Loudness changes induced by a proximal sound: Loudness enhancement, loudness recalibration, or both?

Daniel Oberfeld[a)]
Department of Psychology, Johannes Gutenberg—Universität Mainz, 55099 Mainz, Germany

(Received 21 September 2006; revised 30 January 2007; accepted 30 January 2007)

The effect of a forward masker on the loudness of a target tone in close temporal proximity was investigated. Loudness matches between a target and a comparison tone at the same frequency were obtained for a wide range of target and masker levels. Contrary to the hypothesis by Scharf, Buus, and Nieder [J. Acoust. Soc. Am. **112**, 807–810 (2002)], these matches could not be explained by an effect of the masker on the comparison loudness, which was measured by loudness matches between the comparison and a fourth tone separated in frequency from the comparison and the masker. The data thus demonstrate that a forward masker has an effect on the loudness of a proximal target. The results are compatible with the suggestion by Arieh and Marks [J. Acoust. Soc. Am. **114**, 1550–1556 (2003)] that the masker triggers two processes. The data indicate that the effect of the slower-decaying process resulting in a reduction in the loudness of a following tone saturates at masker-target level differences of 10–20 dB. The faster-decaying process causing loudness enhancement or loudness decrement has the strongest effect at a masker-target level difference of approximately 30 dB. A model explaining this mid-difference hump is proposed. *© 2007 Acoustical Society of America.* [DOI: 10.1121/1.2710433]

PACS number(s): 43.66.Cb, 43.66.Mk, 43.66.Ba [AJO]   Pages: 2137–2148

I. INTRODUCTION

Does a sound presented in close temporal proximity to a target sound have an effect on the loudness of the target, and how can this effect be understood? Effects of a *masker*[1] temporally separated from a *target* by less than 1 s have received considerable interest in the field of auditory intensity processing, although not all phenomena are currently well understood (for recent reviews see Plack and Carlyon, 1995; Scharf, 2001; Oberfeld, 2005). Concerning effects on loudness, several studies conducted in the 1970s reported that a forward masker higher in level than a target following it by less than about 500 ms caused an increase in target loudness (*loudness enhancement*; e.g., Galambos *et al.*, 1972; Zwislocki and Sokolich, 1974), while the loudness of the target was reduced if the masker level was lower than the target level (*loudness decrement*; e.g., Zwislocki and Sokolich, 1974; Elmasian *et al.*, 1980). In the experiments, listeners adjusted the level of a *comparison* presented approximately 1000 ms after the target, until the comparison loudness matched the loudness of the target. Brief tone or noise bursts were used. The comparison was presented at the same frequency as the masker and the target. The upper row labeled "Three-tone task" in Fig. 1 displays a typical temporal structure. Forward maskers higher in level than the target resulted in the comparison level being adjusted to higher levels than in quiet, which was taken as evidence for loudness enhancement of the target. To summarize the experimental results, loudness enhancement increases with the level difference between masker and target, at least for level differences up to 30 dB, amounting to as much as 20 dB (Elmasian and Galambos, 1975; Elmasian *et al.*, 1980). On the other hand, for the masker level fixed to 90 dB sound pressure (SPL), the maximum amount of loudness enhancement was observed at *intermediate* target levels (40–60 dB SPL) in experiments by Zeng (1994) and Plack (1996). At low target levels, the effect of the masker on the loudness level of the target was small, however, representing an analog to the midlevel hump in intensity discrimination (e.g., Zeng *et al.*, 1991; Carlyon and Beveridge, 1993). Oberfeld (2003) showed that forward masking has a significant effect on the loudness of a low-level target, but that this effect is maximal at intermediate masker-target level differences in the range between 20 and 40 dB, resulting in a mid-*difference* hump. Zwislocki and Sokolich (1974) reported that loudness enhancement vanished if the masker-target inter-stimulus interval (ISI) was longer than 400 ms. Loudness enhancement was also observed if the masker followed the target (backward masking; Elmasian *et al.*, 1980; Plack, 1996).

Recently, Scharf *et al.* (2002) raised the question of whether the loudness matches obtained in the three-tone matching task really reflect a change in the loudness of the target. They used a comparison tone much higher in frequency than the masker and the target and found no evidence for loudness enhancement. Scharf *et al.* (2002) interpreted this finding as to showing that in previous studies the masker did not enhance the loudness of the target presented proximally to the masker, but rather caused loudness recalibration (Marks, 1994), or "induced loudness reduction (ILR)," as Scharf *et al.* termed it, in the comparison. The loudness recalibration experiments showed that a moderately strong tone (e.g., 80 dB SPL) reduces the loudness of a weaker tone following with an ISI of more than about 200 ms (for a review see Wagner and Scharf, 2006). The effect was observed only for targets similar in frequency to the masker (Marks,

[a)]Electronic mail: oberfeld@uni-mainz.de

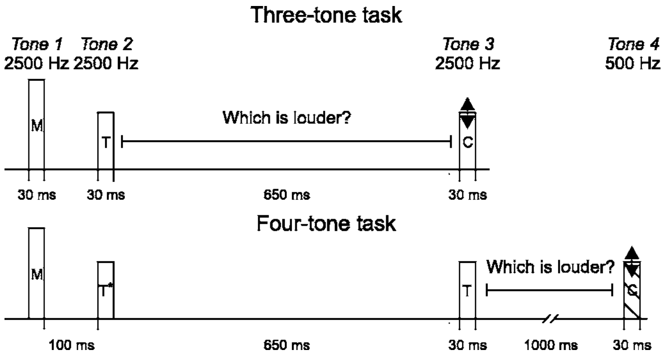

FIG. 1. Trial configurations used in the experiments. In both tasks, the listener decided whether the target tone *T* or the comparison tone *C* had been louder. The level of the target was fixed. The level of the comparison was adjusted by an adaptive procedure. The level of the masker *M* was varied between blocks. In the three-tone task (upper row), the frequency of all tones was 2500 Hz. Listeners produced a loudness match between Tone 2 and Tone 3. In the four-tone task (lower row), listeners produced a loudness match between Tone 3 (2500 Hz) and Tone 4 (500 Hz). The target (Tone 3) was preceded by exactly the same stimuli as the comparison in the three-tone task. Tone 2 (T^*) was identical in level to the target. For the baseline matches, the masker or the masker and the tone T^* were omitted. All stimuli were 30 ms tone bursts.

1994). Scharf *et al.* pointed out that a reduction in comparison loudness would result in the comparison being adjusted to higher levels than the target, that is, in "loudness enhancement" as defined in the three-tone matching task. To explain why the target is not also subject to loudness reduction, however, it was necessary to assume 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 remained unclear, data compatible with the hypothesis were reported by Arieh and Marks (2003a), who presented the masker (80 dB SPL) and the target (60 dB SPL) at 2500 Hz, and the comparison at 500 Hz. The masker had virtually no effect at masker-target ISIs smaller than 200 ms, but caused a reduction in the loudness level of more than 11 dB at ISIs longer than 500 ms. The average loudness match produced for the masker and the comparison presented at the same frequency corresponded to the average difference between the loudness reduction induced in a target presented 575 or 75 ms after the masker.

These recent data make it necessary to reinterpret the results obtained in experiments using the same frequency for all tones. It remains to be shown, however, whether the entire set of loudness enhancement data can be explained by ILR. For instance, Nieder *et al.* (2003) reported that presenting the masker at either 80 dB SPL or 95 dB SPL produced a difference of only about 2 dB in the loudness reduction induced in 70 dB SPL targets. In contrast, Elmasian and Galambos (1975) found the average loudness matches for a 70 dB SPL target obtained with an 80 dB SPL and a 100 dB SPL masker to differ by about 11 dB. Another important issue is the observation of loudness *decrement* (e.g., Elmasian and Galambos, 1975). The loudness of the target being matched by a comparison *lower* in level than the target could be explained by loudness recalibration only if a masker less intense than the target produced an *increase* in comparison loudness, contrary to results by Mapes-Riordan and Yost (1999).

In the present study, the hypothesis that the masker-induced change in loudness level measured with the masker and the comparison sharing the same frequency is due to a reduction in comparison loudness was tested directly in two experiments. Loudness matches in the traditional three-tone loudness matching procedure, depicted in the upper row of Fig. 1, were obtained for masker-target level combinations found to produce the most pronounced changes in loudness level. For the same listeners, the effect of the masker on the loudness of the comparison (Tone 3 in the lower row of Fig. 1) was measured by obtaining loudness matches between the latter tone and a fourth tone presented at a much lower frequency. This made it possible to compare the change in loudness level in the three-tone task to the change in comparison loudness, for each masker-target level combination. An important detail was that in the four-tone task depicted in the in the lower row of Fig. 1, the same two stimuli as in the three-tone task preceded Tone 3, namely the masker and what had been the target. In the experiments by Arieh and Marks (2003a), only one tone (the masker) preceded the test tone. To estimate the reduction in comparison loudness effective in the three-tone task, however, it is important to use exactly the same temporal configuration.

