coefficients,” J. Acoust. Soc. Am. **95**, 423-434.

Riesz, R. (**1933**). “The relationship between loudness and the minimum perceptible increment in intensity,” J. Acoust. Soc. Am. **4**, 211-216.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). „Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input–output functions, tuning curves, and response phases,“ J. Acoust. Soc. Am. **80**, 1364–1374.

Romani, G. L., Williamson, S. J., and Kaufman, L. (**1982**). “Tonotopic organization of the human auditory cortex,” Science **216**, 1339-1340.

Rubin, H. (**1960**). “Auditory facilitation following stimulation at low intensities,” J. Acoust. Soc. Am. **32**, 670-681.

Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (**1992**). „Basilar membrane responses to two-tone and broadband stimuli,“ Philos. Trans. R. Soc. London B **336**, 307–315.

Sachs, M. B., and Abbas, P. J. (**1974**). “Rate versus level functions for auditory-nerve fibers in cats: Tone-burst stimuli,” J. Acoust. Soc. Am. **56**, 1835–1847.

Salvi, R. J., Saunders, S. S., Ahroon, W. A., Shivapuja, B. G., and Arehole, S.(**1986**). “Psychophysical and physiological aspects of auditory temporal processing in listeners with noise-induced sensorineural hearing loss,” in *Basic and Applied Aspects of Noise-Induced Hearing Loss*, edited by R. J. Salvi, D. Henderson, R. P. Hamernik, and V. Colletti (Plenum Press, New York), 179-184.

Sams, M., Hari, R., Rif, J., and Knutila, J. (**1993**). “The human auditory sensory memory trace persists about 10 s: neuromagnetic evidence,” J. Cogn. Neurosci. **5**, 363–370.

Scharf, B. (**1978**). “Loudness,” in *Hearing (Handbook of Perception, Vol. 4)*, edited by E. C. Carterette, and M. P. Friedman (Academic Press, New York), 187-242.

Scharf, B. (**2001**). “Sequential effects in loudness,” in *Fechner Day 2001: Proceedings of the 17th Annual Meeting of the International Society for Psychophysics*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (ISP, Leipzig).

Scharf, B., and Buus, S. (**1986**). “Audition I. Stimulus, physiology, thresholds,” in *Handbook of perception and human performance (Vol. 1: Sensory processes and perception)*, edited by K. R. Boff, L. Kaufmann, and J. P. Thomas (Wiley, New York), 14-1 - 14-71.

Scharf, B., Buus, S., and Nieder, B. (**2002**). “Loudness enhancement: induced loudness reduction in disguise? (L),” J. Acoust. Soc. Am. **112**, 807–810.

Scharf, B., and Houtsma, A. J. M. (**1986**). “Audition II. Loudness, pitch, localization, aural distortion, pathology,” in *Handbook of perception and human performance (Vol. 1: Sensory processes and perception)*, edited by K. R. Boff, L. Kaufmann, and J. P. Thomas (Wiley, New York), 15-1 - 15-60.

Schlauch, R. S. (**1994**). “Intensity resolution and loudness in high-pass noise,” J. Acoust. Soc. Am. **95,** 2171- 2179.

Schlauch, R. S., Clement, B. R., Ries, D. T., and DiGiovanni, J. J. (**1999**). “Masker laterality and cueing in forward-masked intensity discrimination,” J. Acoust. Soc. Am. **105**, 822-828.

Schlauch, R. S., DiGiovanni, J. J., and Ries, D. T. (**1998**). “Basilar membrane nonlinearity and loudness,” J. Acoust. Soc. Am. **103**, 2010-2020.

Schlauch, R. S., Harvey, S., and Lanthier, N. (**1995**). “Intensity resolution and loudness in broadband noise,” J. Acoust. Soc. Am. **98**, 1895-1902.

Schlauch, R. S., Lanthier, N., and Neve, J. (**1997**). “Forward-masked intensity discrimination: duration effects and spectral effects,” J. Acoust. Soc. Am. **102**, 461-467.

Schlauch, R. S., and Wier, C. C. (**1987**). “A method for relating loudness-matching and intensity-discrimination data,” J. Speech Hear. Res. **30**, 13-20.

Schulze, H. H. (**1989**). “The perception of temporal deviations in isochronic patterns,“ Percept. Psychophys. **45**, 291-296.

Smith, R. L. (**1977**). “Short-term adaptation in single auditory nerve fibers: some poststimulatory effects,” J. Neurophysiol. **40**, 1098-1112.

