Does induced loudness reduction explain contextual effects in loudness judgment?

Lawrence Marks

John B. Pierce Laboratory, 290 Congress Avenue, New Haven, CT 06519, USA
marks@jbpierce.org

Briefly presenting an inducing tone of 70-80 dB can substantially reduce the loudness of a subsequent test tone at or near the inducer's frequency, a phenomenon called induced loudness reduction (ILR). The study of ILR emerged from earlier observations on differential contextual effects in loudness judgment: Tones of a given SPL and frequency were judged softer when presented as part of an ensemble of high rather than low SPL tones at the same frequency, relative to judgments of loudness of tones at a different sound frequency. At first, these effects of stimulus context on loudness judgment were assumed to reflect decisional processes, that is, to reflect biases in loudness judgment. On the other hand, ILR is often assumed to reflect a depression in the intensity response of the auditory system. While it is tempting to explain differential contextual effects in loudness judgment wholly in terms of ILR, the properties of ILR and the properties of contextual effects may not be identical, leaving open a possible role for decisional processes as well as sensory processes in contextual effects, and perhaps also in ILR, in both laboratory and 'real-world' settings.

1 Introduction

The acoustical environment is typically in continuous flux. Sounds come and go, changing over time in frequency composition and in intensity. The loudness of a sound heard at any moment reflects not only the physical characteristics of that sound, especially the distribution of acoustical power over the frequency spectrum, but also the history of acoustical stimulation to which the listener has recently been exposed. To say this is to recognize that people may judge a given sound, a sound of fixed physical characteristics, to be louder or softer depending on the physical properties of previous sounds – on how they were perceived and judged. Presenting sound 1 before sound 2 can alter the loudness of sound 2, even when sound 2 follows sound 1 by several hundred milliseconds or several seconds, and even when the total effective energy of sound 1 falls well below the level needed to produce auditory fatigue.

Psychoacoustic research in recent years, using different experimental paradigms, has identified two new time-dependent phenomena: induced loudness reduction and differential context effects. The main questions posed here are these: Are induced loudness reduction and differential context effects simply two faces of the same phenomenon? In other words, does the same mechanism, or do the same mechanisms, underlie both?

2 Induced loudness reduction

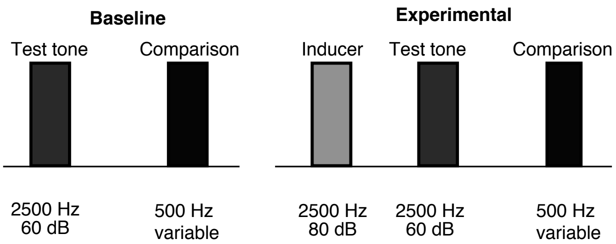

Fig. 1 Stimulus sequences.
Sequences of stimuli used to measure induced loudness reduction.

Presenting a brief tone of 70-80 dB SPL (an 'inducer') can reduce substantially the loudness of a subsequent tone at or near the inducer's frequency [1, 2, 3, 4]. This effect is commonly called induced loudness reduction, or ILR [2, 3]. A paradigm often used to measure ILR is depicted in Fig. 1. The SPL of a comparison tone is varied from trial to trial in order to match the loudness of a test tone in the presence of an inducer that precedes the test tone (experimental condition) and in the absence of the inducer (baseline condition). Typically, the test tone has the same frequency as the inducer, but the comparison tone has a very different frequency, in order to minimize the effect of the inducer on the loudness of the comparison. By definition, ILR equals the difference between the matching SPLs in the baseline and experimental conditions. In general, subjects set the SPL of the comparison lower in the experimental condition relative to baseline, thereby quantifying how the inducer reduces the loudness of the test tone.