II. EXPERIMENT 1: CAN "LOUDNESS ENHANCEMENT" BE EXPLAINED BY LOUDNESS RECALIBRATION?

Experiment 1 tested the hypothesis that the loudness matches obtained with the target and the comparison presented at the same frequency can be explained by a reduction in comparison loudness (Scharf *et al.*, 2002). Target level (L_T) was 60 dB SPL, a level at which both considerable "loudness enhancement" (Zeng, 1994; Plack, 1996; Oberfeld, 2003) and ILR (Mapes-Riordan and Yost, 1999; Arieh and Marks, 2003a; Nieder *et al.*, 2003) have been reported. The maximum masker level was 90 dB SPL, corresponding to the condition producing the largest amount of loudness enhancement (Zeng, 1994; Plack, 1996; Oberfeld, 2003). A 70 dB SPL masker was presented because ILR can be expected to be most pronounced for a 10–20 dB level difference between masker and target (Mapes-Riordan and Yost, 1999). To answer the question of whether loudness *decrement* is due to a change in comparison loudness rather than in target loudness, two maskers lower in level than the target (L_M=40 and 50 dB SPL) were included.

A. Method

Nine students at the Johannes Gutenberg—Universität Mainz participated in the experiment voluntarily (eight female, one male, age 20–27 years). They either received partial course credit or were paid for their participation. All reported normal hearing. For the ear tested, detection thresholds were better than 13 dB hearing level (HL) at all octave frequencies between 0.5 and 4 kHz.

All stimuli were pure tones with a steady-state duration of 20 ms, gated on and off with 5 ms $\cos^2$ ramps. On each trial, listeners decided whether the target (the penultimate tone in the trial) or the comparison (the final tone) had been louder. The level of the target was fixed at 60 dB SPL. The level of the comparison was adjusted by an adaptive procedure. The level of the masker was varied between blocks.

The stimuli were generated digitally, played back via one channel of an RME ADI/S digital-to-analog converter (sampling rate 44.1 kHz, 24 bit resolution), attenuated (TDT PA5), buffered (TDT HB7), and presented to the right ear via Sennheiser HDA 200 headphones calibrated according to IEC 318 (1970). The attenuator setting remained constant within a trial. The experiment was conducted in a single-walled soundproof chamber.

The experiment comprised two tasks. In the *three-tone task* depicted in the upper row of Fig. 1, the frequency of all tones was 2500 Hz. Listeners produced a loudness match between Tone 2 (target) and Tone 3 (comparison). The silent interval between masker offset and target onset was 100 ms. The interval between target offset and comparison onset was 650 ms. In the *four-tone task* (Fig. 1, lower row), listeners produced a loudness match between Tone 3 (target: 2500 Hz) and Tone 4 (comparison: 500 Hz). Tone 3 was preceded by exactly the same two tones as the comparison in the three-tone task. Tone 2 (T^*) was identical in level to the target (Tone 3). The silent interval between Tone 3 and Tone 4 was 1000 ms. For the baseline matches, the masker M or the masker and tone T^* were omitted. Listeners were instructed to ignore the masker and tone T^*. No feedback was provided. The inter-onset interval between the target in a given trial and the target in the following trial was fixed at 5.7 s in both tasks.

A two-interval, two-alternative forced-choice (2I, 2AFC) interleaved-staircase procedure (Jesteadt, 1980) was used. Each experimental block comprised two randomly interleaved tracks. The upper track converged on the comparison level corresponding to the 70.7% *comparison louder* point on the psychometric function. If the listener indicated on two consecutive trials that he or she had perceived the comparison as being louder than the target, the level of the comparison was reduced. After each response indicating that the target had been perceived as being louder than the comparison, the level of the comparison was increased. In the lower track, a 1-down, 2-up rule was used to track the 29.3% *comparison louder* point on the psychometric function. In the three-tone task, the upper and the lower track started with a comparison level 15 dB above and below target level, respectively. In the four-tone task, the two tracks started with the level of Tone 4 15 dB above and below, respectively, the baseline match obtained at the beginning of a given session. The step size was 5 dB until the fourth reversal. The track continued with a step size of 2 dB until ten reversals had occurred or 60 trials had been presented. If in one of the tracks ten reversals had already occurred before the other track had also reached ten reversals, trials from the former track were still presented with an a priori probability of 0.25.

Listeners received only one task in each session. Sessions with the two tasks alternated. In each block, only one masker-target level combination was presented. A session started with a baseline match. In the following blocks, the masker (three-tone task) or the masker and tone T^* (four-tone task) were included. Masker level increased from block to block, starting with $L_M=40$ dB. Therefore, the largest masker level was always presented at the end of a session. The experiment started with a practice session.

For each block, the arithmetic mean of the level differences between comparison and target (L_C-L_T) at all but the first four reversals was computed separately for the upper and for the lower track, with the restriction that for each track an even number of reversals entered the computation (e.g., if 11 reversals had occurred in one of the tracks, reversal 11 was excluded). The arithmetic mean of these two values was taken as the loudness match, corresponding to the comparison level at the point of subjective equality (PSE) minus target level. A run was discarded if the standard deviation of L_C-L_T at the counting reversals was greater than 6 dB in either the upper or the lower track. Three runs were obtained in different sessions for each task and masker level, resulting in a total of six experimental sessions. Time permitting, additional runs were presented in a seventh session if the standard deviation of the loudness matches exceeded 5 dB.

Because the upper and the lower track converge on the 70.7% and the 29.3% *comparison louder* point on the psychometric function, respectively, half the difference between L_C-L_T in the upper track and L_C-L_T in the lower track was taken as a measure of loudness variability, denoted as *jnd* $=(x_{0.707}-x_{0.293})/2$ (Schlauch and Wier, 1987; Zeng, 1994; Plack, 1996).

B. Results and discussion

For the three-tone task, the individual results are displayed in Fig. 2 as the level difference between comparison (Tone 3) and target (Tone 2) at the point of subjective equality (PSE), relative to the baseline match. Positive values indicate that the masker either enhanced the loudness of Tone 2, or reduced the loudness of Tone 3, or both. For the four-tone task, the results are displayed as the level difference between target (Tone 3) and comparison (Tone 4) at the PSE, relative to the baseline match (without masker and tone T^*). Positive values indicate that the masker reduced the loudness of Tone 3. For each block in which a masker was presented, the arithmetic mean of the baseline matches was subtracted from the loudness match; the latter was defined as the arithmetic mean of average L_C-L_T in the upper and in the lower track. For the four-tone task, the resulting value was multiplied by -1.

What can be concluded about the three-tone task (squares in Fig. 2)? For masker levels lower than the target level, most loudness matches (relative to the baseline) were negative. If the masker level was higher than the target level, virtually all loudness matches were positive. Therefore, the traditional interpretation of the results would be that the masker produced loudness decrement or loudness enhancement, depending on its level relative to the target level. The size of the effects was compatible with previous data, with a maximum amount of "loudness decrement" of about 7 dB (Elmasian and Galambos, 1975), a maximum amount of