Smith, R. L. (**1979**). “Adaptation, saturation, and physiological masking in single auditory-nerve fibers,” J. Acoust. Soc. Am. **65**, 166-178.

Stellmack, M. A., and Viemeister, N. F. (**2000**). “Observer weighting of monaural level information in a pair of tone pulses,“ J. Acoust. Soc. Am. **107**, 3382-3393.

Stevens, S. S. (**1955**). “The measurement of loudness,” J. Acoust. Soc. Am. **27**, 815-829.

Stevens, S. S. (**1956**). “The direct estimation of sensory magnitudes - loudness,” Am. J. Psychol. **69**, 153-181.

Suga, N. (**1977**). “Amplitude spectrum representation in the Doppler-shifted-CF processing area of the auditory cortex of the mustache bat,” Science **196**, 64-67.

Suga, N., O’Neill, W. E., and Manabe, T. (**1978**). “Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the mustache bat,” Science **200**, 778–781.

Teich, M. C., and Khanna, S. M. (**1985**). “Pulse-number distribution for the neural spike train in the cat’s auditory nerve,” J. Acoust. Soc. Am. **77**, 1110-1128.

Teich, M. C., and Lachs, G. (**1979**). “A neural-counting model incorporating refractoriness and spread of excitation. I. Application to intensity discrimination,” J. Acoust. Soc. Am. **76**, 1738-1749.

ten Hoopen, G., Boelaarts, L., Gruisen, A., Apon, I., Donders, K., Mul, N., and Akerboom, S. (**1994**). "The detection of anisochrony in monaural and interaural sound sequences," Percept. Psychophys., **56**, 110-120.

Thurstone, L. L. (**1927**). "A law of comparative judgment," Psychol. Rev. **34**, 273-286.

Turner, C. W., Horwitz, C. W., and Souza, P. E. (**1994**). "Forward-masked intensity discrimination measured using forced-choice and adjustment procedures." J. Acoust. Soc. Am. **96**, 2121-2126.

van Noorden, L. P. A. S. (**1977**). "Minimum differences of level and frequency for perceptual fission of tone sequences," J. Acoust. Soc. Am. **61**, 1041-1045.

Viemeister N. F. (**1972**). "Intensity discrimination of pulsed sinusoids: The effects of filtered noise," J. Acoust. Soc. Am. **51**, 1265-1269.

Viemeister, N. F. (**1983**). "Auditory intensity discrimination at high frequencies in the presence of noise," Science **211**, 1206-1208.

Viemeister, N. F. (**1988**). "Psychophysical aspects of auditory intensity coding," in *Auditory Function*, edited by G. Edelman, W. Gall, and W. Cowan (Wiley, New York), 213-241.

Viemeister, N. F., and Bacon, S. P. (**1988**). "Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones," J. Acoust. Soc. Am. **84,** 172-178.

Vorberg, D., and Wing, A. (**1996**). "Modeling variability and dependence in timing," in *Handbook of Perception and Action. Vol. 2: Motor Skills*, edited by H. Heuer and S.W. Keele (Academic Press, London), 181-262.

Vos, P. G. (**1977**). "Temporal duration factors in the perception of auditory rhythmic patterns," *Scientific Aesthetics*, **1**, 183-199.

Wichmann, F. A., and Hill, N. J. (**2001**). „The psychometric function. I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293-1313.

Widin, G. P., Viemeister N. F., and Bacon, S. P. (**1986**). "Effects of forward and simultaneous masking on intensity discrimination," J. Acoust. Soc. Am. **80**, 108-111.

Winslow, R. L., and Sachs, M. B. (**1988**). "Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle," Hear Res. **35**, 165-89.

Woodworth, R. S., and Schlosberg, H. (**1954**). *Experimental psychology* (Holt, Rinehart, and Winston, New York).

Yates, G. K. (**1990**). "Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions," Hearing Res. **50**, 145–162.

Yost, W. A (**2000**). *Fundamentals of Hearing. An Introduction* (Academic Press, San Diego).

Young, E. D., and Barta, P. E. (**1986**). "Rate responses of auditory nerve fibers to tones in noise near masked threshold," J. Acoust. Soc. Am. **79**, 426–442.

Zeng, F.-G. (**1994**). "Loudness growth in forward masking: Relation to intensity discrimination," J. Acoust. Soc. Am. **96,** 2127-2132.

Zeng, F.-G. (**1998**). "Interactions of forward and simultaneous masking in intensity discrimination," J. Acoust. Soc. Am. **103**, 2021-2030.