Research using the paradigm of Fig. 1 and variants of it has shown, for example, that a brief inducer can reduce the loudness of a subsequent test tone by as much as 10 dB or more [1, 5]. An 80-dB SPL inducer can substantially affect test tones with SPLs lying 10-20 dB below the level of the inducer, while having little effect on signals at the same level of the inducer (80 dB), and no effect at all on signals around absolute threshold [1]. Although many studies have measured ILR using inducers with SPLs at 80 dB [1, 2, 4, 5], evidence of some ILR has been reported with inducers of 68-73 dB [6]; no ILR was found, however, with inducers of 40 dB [1].

3 Differential context effects

Condition A 500 Hz	**Condition A 2500 Hz**	**Condition B 500 Hz**	**Condition B 2500 Hz**
35 dB			30 dB
40 dB			35 dB
45 dB			40 dB
50 dB			45 dB
55 dB	**55 dB**	**50 dB**	**50 dB**
60 dB	**60 dB**	**55 dB**	**55 dB**
65 dB	**65 dB**	**60 dB**	**60 dB**
70 dB	**70 dB**	**65 dB**	**65 dB**
	75 dB	70 dB	
	80 dB	75 dB	
	85 dB	80 dB	
	90 dB	85 dB	

Fig. 2 Stimulus ensembles.
Ensembles of stimuli used to produce and measure differential context effects in loudness.

The identification and subsequent study of ILR emerged from earlier observations on what may be called differential contextual effects (DCEs) in loudness judgment. Differential context effects were first reported in the results of a series of experiments in which subjects gave magnitude estimates of the loudness of pure tones that varied in both sound frequency and SPL [7]. These experiments asked in particular how changing the ensemble of SPLs at two sound frequencies presented over the course of an experiment affects loudness judgments.

Consider the following example: In one condition of stimulus context (condition A), the SPLs at a low frequency, say, 500 Hz, are relatively weak and those at a higher frequency, say, 2500 Hz, are stronger, whereas in a second condition (condition B), the SPLs at 500 Hz are relatively strong and those at 2500 Hz are weaker. A subset of four SPLs at 500 Hz is common to both contextual conditions, as is a subset of four SPLs at 2500 Hz [7]. These stimulus contexts are depicted in Fig. 2.

The average judgments of loudness obtained in the two conditions of Fig. 2 appear in Fig. 3 [7]. Note that the four common SPLs at 500 Hz are judged to be softer in condition B than in condition A, whereas all four common stimuli at 2500 Hz are judged to be softer in A than in B. That is, in each condition, loudness is smaller at the frequency whose 'contextual stimuli' are high in SPL.

Findings of many experiments have given the same outcome: The relative loudness of a tone of given SPL and frequency is judged to lower when the ensemble of SPLs at that frequency is strong rather than weak [6, 7, 8]. The results shown in Fig. 3 imply that stimulus context affected loudness more or less equivalently at the two frequencies. From these data it is not possible to determine unequivocally, however, whether the changes in loudness at a given frequency comprise reductions in loudness when the contextual stimuli at that frequency are relatively high, increases in loudness when the contextual stimuli at that frequency are relatively low, or both.

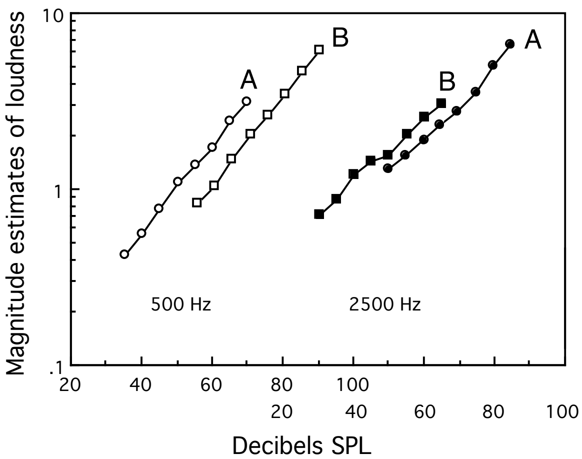

Fig. 3 Differential context effects.
Average magnitude estimates of loudness of 500-Hz and 2500-Hz tones, under two conditions of stimulus context. Data from [7].