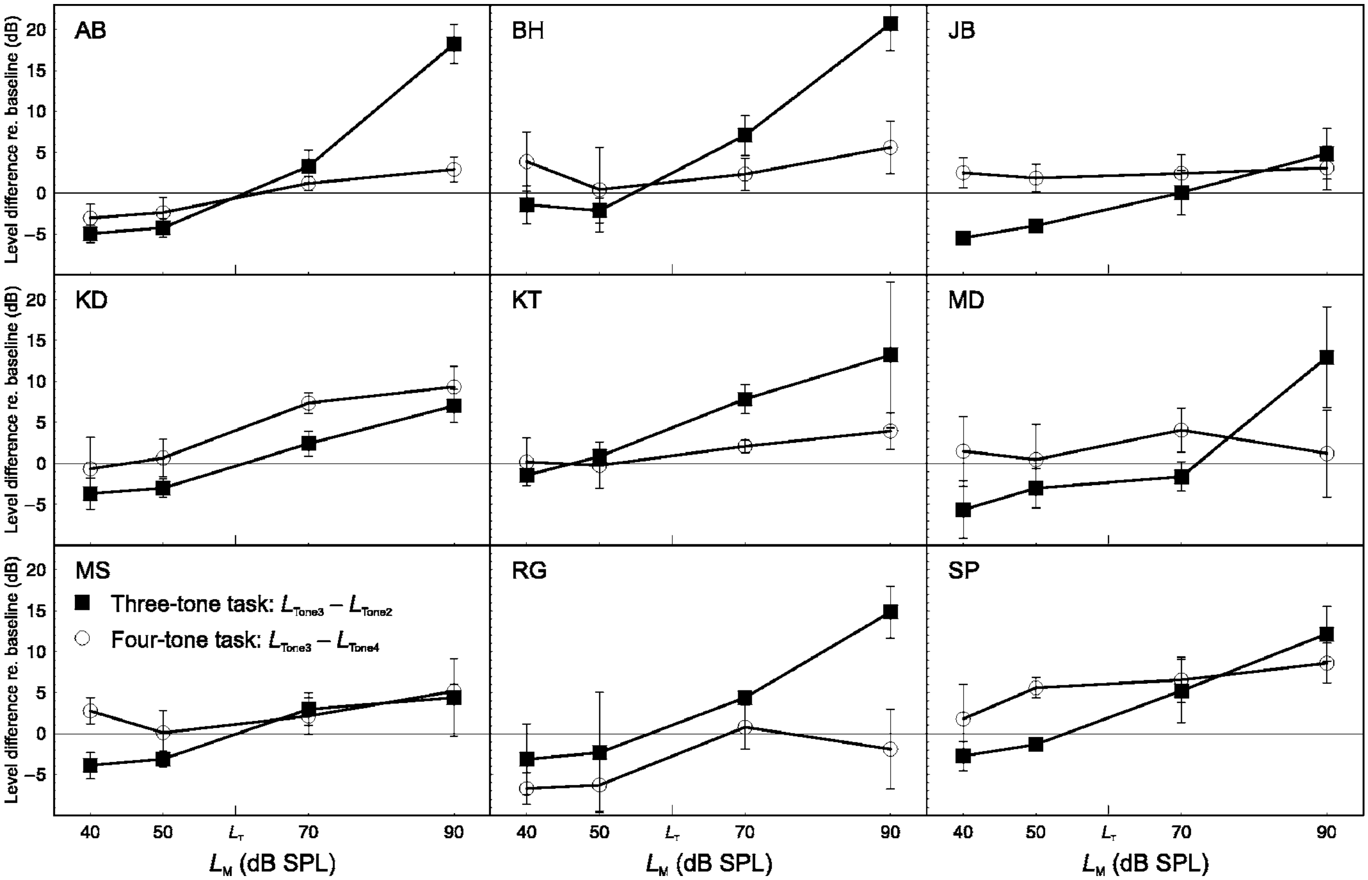

FIG. 2. Experiment 1. Individual masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task), as a function of masker level and task. Squares: level of the comparison (Tone 3) matching the loudness of the target (Tone 2) in the three-tone task, relative to the baseline match. Circles: level of Tone 3 (2500 Hz) matching the loudness of Tone 4 (500 Hz) in the four-tone task, relative to the baseline match. Target level was 60 dB SPL. Panels represent listeners. Error bars show plus and minus one standard deviation of the three or more measurements obtained for each data point.

loudness enhancement of about 20 dB (Elmasian and Galambos, 1975; Oberfeld, 2003), and a considerable amount of inter-individual variability (Plack, 1996).

Can the effects observed in the three-tone task be explained by a change in the loudness of the comparison? If so, the two lines in Fig. 2 representing matches from the two tasks should lie on top of each other. Yet, only a few of the data points are compatible with this prediction. First, for masker levels lower than the target level, the match in the four-tone task (circles in Fig. 2) was close to zero for most listeners (e.g., listener KD), compatible with data by Mapes-Riordan and Yost (1999). Therefore, the *negative* level difference between comparison and target required to produce the loudness match in the three-tone task cannot be explained by a change in comparison loudness in most cases.

For masker levels higher than the target level, the data do in fact indicate a reduction in the loudness of Tone 3 (the comparison in the three-tone task), except for listener RG at

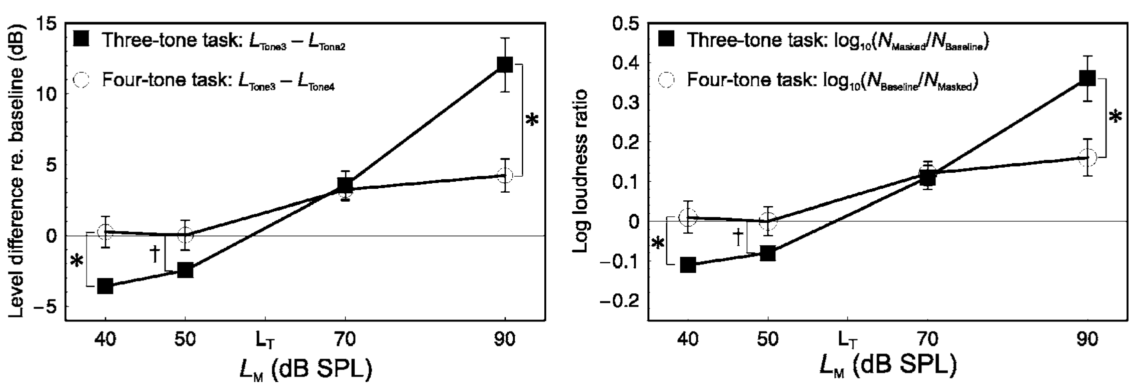

FIG. 3. Experiment 1. Left panel: Mean masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task), as a function of masker level and task. Same symbols as in Fig. 2. Right panel: Mean logarithm of the ratio between comparison loudness at the PSE in forward masking and comparison loudness at the PSE in quiet, as a function of masker level and task. Brackets indicate significant differences (paired-sample t tests; *: $p < 0.05$, †: $p < 0.1$; two tailed). Error bars show plus and minus one standard error of the mean (SEM) of the nine individual estimates.

the highest masker level. The maximum effect was approximately 10 dB. Mean results are displayed in the left panel of Fig. 3. Mean loudness reduction was 4.2 dB (SD=3.5 dB) at L_M=90 dB SPL. This value is smaller than the 11 dB of loudness reduction Arieh and Marks (2003a) found for an 80 dB SPL masker combined with a 60 dB SPL target (50 ms, 2500-Hz tone bursts), but comparable to the amount of ILR reported by Mapes-Riordan and Yost (1999) for the same level combination and 500 ms, 2500 Hz tones. It therefore remains unclear whether the tone T^* interpolated between masker and target in the four-tone task reduced the amount of ILR. With the 90 dB SPL masker, the reduction in Tone 3 loudness was considerably smaller than the loudness enhancement of Tone 2 observed in the three-tone task for six of the nine listeners. With the 70 dB SPL masker, the difference between the two measures was generally smaller.

A repeated-measures analysis of variance (ANOVA) with Huynh-Feldt correction for the degrees of freedom was conducted. The factors were masker level and task. There was a significant effect of masker level [$F(3,24)=72.8$, $p=0.001$, $\tilde{\varepsilon}=0.7$]. As a posthoc analysis, two separate ANOVAs with the factor masker level were conducted. The effect of masker level was significant for both the three-tone and the four-tone task [$F(3,24)=59.83$, $p=0.001$, $\tilde{\varepsilon}=0.46$ and $F(3,24)=11.10$, $p=0.001$, $\tilde{\varepsilon}=0.95$, respectively]. One-sample t tests were conducted for each data point. For the three-tone task, all matches differed significantly from 0 dB ($p<0.05$, two tailed). For the four-tone task, the matches were significantly different from 0 dB only if the masker level was higher than the target level, confirming the observation by Mapes-Riordan and Yost (1999) that a masker lower in level than the target produces no ILR. Loudness decrement thus reflects a reduction in the loudness of the target presented proximally to the masker.

In the two-factorial ANOVA, the effect of task was not significant [$F(1,8)=0.124$]. There was a significant Masker Level×Task interaction, however, demonstrating that the matches of Tone 3 versus Tone 2 (three-tone task) and Tone 3 versus Tone 4 (four-tone task) were not identical [$F(3,24)=20.0$, $p=0.001$, $\tilde{\varepsilon}=0.71$]. Posthoc pairwise comparisons indicated that the two matches obtained at each masker level differed significantly at a masker level of 40 dB SPL [$t(8)=-3.2$, $p=0.012$], and marginally significantly at a masker level of 50 dB SPL [$t(8)=-2.3$, $p=0.054$]. These results confirm that loudness decrement cannot be explained by a change in comparison loudness. At a masker level of 70 dB SPL, the difference between the two different matches was not significant [$t(8)=0.22$]. This observation is compatible with data by Arieh and Marks (2003a) obtained for an 80−dB SPL masker combined with a 60 dB SPL target. For the 90 dB SPL target, however, there was a significant difference between the loudness match obtained in the three-tone and in the four-tone task [$t(8)=3.2$, $p=0.013$]. Thus, the masker had an effect on target loudness at this masker level, contrary to the hypothesis by Scharf *et al.* (2002).