Zeng, F.-G., and Shannon, R. V. (**1995**). “Possible origins of the nonmonotonic intensity discrimination function in forward masking,” Hear. Res. **82**, 216-224.

Zeng, F.-G., and Turner, C. W. (**1992**). “Intensity discrimination in forward masking,” J. Acoust. Soc. Am. **92**, 782-787.

Zeng, F.-G., Turner, C. W., and Relkin, E. M. (**1991**). “Recovery from prior stimulation II: Effects upon intensity discrimination,” Hear. Res. **55**, 223-230.

Zwicker, E. (**1958**). „Über psychologische und methodische Grundlagen der Lautheit,“ Acustica **8**, 237-258.

Zwicker, E. (**1963**). „Über die Lautheit von ungedrosselten und gedrosselten Schallen,“ Acustica **13**, 194-211.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models, 2nd edition* (Springer, Berlin).

Zwislocki, J. (**1965**). “Analysis of some auditory characteristics,” in *Handbook of mathematical psychology*, *Vol. 3*, edited by R. D. Luce, R. B. Bush, and E. Galanter (Wiley, New York), 3-97.

Zwislocki, J. J. (**1969**). “Temporal summation of loudness: An analysis,” J. Acoust. Soc. Am. **46**, 431-441.

Zwislocki, J. J., and Jordan, H. N. (**1986**). “On the relations of intensity jnd’s to loudness and neural noise,” J. Acoust. Soc. Am. **79**, 772-780.

Zwislocki, J. J., Ketkar, M., Cannon, M. W., and Nodar, R. H. (**1974**), “Loudness enhancement and summation in pairs of short sound bursts,” Percept. Psychophys. **16**, 91-95.

Zwislocki, J. J., and Sokolich, W. G. (**1974**). “On loudness enhancement of a tone burst by a preceding tone burst,” Percept. Psychophys. **16**, 87-90.

Appendix

I Derivation of the function representing loudness similarity

The function $f(N_T, N_M)$ presented in Eq. 7-7 has been deducted from a 'memory confusion' model. The underlying idea is that the amount of mergence is monotonically related to the probability of confusing the representations of masker and target loudness (Kallman and Massaro, 1979).

It is assumed that at presentation of masker and target in interval 1, the listener makes an observation of masker loudness and target loudness, respectively. The loudness values N_M and N_T are stored in a noisy memory system and the listener knows from presentation order if mask or target sounded louder. Now, what happens if the listener tries to retrieve target loudness from the store, and he or she has to accomplish this by using solely loudness information? Due to memory noise, the loudness values retrieved from the store are assumed to be instantiations of two normally distributed random variables X_M and X_T, with means N_M and N_T, respectively, and a common standard deviation σ_X (Thurstone's Case *V*, cf. Luce and Galanter, 1963, pp. 215*f*).

If the listener knows that $N_T > N_M$, the probability of correctly identifying the target is

$$p(\text{Correct Identification} \mid N_T > N_M) = p(X_T > X_M) = p(X_T - X_M > 0), \qquad \text{(A-1)}$$

where X_M is the value of masker loudness retrieved from the memory system, and X_T is the corresponding value of target loudness.

Using Thurstone's equation, we find

$$p(X_T - X_M > 0) = \int_{x=0}^{\infty} F_{N_T - N_M, \sigma_{Sim}}(x)\,dx\,, \qquad \text{(A-2)}$$

where N_T is the expected value of target loudness (Eq. 7-3) and N_M is the expected value of masker loudness (Eq. 7-2). As $X_T - X_M$ is a difference between two normally-distributed random variables, it is normally distributed with mean $(N_T - N_M)$ and standard deviation $\sigma_{Sim} = \sqrt{2\,\sigma^2_X - 2\,r(X_M, X_T)\,\sigma^2_X}$, where $r(X_M, X_T)$ is the correlation coefficient. In Eq. A-2, $F_{N_T - N_M, \sigma_{Sim}}$ denotes the normal probability density function with these parameters. For the sake of simplicity, σ_{Sim} is assumed to be constant (independent of N_T and N_M).