To conclude that loudness at 500 Hz is greater in condition A compared to B and that loudness at 2500 Hz is greater in condition B compared to A is to imply the presence of what may be called absolute context effects: that stimulus context affects loudness at both frequencies. In order to infer the presence of absolute effects from the results shown in Fig. 3, however, it is necessary to assume that, on average, the subjects used their rating scale (in this case, an open-ended magnitude-estimation scale) in the same way in both conditions. Although it is reasonable that they did, it is possible that the subjects simply gave bigger numbers in one condition than the other. If so, then loudness might have changed across conditions at one frequency but not at the other. Without knowing how the subjects used the magnitude-estimation scale in the two conditions, it is impossible to determine conclusively whether shifts in stimulus context produce increases in loudness at one frequency, decreases at the other, or both.

Even without assuming that subjects used the response scale in the same way in both conditions, it is possible to infer from the data of Fig. 3 the presence of contextual effects: Changes in stimulus context clearly affected the loudness at (at least) one of the two sound frequencies, relative to loudness at the other frequency. Consider the responses to the two sets of common SPLs. Loudness at 500 Hz was *relatively* greater than loudness at 2500 Hz in condition A compared to condition B. The change across conditions in relative loudness constitutes a differential context effect (DCE).

The presence of DCEs in loudness judgment is not limited to experimental paradigms in which subjects give numerical estimates; DCEs also appear in paradigms in which the ensembles of SPLs at different frequencies vary across conditions but the subjects simply compare loudness in a forced-choice (paired-comparison) task [1, 8].

From the stimulus ensembles depicted in Fig. 2, it is possible to construct, in each condition, A and B, a set of 64 different pairs of tones, pairing each of the 8 possible SPLs at 500 Hz with each of the 8 possible SPLs at 2500 Hz. Then, on every trial, the subject hears the two tones of the pair in sequence and responds by indicating which tone is louder. In this direct (paired) comparison paradigm, 16 stimulus pairs are common to the two contextual conditions, and the paired comparisons of loudness are seen to follow directly from the judgments obtained in the rating paradigm. For any given common pair, subjects choose the 500-Hz tone to be louder than the 2500-Hz tone more often in condition A than condition B [8]. Again, however, it is not possible to discern from these results whether these DCEs constitute reductions in loudness, increases in loudness, or both.

4 Sensory and decisional processes

4.1 Models

It was initially suggested that DCEs in loudness might arise from response biases operating in tasks of numerical judgment [7], but this is unlikely. The presence of DCEs in tasks requiring only direct comparisons of loudness [1, 8], for example, helped eliminate numerical response bias as a possible explanation. But these findings and others still leave open the possibility that a broader class of decisional processes, not just biases in numerical response, might partly or wholly generate the context effects. According to a more general decisional hypothesis, the first few context-inducing stimuli might not change the sensory representations of loudness of any of the stimuli, but might change the decisional processes by which subjects judge loudness; that is, context may change the rules that operate in the underlying decision space. Sensory and decisional models of DCEs are characterized in Fig. 4.

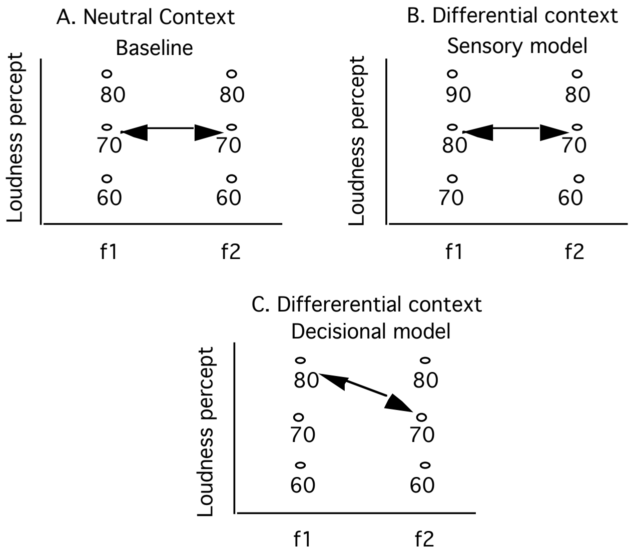

Fig. 4 Representations of loudness in decision space. Decision spaces for perception and judgment of loudness before and after exposure to context-inducing stimuli, according to sensory and decisional models.