The upper panel of Fig. 4 shows the loudness matches in the three-tone task plotted against the loudness matches in the four-tone task. Scharf *et al.* (2002) predicted that the reduction in Tone 3 loudness measured in the four-tone task

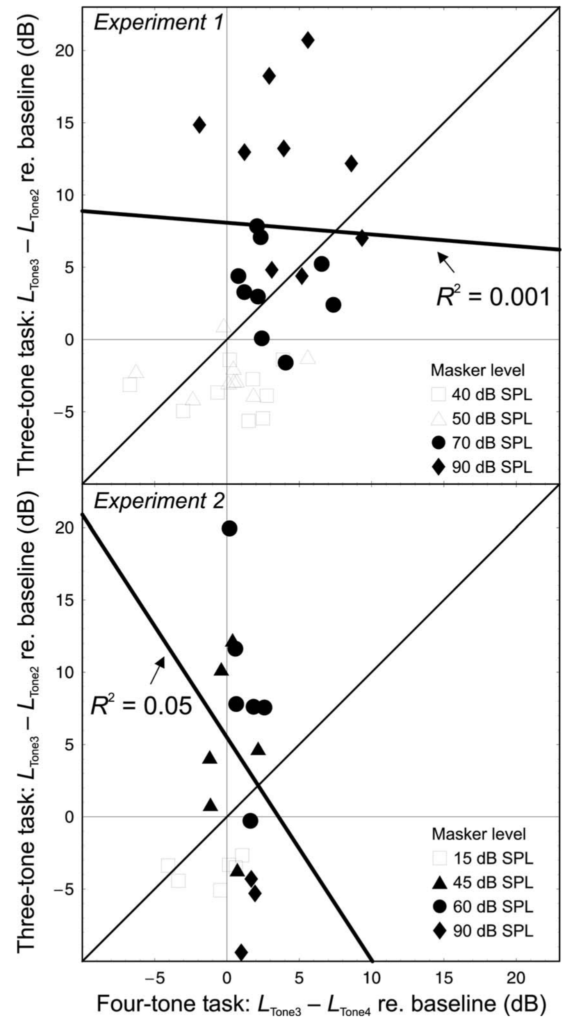

FIG. 4. Scatterplot of individual masker-induced changes in the loudness level of Tone 3 in the four-tone task (x axis), and of Tone 2 in the three-tone task (y axis). Upper panel: Experiment 1. Lower panel: Experiment 2. Each data point represents one listener and one masker level. The different masker levels are indicated by the different symbols. The thick line shows the best fitting linear regression line, computed only for the data points where masker level was higher than target level (filled symbols). All of these data points should lie on the diagonal if the loudness match in the three-tone task was determined completely by the loudness reduction of the comparison.

be identical to the change in loudness level measured in the three-tone task, that is, that all data points where the masker was higher in level than the target (filled symbols in Fig. 4) should lie on the diagonal. The thick line shows the best-fitting linear regression line. The Pearson product-moment correlation was not significant ($r=-0.04$, $N=18$), confirming the pronounced differences between the loudness matches in the two tasks. In an additional analysis, the data points where $L_M < L_T$ were also included. The correlation between the two measures was significant ($r=0.40$, $N=36$, $p=0.017$), but the proportion of the variance of the three-tone matches explained by the four-tone matches was still only $R^2=0.16$.

Taken together, the data show that the effects of a forward masker on the loudness matches in experiments presenting the masker and the comparison at the same frequency can only partly be attributed to an effect on comparison loudness. Thus, a forward masker induces loudness changes in a proximal target, but the loudness matches are contaminated by ILR of the comparison if the masker and the comparison share the same frequency.

There are three issues which might be a reason to qualify the above interpretation of the loudness matches. The first issue arises because the loudness matches in the three-tone and in the four-tone task were obtained for different comparison frequencies. The second issue is a potential effect of the sequence of comparison levels presented during the course of a session or during an experimental block. The third issue is related to the variation in the level of Tone 3 introduced by the adaptive procedure in the three-tone task.

Concerning the first issue, from the difference between the comparison level at the PSE in forward masking and the comparison level at the PSE in quiet observed in the three-tone task on the one hand and the four-tone task on the other hand, the conclusion was drawn that the effect of the masker on the loudness level of the target was larger than can be explained by a reduction in comparison loudness. However, the direct comparison of the level difference in dB at 2500 and 500 Hz is valid only if the slope of the loudness functions (*Lf*) for 500 and 2500 Hz tones is identical. A recent estimate of the slope of the *Lf* at various frequencies has been provided by Suzuki and Takeshima (2004), who estimated the exponent of the power function relating sound intensity and loudness at levels above 30 dB SPL to be 0.31 at 500 Hz and 0.29 at 2500 Hz (see Fig. 7 in Suzuki and Takeshima, 2004). To test for a potential effect of this difference in the exponents on the conclusion presented above, the ratio between the comparison loudness at the PSE in forward masking and the comparison loudness at the PSE in quiet (baseline condition) was estimated for each individual block. The loudness function proposed by Zwislocki (1965) was used, because it provides a better account of loudness near threshold than the loudness function used by Suzuki and Takeshima (cf. Buus *et al.*, 1998). According to Zwislocki (1965)

$$N = a[(p^2 + 2.5p_t^2)^\alpha - (2.5p_t^2)^\alpha], \tag{1}$$

where N is loudness, p is sound pressure, p_t is sound pressure at the detection threshold, α is the frequency dependent slope of the *Lf*, and a is a scale constant. Individual detection thresholds for 500 and 2500 Hz tones (duration 30 ms, including 5 ms $\cos^2$ ramps) in quiet were used. As no detection thresholds were available for listener RG at 2500 Hz and for listener KT at both frequencies, the average threshold of the remaining listeners at the respective signal frequency was used in these three cases. The ratio between comparison loudness at the PSE in forward masking and comparison loudness at the PSE in the baseline condition was estimated using Eq. (1)

$$\frac{N_{\text{Masked}}}{N_{\text{Baseline}}} = \frac{(p_{\text{Masked}}^2 + 2.5p_t^2)^\alpha - (2.5p_t^2)^\alpha}{(p_{\text{Baseline}}^2 + 2.5p_t^2)^\alpha - (2.5p_t^2)^\alpha}, \tag{2}$$

where p_{Masked} and p_{Baseline} denote the sound pressure of the comparison matching the loudness of the target in the presence of the forward masker and in quiet, respectively. Because the ratio $N_{\text{Masked}}/N_{\text{Baseline}}$ ranges from unity to infinity for N_{Masked} greater than N_{Baseline}, but only from zero to unity for N_{Masked} smaller than N_{Baseline}, the logarithm of the ratio was used in the analyses. For the data obtained in the four-tone task, the ratio $N_{\text{Masked}}/N_{\text{Baseline}}$ was inverted so that positive values of the log loudness ratio indicate a masker-induced reduction in the loudness of Tone 3.

As the right panel of Fig. 3 shows, the mean log loudness ratios exhibited the same pattern as the masker-induced changes in loudness level displayed in the left panel. A repeated-measures ANOVA with the factors masker level and task was conducted. There was a significant effect of masker level $[F(3,24)=66.9,\ p=0.001,\ \tilde{\varepsilon}=0.75]$. The Masker Level×Task interaction was also significant $[F(3,24)=13.9,\ p=0.001,\ \tilde{\varepsilon}=0.74]$, demonstrating that the masker-induced loudness changes were not identical in the two tasks. The effect of task was not significant $[F(1,8)=0.01]$. Post-hoc pairwise comparisons indicated that the log loudness ratios obtained in the two tasks differed significantly at a masker level of 40 dB SPL $[t(8)=-3.1,\ p=0.015]$, and at a masker level of 90 dB SPL $[t(8)=2.5,\ p=0.036]$. At a masker level of 50 dB SPL, the difference was marginally significant $[t(8)=-2.2,\ p=0.061]$. At a masker level of 70 dB SPL, the difference was not significant $[t(8)=-0.26]$. This is the same pattern of statistical results that had been observed for the changes in loudness level.