For $N_T < N_M$, the probability of a correct identification is $p(\text{correct identification} \mid N_T < N_M)$

$$p(X_T < X_M) = p(X_M - X_T > 0) = \int_{x=-\infty}^{0} F_{N_T - N_M, \sigma_{Sim}}(x)\,dx\,. \qquad \text{(A-3)}$$

From Eqs. A-2 and A-3 it follows that

$$p(\text{Correct Identification}) = \begin{cases} \int_{x=0}^{\infty} F_{N_T - N_M, \sigma_{Sim}}(x)dx & \text{if } N_T > N_M \\ \int_{x=-\infty}^{0} F_{N_T - N_M, \sigma_{Sim}}(x)dx & \text{if } N_T < N_M \end{cases}, \qquad (\text{A-4})$$

which can be written as

$$p(\text{Correct Identification}) = \frac{1}{2}\left(1 + Erf\left(\frac{|N_M - N_T|}{\sqrt{2}\sigma_{Sim}}\right)\right), \qquad (\text{A-5})$$

where $\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_{z=0}^{x} e^{-z^2} dz$.

As the weight p_M assigned to masker loudness is assumed to be monotonically related to the probability of falsely retrieving masker instead of target information from the store, less weight will be assigned to masker loudness if masker and target loudness differ strongly. If they are identical, the maximum amount of mergence will be effective. The effect of other differences between masker and target (e.g., duration, frequency) are included simply by scaling the loudness-related confusion probability predicted by Eq. A-5 so that the maximum probability of assessing masker loudness is just p_{Max},

$$p_M = p_{Max} \cdot 2 \cdot [1 - p(\text{correct identification})]. \qquad (\text{A-6})$$

If N_M equals N_T, p(correct identification) is at its minimum value (0.5) and masker weight p_M is equal to p_{Max}. If identification on the basis of loudness is perfect [p(correct identification) = 1], p_M is zero. Using Eqs. A-5 and A-6, we arrive at

$$p_M = p_{Max}\left(1 - \text{Erf}\left(\frac{|N_M - N_T|}{\sqrt{2}\sigma_{Sim}}\right)\right). \qquad (\text{A-7})$$

It seems reasonable to require that mean masker and target loudness differing by, e.g., 10% result in the same value of p_M independent of target loudness. In the present form, however, if N_T is large, p_M decreases more rapidly with the relative distance between N_T and N_M. To obtain the desired behavior, Eq. A-7 is modified by simply using logarithmically transformed loudness values,

$$p_M = p_{Max}\left(1 - \text{Erf}\left(\frac{|\log_{10} N_M - \log_{10} N_T|}{\sqrt{2}\sigma_{Sim}}\right)\right). \qquad (\text{A-8})$$

For stimuli differing only in presentation level, the similarity parameter σ_{Sim} is modeled to be constant, i.e., to be independent of loudness. As $\log_{10} N_M - \log_{10} N_T$ is equal to $\log_{10}(N_M/N_T)$, masker weight now depends on the ratio rather than on the difference between masker and target loudness if p_{Max} is also constant (Fig. 39).

II Glossary

II.1 Acoustical sound and sound pressure

Sound is propagated in the form of waves. If the medium is air, the motion of air molecules results in local changes in pressure. Sound pressure is measured as the root-mean-square (RMS) value of the deviations from static pressure (Scharf and Buus, 1986, p. 14-4). It is normally denoted as p or P and measured in units of Pascal. In a microphone, the displacement of the diaphragm is proportional to the pressure of the sound field. The displacement generates a proportional electric voltage. Consequently, the microphone output voltage V is proportional to instantaneous pressure.

II.2 Intensity

A sound wave carries energy. Its power (the rate of energy, measured in Watts, 1 $W = 1\ J/s$) is proportional to the square of sound pressure. Its intensity is the power flow per unit area of wavefront,

$$I = p^2/(\rho\, c), \qquad \text{(A-9)}$$

where p is RMS sound pressure, ρ is the density of the medium (approximately 1.21 kg/m^3 for air at 20°C) and c is the speed of sound in the medium (approx. 344 m/s for air at 20° C). If SI units are used, the unit of intensity is W/m^2.

II.3 Intensity level, sound pressure level

Due to the large dynamic range of our auditory system, intensity is almost exclusively expressed on a logarithmic scale relative to reference intensity I_0,

$$L = 10\ \log_{10}(I/I_0)\ \text{dB}, \qquad \text{(A-10)}$$

where I denotes intensity in W/m^2. The measure L is termed intensity level, its unit is decibels (dB). Most frequently $I_0 = 10^{-12}\ W/m^2$ is used, corresponding roughly to the threshold of human hearing at frequencies around 4 kHz. In this case, L is denoted as "sound pressure level" with unit "dB SPL".