The decision spaces sketched in Fig. 4 could underlie judgments of loudness of various kinds (magnitude estimates, direct comparisons). In the figure, different sound frequencies appear as the abscissa, and representations of sound intensity, that is, representations of loudness, appears as the ordinate.

The panel in the upper left depicts the representations of loudness at two frequencies, f1 and f2 under neutral conditions – for example, prior to the presentation of any differential context-inducing stimuli. For simplicity, it is assumed that in neutral conditions, equal SPLs at f1 and f2 produce equal loudness and equal judgments of loudness (as indicated by the arrow): A 70-dB stimulus at f1 is judged equal to a 70-dB stimulus at f2. The panels in the upper right and at the bottom show two possible organizations of decision space after exposure to context-inducing stimuli in which SPLs are greater at f1 than f2.

The panel in the upper right characterizes a sensory model, which assumes that the context-inducing stimuli reduce loudness at f1 relative to loudness at f2. As a result, SPL must be greater at f1 than at f2 to produce equal loudness (displacement of scales at f1 and f2) and equal judgments of loudness (horizontal arrow). As in the baseline condition, the sensory model assumes that subjects 'read off' relative loudness by comparing values at different frequencies on a vector that is orthogonal to the ordinate.

Finally, the panel at the bottom characterizes a decisional model. This decisional model assumes that the context-inducing stimuli have no effect on the representations of loudness – the locations of the stimuli at f1 and f2 are identical to those in the baseline condition. Instead, the decisional model assumes that the context-inducing stimuli lead the subjects to make their judgments by comparing values at f1 and f2 along a vector that is not orthogonal to the ordinate, but tilted so that a greater value on the ordinate is required at f1 than f2 to judge the loudness as equal. Thus, Fig. 4 provides a spatial representation of a decisional process by which the stimulus context could produce DCEs. Intense stimulation at f1 leads subjects to require a higher level of loudness at f1 relative to f2 in order to report matching loudness at the two frequencies.

4.2 Evidence

A preponderance of evidence has led to the conclusion that DCEs do represent changes in the underlying sensory representations of loudness per se, and not (just) changes in decisional processes [5, 9, 10]. In part, this evidence comes from experiments using contextual paradigms, for instance, from a study showing that varying relative SPLs of tones at 500 and 2500 Hz not only modifies relative loudness judgments, as in Fig. 3, but also modifies response times and errors in a speeded choice task in which subjects have to identify the sound frequency [11]. Importantly, the effects of context on rapid choice responses could not be attributed to a trade-off between response times and errors, such speed-accuracy trade-offs being the hallmarks of changes in decision criteria.

Even if sensory processes do provide the main source of DCEs, it is nevertheless possible that decisional mechanisms contribute to DCEs, especially when subjects give estimates of loudness. One early finding supports this possibility: DCEs were notably smaller when subjects judged loudness using the method of absolute magnitude estimation rather than traditional magnitude estimation [7]. Instructions associated with absolute magnitude estimation eschew any reference to perceptual ratios but simply ask subjects to assign numbers whose subjective magnitudes 'match' the magnitudes of the loudness sensations. It has been argued that these instructions reduce tendencies for subjects to make relative judgments [12], and tendencies to make relative judgments might well contribute to DCEs.

5 Induced loudness reduction and differential context effects

The questions posed at the end of the Introduction were: Are ILR and DCEs simply two faces of the same phenomenon? Does the same mechanism, or do the same mechanisms, underlie both?