Concerning the second issue, in a paper published after Experiments 1 and 2 had been completed, Epstein and Gifford (2006) reported that obtaining the baseline match at the beginning of a session may result in an underestimation of ILR. Because the comparison level will be higher during the baseline match than in a subsequent block in which a condition producing ILR is presented, the comparisons in the second block might be subject to ILR caused by the comparisons presented in the first block, due to the slow recovery from ILR (Arieh *et al.*, 2005; Epstein and Gifford, 2006). Epstein and Gifford found the estimate of ILR to be about 3 dB smaller if the experimental condition (80 dB SPL masker, 70 dB SPL target) was run immediately after the baseline match rather than after a delay of 15 or 120 min. For the present experiment, there might thus have been a carry over effect from the first block (baseline) to the second block (40 dB SPL masker). Moreover, as the loudness match with the 40 dB SPL and the 50 dB SPL masker was rather similar to the baseline match for most listeners, a carry over effect from Block 2 to Block 3, and from Block 3 to Block 4 cannot be precluded either. On the other hand, as the minimum duration of Block 4 presenting the 70 dB SPL masker was 5 min, which is larger than the ILR recovery time of 130 s (Arieh *et al.*, 2005), it can be assumed that the problem was not relevant for the block presenting the 90 dB SPL masker. Even if it had been, the 1–4 dB underestimation of ILR reported by Epstein and Gifford (2006) cannot account for the 8 dB discrepancy between the loudness reduction of Tone 3 (measured in the four-tone task) and the change in loudness level observed in the three-tone task. The results by Epstein and Gifford also point to a potential problem associated with the interleaved-staircase procedure used in the current study as well as by Mapes-Riordan and Yost (1999). The comparisons in the upper track are on average higher in level

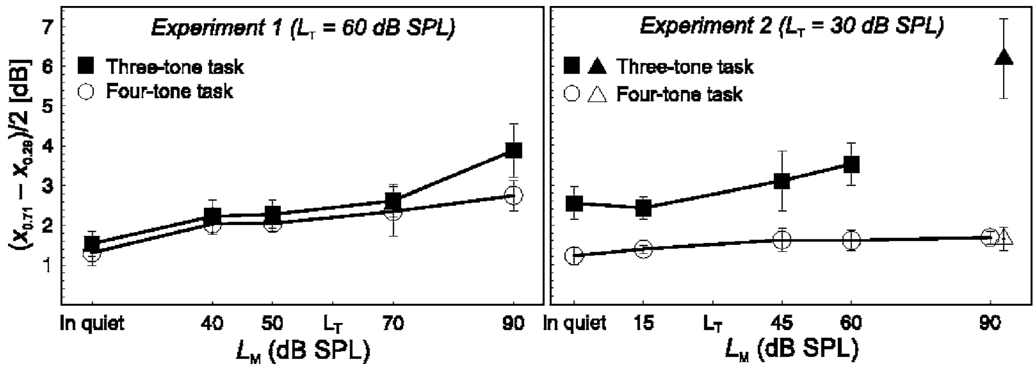

FIG. 5. Experiment 1. Mean loudness variability measured as half the difference between average reversal level in the upper and in the lower track $[(x_{0.71} - x_{0.29})/2]$, as a function of masker level and task. Squares: three-tone task. Circles: four-tone task. Left panel: Experiment 1. Right panel: Experiment 2. Only three listeners were tested with the 90 dB SPL masker in the three-tone task. The filled and open triangles show their average data in the three-tone task and the four-tone-task, respectively. Error bars show ±1 SEM.

than the comparisons in the lower track, so that the latter might be subject to ILR. The same effect would also apply to the baseline matches, however. Because ILR is defined as the loudness match in a block presenting a masker minus the baseline match, a potential loudness reduction of the tones in the lower track would cancel, provided that the level difference between upper and lower track does not differ between the forward masked conditions and the baseline condition. The average difference (with the *SD* in parentheses) between mean comparison level in the upper track and mean comparison level in the lower track was 8.4 dB (2.8 dB), 9.2 dB (2.8 dB), and 7.9 dB (2.7 dB) for blocks presenting the 70 dB SPL masker, the 90 dB SPL masker, and the baseline condition, respectively. Thus, the level difference in blocks presenting the 90 dB SPL masker was on average 1.4 dB larger than in the baseline condition, but it is unlikely that this small difference significantly affected the estimate of loudness reduction induced in Tone 3. Note also that because the loudness match in a given block is defined as the arithmetic mean between the average reversal levels in the two tracks, any influence of a change in loudness level of the lower track on the estimate will be attenuated by a factor of 2.

Turning to the third issue, the induced reduction in Tone 3 loudness was measured in the four-tone task with the level of Tone 3 fixed at target level. In the three-tone task, however, the level of Tone 3 (the comparison) was varied by the adaptive procedure, so that it could have levels above and below target level. If now the 90 dB SPL masker caused a disproportionately greater amount of ILR at lower levels of Tone 3, then the reduction in the loudness of Tone 3 could have been greater in the three-tone task than in the four-tone task, due to Tone 3 being presented at levels below the target level in the lower track of a three-tone block. This would result in the impression of additional loudness enhancement above that predicted by ILR on the four-tone task. To estimate the importance of this issue, mean comparison level in the lower track was computed for each block obtained in the three-tone task, for all trials following the fourth reversal (i.e., the part of the track that was used in the calculation of the loudness match $L_C - L_T$). At the 90 dB SPL masker level, mean comparison level in the counting part of the lower track was lower than target level in 11 of the total of 32 blocks only, with a minimum value of 7.1 dB below target level. In 18 blocks, on the other hand, mean comparison level in the lower track was at least 5 dB higher than target level. Thus, on average ILR of Tone 3 should have been *smaller* in the three-tone than in the four-tone task. With the 11 blocks in which mean comparison level in the counting part of the lower track was lower than target level excluded, the mean loudness match (relative to the baseline condition) obtained with the 90 dB SPL masker was 14.7 dB ($SD = 4.9$ dB) and 4.4 dB ($SD = 3.7$ dB) for the three-tone and the four-tone task, respectively. The difference between these two matches was significant [$t(7) = 4.05$, $p = 0.005$; note that only eight of the nine listeners contributed to this analysis], just as in the original analysis presented above. The variation in the level of Tone 3 in the three-tone task thus has no implications for the interpretation of the results.

The mean loudness variability is displayed in the left panel of Fig. 5. Carlyon and Beveridge (1993) suggested that in intensity discrimination experiments, the *jnd* elevation caused by a forward masker might be related to the masker-induced loudness change. In fact, loudness enhancement and intensity-difference limens or loudness variability were found to be (weakly) correlated (Zeng, 1994; Plack, 1996; Oberfeld, 2005). For the present data, the average increase in the loudness variability observed with the 90 dB SPL masker was only 2.3 dB in the three-tone task, while elevations of 4–15 dB in the intensity DL have been observed for 80–90 dB SPL maskers combined with a 60 dB SPL standard (e.g., Plack *et al.*, 1995; Oberfeld, 2005). The data speak against a one-on-one relation between the masker-induced loudness change and loudness variability: Paired-sample *t* tests indicated that for both tasks, the loudness variability was significantly larger in forward masking than in quiet at all masker levels except the lowest ($p < 0.05$, two tailed). Thus, there was an effect on the loudness variability in conditions producing loudness enhancement, ILR, and virtually no effect on the loudness (cf. Fig. 3, left panel). For the three-tone task, a repeated-measures ANOVA conducted for the forward-masked conditions showed that the increase in the loudness variability with masker level was significant

$[F(3,24)=4.4$, $p=0.038$, $\tilde{\varepsilon}=0.56]$. For the four-tone task, an ANOVA conducted for the data obtained under forward masking showed no significant effect of masker level $[F(3,24)=1.1]$.

III. EXPERIMENT 2: ENHANCEMENT VERSUS RECALIBRATION AT A LOW TARGET LEVEL

Marks (1996), Arieh and Marks (2003a), and Wagner and Scharf (2006) suggested that ILR occurs only if the masker is presented at a relatively high sound pressure level (i.e., above 60 dB SPL). It therefore seemed unlikely that the change in the loudness level of a 25 dB SPL target caused by 40 and 55 dB SPL maskers reported by Oberfeld (2003) was due to a reduction in comparison loudness. In Experiment 2, exactly the same design as in Experiment 1 was used, but a 30 dB SPL target was presented. Masker-target level differences ranged from −15 to +60 dB. The effect of the masker on the loudness level in the three-tone task was expected to be most pronounced at intermediate masker-target level differences (Oberfeld, 2003).

A. Method

The same stimuli, apparatus and procedure as in Experiment 1 were used, except for the lower target level and the different masker levels.

Seven students took part in the experiment voluntarily; only one of them (KD) had participated in Experiment 1. The listeners either received partial course credit or were paid for their participation. All reported normal hearing. For the ear tested, detection thresholds were better than 10 dB HL at all octave frequencies between 0.5 and 4 kHz. One of the listeners showed a systematic shift of the loudness matches in the three-tone task during the course of the experiment. With the 60 dB SPL masker, for instance, she adjusted the comparison to a level 33.7 dB above target level in the first session presenting the three-tone task. In the following sessions, the level difference between comparison and target required for the loudness match gradually shifted toward negative values. In the last session presenting the three-tone task, it was −13 dB, resulting in a range of more than 45 dB. A comparable pattern was observed with the 15 dB SPL and the 45 dB SPL masker in the three-tone task. The data of this listener were excluded from the analysis. The remaining listeners (five female, one male) ranged in age between 19 and 26 years.