From Eq. A-10 it follows that

$$L = 10\ \log_{10}\frac{p^2/(\rho c)}{p_0^{\,2}/(\rho c)}\ \text{dB} = 20\ \log_{10}(p/p_0)\ \text{dB}, \qquad \text{(A-11)}$$

where p_0 is reference sound pressure. The value corresponding to $I_0 = 10^{-12}\ W/m^2$ is $p_0 = 20$ μPa, if the medium is air at room temperature. For comparison, atmospheric pressure is approximately 10^5 Pa.

Note that the prefactor to be applied in a decibel calculation is 10 for energylike or fluxlike quantities, which are real and never negative (Hartmann, 1995). Examples are acoustical or electrical energy, intensity, or power. For dynamical quantities, which may be also negative or complex, the appropriate prefactor is 20. This factor applies to pressure or voltage.

II.4 Sensation level (SL)

Sensation level is sound pressure level expressed relative to a listener's detection threshold for the sound,

$$L_{SL} = L_{SPL} - L_{TH} \text{ dB SL}, \qquad \text{(A-12)}$$

where L_{TH} is sound pressure level at threshold.

II.5 Hearing level (HL)

Hearing level is also called hearing threshold level (HTL). It is detection threshold measured in dB SPL minus the average threshold of normal listeners for a specific sound (as published, e.g., in the international norm ISO 389-1). The unit is "dB HL".

Note that hearing level depends on frequency and on the sound source, but not on the listener.

II.6 Coherent and non-coherent summation

If two equal-frequency tones are added in phase, their sound pressures add linearly. If their sound pressure levels are identical ($L_1 = L_2$), their pressures are also identical ($p_1 = p_2$). Total level L_{Total} is 6 dB greater than the individual levels in this situation,

$$L_{Total} = 20 \log_{10}\left(\frac{p_1 + p_2}{p_0}\right) \text{ dB SPL} = 20 \log_{10}\left(\frac{2p_1}{p_0}\right) \text{ dB SPL} = L_1 + 6 \text{ dB}. \qquad \text{(A-13)}$$

If two tones of identical frequency are added with a 90° phase shift between them ("quadrature phase"), their intensities rather than their pressures add linearly. The same applies to the addition of sounds differing in frequency. In these conditions, if two tones of the same level are added, total level is only 3 dB above the individual levels,

$$L_{Total} = 10 \log_{10}\left(\frac{I_1 + I_2}{I_0}\right) \text{ dB SPL} = 20 \log_{10}\left(\frac{2I_1}{I_0}\right) \text{ dB SPL} = L_1 + 3 \text{ dB}. \qquad \text{(A-14)}$$

II.7 Loudness

Loudness is, by definition, the perceived intensity of a sound (Scharf, 1978). A more empirical definition was given by Plack and Carlyon (1995): "Loudness is the sensation that corresponds most closely to the physical measure of sound intensity" (p. 124). As loudness is a subjective quality, it is measured by different methods and on different scales, all subject to discussion.

Note that loudness matching does not provide a direct measure of sensation magnitude. The most straightforward methods to derive a subjective scale are **magnitude estimation** or **magnitude production.** Stevens (1955, 1956) defined loudness of 40 dB 1-kHz tone as 1 sone and loudness at threshold as 0 sone.

In his magnitude estimation experiments, he found loudness in of tones well above threshold to depend on intensity I as

$$S = k \cdot I^{0.3} = k' \cdot p^{0.6}, \qquad (\text{A-15})$$

where k and k' are scale constants and p is pressure.

From this *loudness function*, i.e., the function relating physical intensity and perceived loudness, it follows that the sound pressure level of 1-kHz tone in plane field has to be increased by 10 dB in order to enlarge the sensation of loudness by a factor of two. The specified loudness function holds only for levels above 40 dB SPL. Below 40 dB SPL, level differences needed for doubling/halving loudness are smaller. This can be accounted for by a corrected loudness function (Scharf, 1978)

$$S = k' (P - P_{Th})^{2\theta}, \qquad (\text{A-16})$$

where P_{Th} is pressure at threshold. This function is steeper near threshold. The estimates of the exponent θ vary with sound parameters, experimental procedure, and also from experiment to experiment. For example, Hellman and Zwislocki (1963) found a value of 0.27 for 1-kHz pure tones at stimulus levels above 40 dB SPL.

A model for loudness of pure tones accounting also for the effects of simultaneous masking was proposed by Zwislocki (1965). This model is discussed in Chapter 7.1.1 above.

For complex stimuli (sounds containing more than one frequency component), Zwicker (1958, 1963, cf. Zwicker and Fastl, 1999, p. 223ff) formulated a model based on excitation patterns. An advanced model using the ideas of Zwicker was proposed by Moore, Glasberg, and Baer (1997).