If the answers to these questions are positive, then ILR and DCEs should have the same quantitative properties. For example, if ILR involves only reductions in loudness, as its label implies and as seems to be the case, then so too should DCEs. Unfortunately in this regard, there are few data if any directly pertinent to DCEs. Most of the evidence pointing to reductions in loudness actually comes from studies using variants of ILR paradigms rather than DCE paradigms. For example, listening to loud tones at f1 decreases the probability that a subsequent test tone at f1 is judged as loud as a previously matching tone at f2; but listening to soft tones has no effect [6]. The conclusion that DCEs constitute only reductions in loudness comes largely from studies that essentially tested ILR, so the conclusion assumes, at least implicitly, that DCEs depend wholly on ILR.

Are there any serious challenges to the hypothesis that ILR wholly underlies DCEs? To be sure, a great deal of evidence implies that ILR and DCEs arise *largely* from a common mechanism or set of mechanisms, whatever it (or they) may be. First and foremost, the phenomenal characteristics of ILR and DCEs bear several resemblances. Both ILR and DCEs require the presentation of moderately intense inducers on the one hand [1], or context-inducing stimuli on the other [8]. Both are frequency specific [13, 14], although ILR at least seems to be fairly broadly tuned in sound frequency [14]. And both can last for relatively long periods of time, at least up to several minutes [4, 15].

5.1 Intensity tuning in ILR

But are the quantitative psychophysical properties of ILR and DCEs *identical*? Perhaps not. With an inducer of 80 dB SPL, Mapes-Riordan and Yost [1] first reported evidence of what may be termed 'intensity tuning' of ILR. Inducers of 80 dB reduced the loudness of subsequent test tones at 60-70 dB by as much as 11 dB (though sometimes by less), but reduced the loudness of test tones at 40 dB by only 4 dB, reduced the loudness of test tones at 80 dB by only 1-3 dB, and did not affect threshold at all. Although results vary somewhat from study to study, the upshot is that ILR requires inducers of at least 65-80 dB and mainly affects test tones 10-20 dB lower, in the region 50-70 dB. And the intensity tuning of ILR may be even sharper than this if, as Epstein [10] has argued, the small amount of ILR reported with inducers and test tones at 80 dB reflects an adventitious reduction in the loudness of the comparison tone. Reanalyzing magnitude estimations of loudness reported by Hellman and Zwislocki [16], Epstein also noted a strong tendency for tones of 50-70 dB in one subject (40-60 dB in another) to be judged louder on the first replicate than on subsequent replicates – consistent with the hypothesis that, over the course of the sessions, ILR occurred, affecting mainly the loudness of tones of medium intensity, relative to lower or higher ones. If relatively sharp intensity tuning characterizes ILR, and if DCEs depend exclusively on ILR, then similarly sharp intensity tuning should also characterize DCEs.

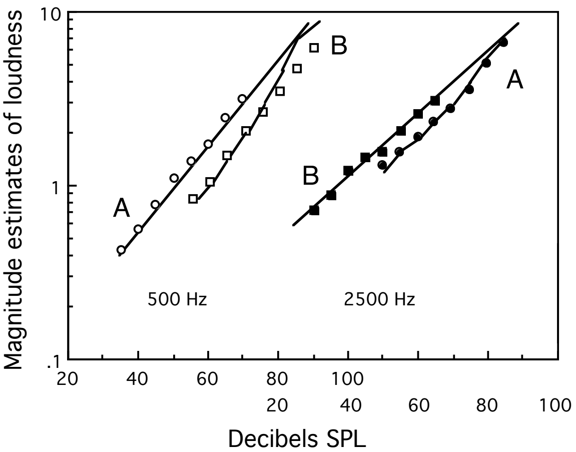

Fig. 5 Differential stimulus context effects predicted by induced loudness reduction (ILR), assuming ILR shows modest intensity tuning. Data from [7], as in Fig. 3.