Detection thresholds were obtained for 500 Hz and 2500 Hz tones with a duration of 30 ms including 5 ms $\cos^2$ ramps. A 2I, 2AFC, adaptive procedure was used (3–down, 1–up rule; Levitt, 1971). Three runs were obtained per condition. For the 500 Hz tones presented in quiet, the individual thresholds ranged from 11.9 to 21.6 dB SPL ($M=17.2$ dB SPL, $SD=4.0$ dB). For the 2500 Hz tones presented in quiet the individual thresholds ranged from 4.5 to 9.6 dB SPL ($M=7.3$ dB SPL, $SD=2.1$ dB). The 2500 Hz tones were also presented with 30 ms, 2500 Hz foward maskers and a masker-signal ISI of 100 ms. Mean thresholds (with *SD*s in parentheses) were 16.9 dB SPL (7.9 dB) and 19.1 dB SPL (8.3 dB) at a masker level of 60 dB SPL and 90 dB SPL, respectively. For listener KS, the average threshold measured with the 90 dB SPL masker was 31.6 dB SPL and thus above the level of the target in the loudness matching task. For listener TG, even the 60 dB SPL masker caused a large threshold elevation to 29.6 dB SPL, although she reported to clearly hear the target in three-tone loudness matching blocks presenting this masker level. No threshold was obtained for this listener in the condition presenting the 90 dB SPL masker. Listener MM reported that she was not able to distinguish the target from the 90 dB SPL masker in the three-tone task, but only perceived some sort of echo, even though the 30 dB SPL target was well above her forward masked detection threshold (19.4 dB SPL). These three listeners were not tested with the 90 dB SPL masker in the three-tone task.

B. Results

Individual data are shown in Fig. 6. In the three-tone task and at a masker level of 15 dB SPL, all listeners adjusted the level of the comparison to a lower level than in the baseline condition, indicating loudness decrement (squares in Fig. 6). For the 45 dB SPL and the 60 dB SPL masker, the level of the comparison matching the loudness of the target was higher than in the baseline condition for all listeners but SD, indicating loudness enhancement. Note that only three listeners were tested with the 90 dB SPL masker. The matches indicated a reduction in loudness rather than loudness enhancement in this condition. These results are in accordance with the expected mid-difference hump.

In the four-tone task (circles in Fig. 6), the change in loudness level effected by the masker was smaller than in the three-tone task. It was also smaller than for the 60 dB SPL target presented in Experiment 1. This observation is compatible with the assumption that ILR is most pronounced for targets presented at an intermediate level (Mapes-Riordan and Yost, 1999; Arieh and Marks, 2003a; Wagner and Scharf, 2006), although no study systematically measured ILR for a low-level target combined with different masker levels. The data are not compatible with the hypothesis by Scharf *et al.* (2002), according to which in Fig. 6, the two lines representing matches from the two tasks should lie on top of each other.

Mean data are shown in the left panel of Fig. 7. An ANOVA with the within-subjects factors masker level and task showed a significant Masker Level × Task interaction $[F(2,10)=16.6$, $p=0.002$, $\tilde{\varepsilon}=0.79]$. The data obtained with the 90−dB SPL masker were excluded from this analysis because only three of the six listeners had been tested in this condition. Pairwise comparisons indicated that the matches obtained in the two tasks differed significantly at a masker level of 15 dB SPL $[t(5)=-3.2$, $p=0.024]$, 60 dB SPL $[t(5)=-2.65$, $p=0.046]$, and 90 dB SPL $[t(2)=-6.0$, $p=0.027]$. For the 45 dB SPL masker, the difference was not significant $[t(5)=1.84]$. The main effect of task was not significant $[F(1,5)=2.7]$. There was a significant main effect of masker level $[F(2,10)=21.1$, $p=0.001$, $\tilde{\varepsilon}=0.81]$. As a posthoc analysis, two separate repeated measures ANOVAs with the factor masker level were run. The effect of masker level was significant for both the three-tone and the four-tone task

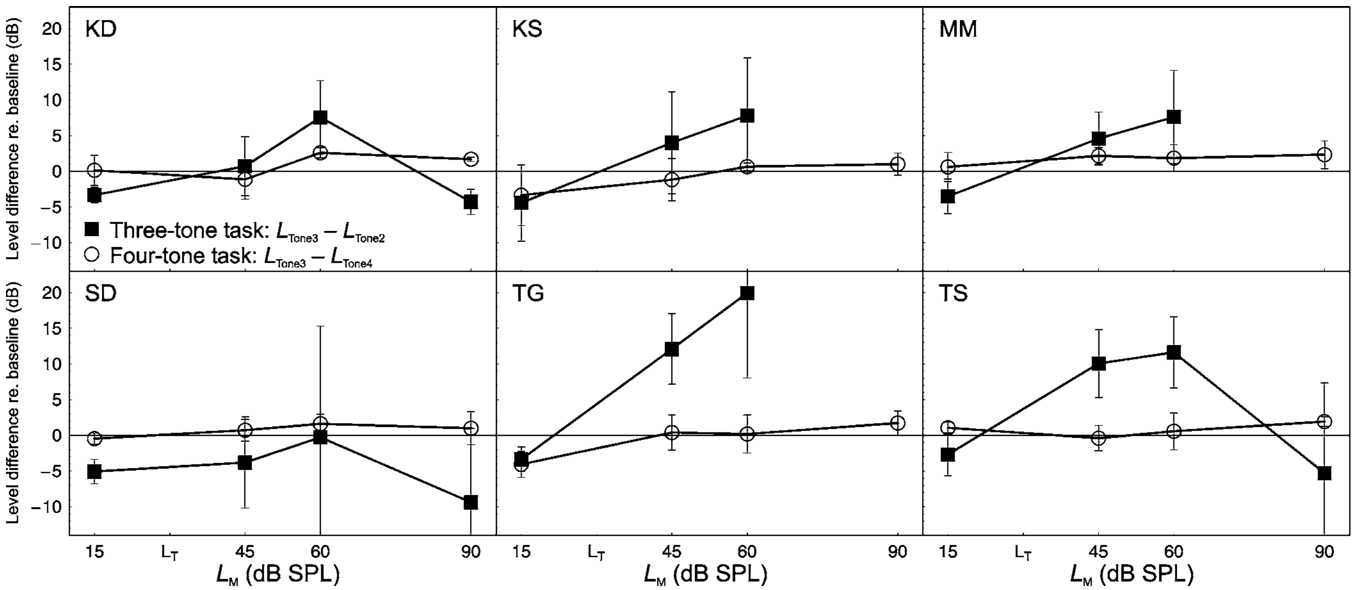

FIG. 6. Experiment 2. Individual masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task), as a function of masker level and task. Target level was 30 dB SPL. Same format as Fig. 2.

[$F(2,10)=21.6$, $p=0.001$, $\tilde{\varepsilon}=0.72$ and $F(3,15)=6.2$, $p=0.006$, $\tilde{\varepsilon}=1.0$, respectively]. One-sample t tests conducted for each data point showed that for the three-tone task, only the matches obtained with the 15 dB SPL and the 60 dB SPL masker differed significantly from 0 dB. For the four-tone task, the matches indicated significant loudness reduction only for the 60 dB SPL and the 90 dB SPL masker.

The lower panel of Fig. 4 shows the loudness matches in the three-tone task plotted against the loudness matches in the four-tone task. Each data point represents one masker level and one listener. For the data points where the masker was higher in level than the target (filled symbols), the correlation between the two measures was not significant ($r=-0.23$, $N=15$). The thick line shows the corresponding regression line. If the data points where $L_M < L_T$ were also included, the correlation between the two measures was again not significant ($r=0.14$, $N=21$).

The data were additionally analyzed in terms of the ratio between comparison loudness at the PSE in forward masking and comparison loudness at the PSE in quiet, using Eq. (2) and individual detection thresholds. As it can be seen in the right panel of Fig. 7, the mean log loudness ratios exhibited approximately the same pattern as the loudness matches displayed in the left panel of Fig. 7. In a repeated-measures ANOVA with the factors masker level and task, the Masker Level × Task interaction was significant [$F(2,10)=12.2$, $p=0.005$, $\tilde{\varepsilon}=0.77$], indicating that the masker-induced loudness changes were not identical in the two tasks. Note that the data obtained with the 90 dB SPL masker were again excluded from the analysis. Posthoc pairwise comparisons indicated that the log loudness ratios observed in the two tasks differed significantly at a masker level of 15 dB SPL [$t(5)=-2.91$, $p=0.033$], and 90 dB SPL [$t(2)=-5.96$, $p=0.027$]. For the 45 dB SPL masker, the difference was not significant [$t(5)=1.62$]. This pattern was also observed in the analysis of the masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task). Unlike in the latter analysis, however, the difference was not

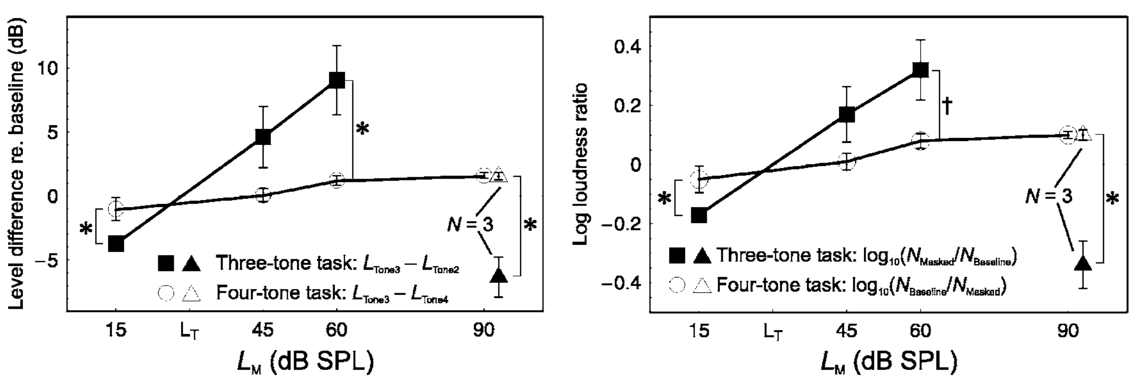

FIG. 7. Experiment 2. Left panel: Mean masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task), as a function of masker level and task. Right panel: Mean logarithm of the ratio between comparison loudness at the PSE in forward masking and comparison loudness at the PSE in quiet (baseline condition), as a function of masker level and task. Squares: three-tone task. Circles: four-tone task. Only three listeners were tested with the 90 dB SPL masker in the three-tone task. The filled and open triangles show their average data in the three-tone task and the four-tone-task, respectively. Brackets indicate significant differences (*: $p<0.05$, †: $p<0.1$; two tailed). Error bars show ±1 SEM of the individual estimates.

significant at the 60 dB SPL masker level [$t(5)=-2.01$, $p=0.10$]. For this masker level, the significant difference between the change in the loudness level of the target in the three-tone task and the change in the loudness level of Tone 3 in the four-tone task must thus be viewed with some caution.

As discussed above, the fact that the level of Tone 3 was fixed in the four-tone task but varied in the three-tone task presents a potential problem for the interpretation of the data. At the 60 dB SPL masker level, at which the masker-induced change in loudness level was significantly larger in the three-tone than in the four-tone task, the mean comparison level in the counting part of the lower track was lower than the target level in ten of the total of 24 blocks obtained in the three-tone task. With the former 10 blocks excluded, the mean loudness match obtained with the 60 dB SPL masker was 14.9 dB ($SD=5.9$ dB) and 1.2 dB ($SD=0.9$ dB) for the three-tone and the four-tone task, respectively. The difference between these two matches was significant [$t(5)=13.7$, $p=0.004$]. Thus, the difference between the masker-induced changes in loudness level observed in the two tasks cannot be attributed to the variation in comparison level in the three-tone task.

Taken together, the results from Experiment 2 (30 dB SPL target) confirm the conclusion for the 60 dB SPL target (Experiment 1) that the masker-induced changes in loudness level measured in a three-tone matching task with comparison frequency equal to masker frequency cannot be explained exclusively by a reduction in the loudness of the comparison. Instead, the data demonstrate that the masker had an effect on the loudness of the proximal target.

The mean loudness variability is displayed in the right panel of Fig. 5. An ANOVA with the within-subjects factors masker level and task was conducted. The data obtained with the 90 dB SPL masker were excluded from the analysis. The effect of masker level was only marginally significant [$F(3,15)=2.9$, $p=0.071$, $\tilde{\varepsilon}=1.0$]. The *jnd* obtained with the 90 dB SPL masker in the three-tone task for three listeners did not differ significantly from the *jnd* in quiet, either [$t(2)=2.85$]. The effect of the forward masker on the loudness variability was thus generally smaller than in Experiment 1, which is compatible with reports that an intense masker (e.g., 90 dB SPL) has a larger effect on the loudness variability for midlevel than for low-level targets (Zeng, 1994; Plack, 1996; Oberfeld, 2005). The loudness variability was significantly larger in the three-tone than in the four-tone task [$F(1,5)=11.3$, $p=0.020$]. The Masker Level $\times$ Task interaction was not significant [$F(3,15)=0.6$].

IV. GENERAL DISCUSSION

The present study tested the hypothesis by Scharf *et al.* (2002) that the change in loudness level measured in experiments presenting the masker and the comparison at the same frequency does not reflect an effect of the masker on the loudness of the proximal target, but rather a reduction in the loudness of the comparison. The results show that a forward masker does induce changes in target loudness, although Scharf *et al.* (2002) conjectured correctly that there are also effects on comparison loudness. Maskers lower in level than the target had no effect on comparison loudness, so that the observed loudness decrement must represent a change in target loudness. For a masker-target level difference of 30 dB, the increase in the loudness level of the target observed in the three-tone task was significantly larger than the loudness reduction of the comparison, demonstrating loudness enhancement of the target. The correlation between the change in the loudness level measured in the three-tone task and the reduction of the comparison loudness level was not significant.

A. Two-process model for the loudness changes caused by a proximal sound

The data collected in the current study are evidence supporting the hypothesis by Arieh and Marks (2003a) and Arieh *et al.* (2004) that the masker triggers two processes, loudness enhancement and loudness recalibration. They can also be used to refine the two-process model. Experiment 1 showed that for a 70 dB SPL masker combined with a 60 dB SPL target, the loudness match in the three-tone task can be explained by a loudness reduction of the comparison. This result is compatible with the observation by Arieh and Marks (2003a) that at a masker-target ISI of 75 ms, an 80 dB SPL masker had no effect on the loudness match between a 60 dB SPL target and a comparison presented at a much lower frequency. Arieh and Marks noted that this finding can be interpreted in two different ways: either both loudness enhancement and loudness recalibration of the target were absent, or both effects were present but equally strong, so that they cancelled. Put differently, Arieh and Marks (2003a) suggested that the masker might trigger two processes, both with a fast onset: loudness enhancement with a decay time of some 100 ms, and loudness recalibration with a decay time of several seconds. This would account for their finding of only very small loudness changes at masker-target ISIs smaller than 200 ms, without the need to assume a delayed onset of loudness recalibration, which would be at odds with the fast onset of inhibitory processes observed at various stages of the auditory system (see Arieh and Marks, 2003a, for a discussion). Moreover, Arieh *et al.* (2004) reported that an 80 dB SPL masker caused "residual loudness recalibration." In the first block of their experiment, the loudness of a 60 dB SPL target was unaffected when it followed the masker by 100 ms. However, a directly following experimental block in which the masker was omitted showed a reduction in target loudness relative to the baseline match. To explain the latter result under the assumption that the masker had caused only recalibration, but no enhancement, it would be necessary to assume that in the first block, the maskers triggered the inhibitory process, but that the target following the masker by 100 ms was somehow protected from its effect (Scharf *et al.*, 2002). Additionally, why should a 70 or an 80 dB SPL masker cause no loudness enhancement in a 60 dB SPL target, while Experiment 1 demonstrated that a 90 dB SPL masker does? The two-process hypothesis can resolve this puzzle, simply by assuming that for maskers 10–20 dB higher in level than a mid-level target, loudness recalibration and loudness enhancement cancel.

The results obtained in the present study demonstrate that the dependence of recalibration and enhancement on the masker-target level combination is not identical. For instance, a 90 dB SPL masker causes more enhancement than recalibration in a 60 dB SPL target. The two-process model can thus be refined as follows. The process causing loudness enhancement or loudness decrement (Elmasian *et al.*, 1980) is effective if two tones are presented within a temporal window of about 400 ms (Zwislocki and Sokolich, 1974; Arieh and Marks, 2003a), because loudness enhancement is observed with both forward and backward maskers (e.g., Elmasian and Galambos, 1975). The effect is maximal at intermediate masker-target level differences (Experiment 2; Zeng, 1994; Plack, 1996; Oberfeld, 2003). It can be assumed that loudness enhancement is effective only if the masker and the target are similar in frequency (Zwislocki and Sokolich, 1974). These effects cannot be explained by mechanisms located in the auditory periphery, because at early processing stages, forward maskers have been found to *reduce* rather than to enhance the neural response to a following test tone (e.g., Bauer *et al.*, 1975; Harris and Dallos, 1979; Relkin *et al.*, 1995), and because it does not seem possible that a backward masker following the target by 100 ms alters the representation of the target in the auditory nerve. A model based on more centrally located mechanisms was proposed by Elmasian *et al.* (1980). They suggested that the loudness representations of the masker and the target are merged automatically. Applied to the three-tone matching task, it follows that the initial value of target loudness is no longer available at the presentation of the comparison, but that the listener will instead compare a weighted average of the masker loudness and the target loudness with the loudness of the comparison. The mergence hypothesis can explain why the loudness is reduced if the masker is less intense than the target, while a more intense masker results in enhancement—in other words, why the loudness of the target always seems to be shifted towards masker loudness.[2] On the other hand, the mergence hypothesis alone cannot account for the mid-level hump in loudness enhancement (Zeng, 1994; Plack, 1996). Given a constant masker level of, e.g., 90 dB SPL, loudness enhancement should increase monotonically with decreasing target level if a simple weighted average between masker loudness and target loudness was used in the loudness match. The same argument applies to the mid-difference hump observed for three listeners in Experiment 2. Oberfeld (2005) proposed that it is possible to resolve these problems by assuming that the effect of the masker depends on the *perceptual similarity* between masker and target, that is, that the masker loudness will receive a smaller weight if the masker and the target differ strongly in, e.g., spectral content, duration, or loudness. Effects of the masker-standard similarity on intensity-difference limens were reported by Schlauch *et al.* (1997, 1999). The finding by Elmasian and Galambos (1975) that a diotic masker combined with a monaural target produced a smaller amount of loudness enhancement than an ipsilateral masker is also compatible with a similarity effect.