5.2 Intensity tuning in DCEs

Let us consider a DCE experiment with the stimulus ensembles shown in Fig. 2 and let us make three assumptions. First, assume that ILR is substantial where SPLs are high (2500 Hz in condition A and 500 Hz in condition B), but negligible where SPLs are low (500 Hz in condition A and 2500 Hz in condition B). Second, assume that intensity tuning is modestly sharp; in particular, let us assume ILR equal to the average values found by Mapes-Riordan and Yost [1] using an inducer at 80 dB: ILR of 2, 7, 9, and 4 dB with test tones of 80, 70, 60, and 40 dB, respectively. And third, let us assume that ILR can account fully for DCEs. From these three assumptions, it should be possible to predict the pattern of data of Fig. 3.

Replotted in Fig. 5 are the data of Fig. 3, together with the predictions. These predictions were generated by fitting straight lines (power functions) to the data at 500-Hz in condition A and 2500 Hz in condition B, assuming neither to show ILR, and then by reducing loudness at 500 Hz in condition B and 2500 Hz in condition A by the average amounts of ILR given above. Critical are the predictions for these latter two functions.

The model of ILR predicts loudness fairly well at 2500 Hz in condition A, but does less well predicting loudness at 500 Hz in condition B. It is worth noting that if intensity tuning of ILR is even sharper than what is assumed here, for instance, if ILR is wholly absent at 80 dB [10], then the predictions made from ILR would fare even worse.

A substantial number of experiments have manipulated the contextual sets of two or more sound frequencies [6, 7, 8, 13], and the upshot is that DCEs tend to be relatively constant when the effects are calculated in dB – an outcome that is hard to square with sharp intensity tuning of ILR. Indeed, loudness functions observed under different contextual conditions suggest that intense context-inducing stimuli at a given frequency affect loudness more or less proportionally at all SPLs (as in Fig. 3), perhaps even affecting the loudness of the inducers themselves. In brief, measures of ILR appear to underestimate the magnitude of DCEs observed at relatively high levels of SPL.

To be sure, the assumptions that underlie the predictions here are oversimplified and undoubtedly imprecise. The predictions rely on ILR measured with inducers at 80 dB, but contextual designs typically present multiple SPLs, several of which may produce ILR. How ILR may combine across inducers differing in SPL is unknown. But even assuming that, in contextual designs, several stimuli act as inducers, it is difficult to account for the evidence of that DCEs can be as substantial at 70-80 dB as at lower SPLs.

6 Decisional processes in DCEs

Differential effects of stimulus context are readily observed when subjects rate loudness. There is little doubt that ILR accounts for the lion's share of the changes in loudness judgment produced by reciprocally shifting sets of SPLs at two frequencies. The question at hand, however, is whether ILR accounts *entirely* for DCEs measured in loudness-rating experiments (and perhaps in other tasks). The evidence at hand suggests that it may not.

Effects of stimulus context are well known to pervade psychophysical judgments, including judgments of loudness, and some contextual effects doubtless reflect high-level decisional or cognitive processes. Among these are the well-known contrast effects and assimilation effects. With contrast effects, which sometimes can mimic adaptation, an intense stimulus on trial n-1 reduces the judgment to a weaker stimulus on trial n, whereas a weak stimulus on trial n-1 increases the judgment to a stronger stimulus on trial n. Assimilation effects act in reverse; An intense stimulus on trial n-1 increases the judgment of intensity of a weaker stimulus on trial n, whereas a weak stimulus on trial n-1 decreases the judgment of intensity of a stronger stimulus on trial n.

Assimilation effects can be potent in loudness judgments. When the set of SPLs of a tone shifts abruptly within a test session, the judgments of loudness reveal assimilative context effects that can be equivalent to more than 5 or even 10 dB [6]. Shifting the set of SPLs up or down increases or decreases, respectively, the loudness rating of tones that are common to the different sets – the reverse of what happens in ILR. Assuming that ILR occurs too in these experiments (the highest SPLs in the experiments just cited were 85 and 90 dB), the net assimilation effect observed presumably equalled the difference between the magnitude of the underlying 'true' assimilation and the magnitude of the offsetting ILR. Most importantly, assimilation effects, unlike ILR, appear to be largely, perhaps entirely, a result of decisional processes [17].