The second process, causing loudness recalibration, is assumed to have a fast onset, but a slow decay in the order of several seconds (Arieh and Marks, 2003a; Arieh *et al.*, 2005). It is effective only for tones similar in frequency (Marks, 1994), and only for masker durations longer than or equal to target duration (Nieder *et al.*, 2003). It reaches its maximum at masker levels 10–20 dB above target level and does not further increase with masker level (Mapes-Riordan and Yost, 1999; Nieder *et al.*, 2003). The effect is larger if the target is presented at intermediate rather than at low levels (Experiments 1 and 2; Mapes-Riordan and Yost, 1999). Loudness recalibration is viewed as a centrally based, adaptation-like process (Marks, 1996; Arieh and Marks, 2003b), although the exact nature of the mechanism remains unclear (cf. Wagner and Scharf, 2006).

Under the assumption that loudness enhancement decays within about 400 ms following masker onset, but that ILR remains constant for several seconds, it is possible to independently estimate the amount of loudness enhancement and ILR a masker causes in a proximal target. The estimate of ILR would be the loudness reduction measured with the masker-target ISI well above 500 ms in a three-tone loudness matching task presenting the comparison at a different frequency than the masker. In this case, loudness enhancement of the target can be assumed to be absent. The estimate of loudness enhancement would be based on the loudness change measured at the short masker-target ISI in the same three-tone loudness matching task. Because according to the two-process model the masker causes both loudness enhancement and ILR in a proximal target, the estimate of loudness enhancement would be the change in loudness level at the short interval *plus* the change in loudness level measured with the long masker-target ISI. According to this rationale, loudness matches (with masker and comparison differing in frequency) obtained by Arieh and Marks (2003a) for a masker-target ISI of 75 and 1650 ms indicate that an 80 dB SPL masker caused on average 11.1 dB of ILR in a 60 dB-target, and 11.2 dB of loudness enhancement. The latter value is roughly compatible with the change in loudness level in similar conditions found in experiments where masker and comparison shared the same frequency (e.g., Elmasian and Galambos, 1975). In fact, if the two-process model is valid, it would follow that the loudness matches obtained with the comparison and the masker presented at the same frequency are actually an estimate of loudness enhancement of the proximal target. According to the model, the match reflects loudness enhancement of the target, ILR of the target, and ILR of the comparison. Because the latter two effects are assumed to be identical in size, they should cancel. The necessary assumption would be that the target interpolated between masker and comparison does not influence the loudness recalibration induced in the comparison, however. Additional data are necessary to test this hypothesis.

ACKNOWLEDGMENTS

I am grateful to Armin Kohlrausch for pointing out the importance of measuring the reduction in comparison loudness with exactly the same two tones preceding it as in the three-tone loudness matching task. I thank Andrew Oxenham, Yoav Arieh, and an anonymous reviewer for helpful comments concerning an earlier version of this manuscript.

[1]In some other studies, the alternative terms *conditioner* or *inducer tone* have been used for the masker, and the terms *test tone* or *standard* for the target.

[2]The concept underlying the mergence hypothesis is similar to adaptation-level theory (Michels and Helson, 1954) and the sensation weighting model by Hellström (1977). These models explain the time-order error (Fechner, 1860) by assuming that if two sequentially presented sounds are compared for their loudness, the representations used in the comparison process are influenced by the context, for example, by the average loudness of all preceding stimulation. The representation of target loudness is therefore assumed to be a weighted average between the momentary sensation and "adaptation level." Hellström (1985) noted that loudness enhancement and decrement can be explained if one assumes that the masker influences the adaptation level effective for the target. If so, the remembered loudness of the target would drift towards masker loudness during the target-comparison interval. The explanation rests on the (reasonable) assumption that the adaptation level effective for the target is influenced more strongly by the masker than the adaptation level effective for the comparison, due to the relative temporal proximity between masker and target.

Arieh, Y., Kelly, K., and Marks, L. E. (**2005**). "Tracking the time to recovery after induced loudness reduction (L)," J. Acoust. Soc. Am. **117**, 3381–3384.

Arieh, Y., Mailloux, J. R., and Marks, L. E. (**2004**). "Loudness recalibration at short ISI: A closer look," J. Acoust. Soc. Am. **115**, 2600(A).

Arieh, Y., and Marks, L. E. (**2003a**). "Time course of loudness recalibration: Implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.

Arieh, Y., and Marks, L. E. (**2003b**). "Recalibrating the auditory system: A speed- accuracy analysis of intensity perception," J. Exp. Psychol. Hum. Percept. Perform. **29**, 523–536.

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.

Buus, S., Müsch, H., and Florentine, M. (**1998**). "On loudness at threshold," J. Acoust. Soc. Am. **104**, 399–410.

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.

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.

Epstein, M., and Gifford, E. (**2006**). "A potential carry-over effect in the measurement of induced loudness reduction," J. Acoust. Soc. Am. **120**, 305–309.

Fechner, G. T. (**1860**). *Elemente der Psychophysik* (Breitkopf und Härtel, Leipzig).

Galambos, R., Bauer, J., Picton, T., Squires, K., and Squires, N. (**1972**). "Loudness enhancement following contralateral stimulation," J. Acoust. Soc. Am. **52**, 1127–1130.

Harris, D. M., and Dallos, P. (**1979**). "Forward-masking of auditory nerve fiber responses," J. Neurophysiol. **42**, 1083–1107.

Hellström, Å. (**1977**). "Time errors are perceptual. An experimental investigation of duration and a quantitative successive-comparison model," Psychol. Res. **39**, 345–388.

Hellström, Å. (**1985**). "The time-order error and its relatives: Mirrors of cognitive processes in comparing," Psychol. Bull. **97**, 35–61.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.

IEC 318 (**1970).** *An IEC artificial ear, of the wide band type, for the calibration of earphones used in audiometry* (International Electrotechnical Commission, Geneva).

Levitt, H. (**1971**). "Transformed up-down procedures in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

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.

Marks, L. E. (**1996**). "Recalibrating the perception of loudness: Interaural transfer," J. Acoust. Soc. Am. **100**, 473–480.

Michels, W. C., and Helson, H. (**1954**). "A quantitative theory of time-order effects," Am. J. Psychol. **67**, 327–334.

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.

Oberfeld, D. (**2003**). "Intensity discrimination and loudness in forward masking: The effect of masker level," in *Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 29 Jahrestagung für Akustik DAGA '03, Aachen*, edited by Deutsche Gesellschaft für Akustik (Berlin: DEGA), pp. 606–607.

Oberfeld, D. (**2005**). *Alteration of Intensity Resolution and Loudness by Sounds Presented in Temporal Proximity*. Doctoral Dissertation, Technical University Berlin. Web site of the German National Library: http://nbn-resolving.de/urn:nbn:de:kobv:83-opus-10616.

Plack, C. J. (**1996**). "Loudness enhancement and intensity discrimination under forward and backward masking," J. Acoust. Soc. Am. **100**, 1024–1030.

Plack, C. J., and Carlyon, R. P. (**1995**). "Loudness perception and intensity coding," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 123–160.

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.

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 (Papst, Berlin), pp. 254–259.

Scharf, B., Buus, S., and Nieder, B. (**2002**). "Loudness enhancement: Induced loudness reduction in disguise? (L)," J. Acoust. Soc. Am. **112**, 807–810.

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., 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.

Suzuki, Y., and Takeshima, H. (**2004**). "Equal-loudness-level contours for pure tones," J. Acoust. Soc. Am. **116**, 918–933.

Wagner, E., and Scharf, B. (**2006**). "Induced loudness reduction as a function of exposure time and signal frequency," J. Acoust. Soc. Am. **119**, 1012–1020.

Yost, W. A. (**2000**). *Fundamentals of Hearing. An Introduction* (Academic, San Diego).

Zeng, F.-G. (**1994**). "Loudness growth in forward masking: Relation to intensity discrimination," J. Acoust. Soc. Am. **96**, 2127–2132.

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.

Zwislocki, J. (**1965**). "Analysis of some auditory characteristics," in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. B. Bush, and E. Galanter (Wiley, New York), Vol. 3, pp. 3–97.

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.