The other side of this coin is that when sets of SPLs at two frequencies are shifted reciprocally so that DCEs result, the judgments of loudness are likely to reflect assimilation in judgment as well as ILR. The design of the experiment that produced the data of Figs. 3 and 5 aimed to equalize the effects of assimilation across stimuli within each condition: To accomplish this, in each test session, every stimulus followed every stimulus in the ensemble, including itself, exactly once [7]. Even so, it is not certain that assimilation effects depend on only the previous stimulus (or, more likely, on the response to the previous stimulus). More importantly, the discrepancies between the measurements of DCEs and the predictions from ILR shown in Fig. 5 could conceivably reflect, at least in part, intensity-dependent variations in loudness assimilation.

7 Decisional processes in ILR?

Epstein [10] has pointed out how, in some ILR designs, there could be adventitious loudness reduction of the comparison tone, which in turn would lead to spurious overestimation of ILR in the test tone. It is plausible that decisional processes could affect observed measures of DCEs. And if so, then might not decisional processes also play a role in paradigms of ILR?

8 Conclusion

Induced loudness reduction and differential context effects are undoubtedly related, and it is reasonable to start with the hypothesis that the same mechanisms underlie both. A sensory mechanism of induced loudness reduction undoubtedly figures prominently in both ILR and DCEs, but decisional mechanisms, perhaps even the same decisional mechanism, may also contribute to both. If this is so, then both the sensory mechanisms of loudness reduction and the decisional mechanisms of loudness judgment doubtless operate not only in laboratory experiments, but also in everyday life outside the laboratory – in the perception of speech, music, and noises in homes, workplaces, and on city streets.

References

[1] D. Mapes-Riordan, W.A. Yost, "Loudness recalibration as a function of level", *J. Acoust. Soc. Am.* 106, 3506-3511 (1999)

[2] B. Nieder, S. Buus, M. Florentine, B. Scharf, "Interactions between test- and inducer-tone durations in induced loudness reduction", *J. Acoust. Soc. Am.* 114, 2846-2855 (2003)

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

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

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

[6] L.E. Marks, "Contextual processing of unidimensional and multidimensional auditory stimuli", *J. Exp. Psychol. Hum. Percept. Perform.* 19, 227-249 (1993)

[7] L.E. Marks, "Magnitude estimation and sensory matching", *Percept. Psychophys.* 43, 511-525 (1988)

[8] L.E. Marks, "'Recalibrating' the auditory system: The perception of loudness", *J. Exp. Psychol. Hum. Percept. Perform.* 20, 382-396 (1994)

[9] L.E. Marks, Y. Arieh, " Differential effects of stimulus context in sensory processing", *Eur. Rev. Appl. Psychol.* 56, 213-221 (2007)

[10] M. Epstein, "An introduction to induced loudness reduction", *J. Acoust. Soc. Am. (online)* 122, EL74-EL80 (2007)

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

[12] J.J. Zwislocki, "Group and individual relations between sensation magnitudes and their numerical estimates", *Percept. Psychophys*. 33, 460-468 (1983)

[13] L.E. Marks, E. Warner, "Slippery context effect and critical bands", *J. Exp. Psychol. Hum. Percept. Perform.* 29, 523-536 (2003)

[14] J. Marozeau, M. Epstein, "Induced loudness reduction as function of frequency difference between test tone and inducer", *Percept. Psychophys*. 70, 725-731 (2008)

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

[16] R.P. Hellman, J.J. Zwislocki, "Monaural loudness functions at 1000 cps and interaural summation", *J. Acoust. Soc. Am.* 35, 856-865 (1963)

[17] L.M. Ward, "Critical bands and mixed-frequency scaling: Sequential dependencies, equal-loudness contours, and power function exponents", *Percept Psychophys.* 47, 551-562 (1990)