

PUBLISHED IN
PERCEPTION AND PSYCHOPHYSICS,
2008, 70(4), pp725-731.

**Induced Loudness Reduction as function of
frequency difference between test tone and
inducer**

by

Jeremy Marozeau¹, Michael Epstein²,

1 Communication Research Laboratory

2. Auditory Modeling and Processing Laboratory

Dept. of Speech-Language Pathology and
Audiology (106A FR)

Northeastern University

360 Huntington Avenue, Boston, MA 02115 U.S.A.

E-mail: j.marozeau@neu.edu

Tel : 617 373 2587

*A portion of this work was presented at 151st meeting of the
Acoustical Society of America [Marozeau, J., Epstein, M., and
Florentine, M. (2006): "Induced Loudness Reduction as a
function of frequency separation"]*

ABSTRACT

When a high-intensity tone (inducer) is followed by a moderate-intensity tone (test tone), the loudness of the latter is reduced. This phenomenon, Induced Loudness Reduction (ILR), depends on the frequency separation of the two tones; as the difference in frequency increases, the amount of ILR decreases. However, the precise course of this decrease is not well known. This paper presents two experiments that address this question. In the first experiment, the amount of loudness reduction produced by a 2.5 kHz 80-dB-SPL inducer was measured with the frequency of the test tone swept from 800 Hz to 6 kHz. In the second experiment, the amount of ILR was measured with the same inducer and with test tones set at 2, 2.5, 3, and 4 kHz. Both experiments show that some ILR occurs at frequency separations as wide as four critical bands. [Supported by NIH/NIDCD grant R01DC02241.]

I. INTRODUCTION

Psychoacoustical experiments examining loudness have always been prone to unexplained variability, typically attributed to biases resulting from subtle procedural differences between experiments (see Poulton, 1989 for review). Very often, these biases are explained using high-level perceptual concepts like learning effects, range effects, or subject preferences. More recently, a number of studies have examined the precise contexts in which some of these types of discrepancies appear and have found specific patterns. In many loudness experiments, the loudness of a sound may be reduced when it is preceded by a higher-intensity sound (inducer) at a close frequency with an equal or longer duration. This phenomenon, initially noted as a "slippery-context" effect (Marks, 1992), has been more recently termed Induced loudness reduction (**ILR**) or loudness recalibration (Arieh and Marks, 2001; Mapes-Riordan and Yost, 1997; Mapes-Riordan and Yost, 1999; Marks, 1994).

Induced Loudness Reduction is a pervasive effect of debated origin and has the potential to have unexpected effects on data collected during loudness experiments involving the presentation of a range of levels, particularly any experiment that groups moderate- and high-level presentations. Notably, it has been shown that the amount of ILR is greater at moderate levels (Mapes-Riordan and Yost, 1999). Because the effects of ILR accumulate over time, results obtained in the early portions of experiments, prior to significant sound exposure, may differ markedly from results obtained in later portions of experiments (Arieh *et al.*, 2005). Epstein (2007) examined the results of a magnitude estimation study by Hellman and Zwislocki (1963). The magnitude-estimation data indicated that over the course of the experiment, loudness changed more for moderate-level sounds than it did for low- and high-level sounds. The elucidation of this effect raises important questions regarding the "true" nature of the loudness function, particularly the shallow, compressive region seen at moderate levels, the origin of which is still in debate (Buus *et al.*, 1999; Epstein and Florentine, 2005; Ward *et al.*,

1996; Parker and Schneider, 1994; Schneider and Parker, 1990; Braida and Durlach, 1972; Florentine and Epstein, 2006). Additionally, loudness calculation and modeling for sound has typically been done a priori using only the present physical stimulus parameters. In fact, the effects of ILR provide significant motivation to generate loudness models that include not only immediate physical parameters, but also long-term context parameters.

It is however difficult to model the effects of ILR on a particular experiment because of the dependence on several factors including: (1) tone levels (Mapes-Riordan and Yost, 1999); (2) frequency separation between inducer and test tones (Marks, 1994); (3) durations of inducer and test tones (Nieder *et al.*, 2003); (4) time separation between inducer and test tones (Arieh and Marks, 2003); (5) number of exposures to inducers (Arieh *et al.*, 2005); and (6) individual differences (Epstein and Gifford, 2006; Wagner and Scharf, 2006).

Previously, the effects of frequency on ILR were not thoroughly examined. Marks and Warner (1991) showed that loudness functions are affected by the frequency-specific context of measurements and Marks (1994) showed that loudness matches between different frequency tones are affected by the frequency separation. These studies imply that inducers with nearly the same frequency as the test tone cause maximal ILR and inducers that are at very different frequencies from the test tone do not cause ILR. However, the relationship between the frequencies of the inducer and test tone and the amount of ILR has not been directly tested and is not well understood for small-to-moderate frequency spacings. Many previous loudness experiments are based on loudness matches between sounds at nearby frequencies. In order to determine what possible influence ILR may have had on those data, it is necessary to gain a further understanding of how ILR changes as a function of frequency spacing. Therefore, the present study examines the amount of ILR as a function of the frequency separation between the inducer and test tone. The first experiment, a modified equal-loudness-tracking Békésy procedure, examines this relationship by sweeping in small frequency

increments. Then, a follow-up experiment examines the amount of ILR at specific frequency spacings.

The frequency spacings are defined in this paper in term of ERBn (Equivalent-Rectangular Bandwidth number) (ANSI, 2007). The frequency is related to the ERBn by the following formula:

$$\text{ERBn} = 21.366 \log_{10}(0.004368f + 1).$$

When two tones are presented simultaneously, different perceptual phenomena will occur depending on whether the frequency separation of the tones is smaller or greater than a certain value. This value is known as a critical band. The ERBn can be used to calculate whether the two tones are within the same critical band. If the frequency difference in ERBn is smaller than one, then the tones are in the same critical band. If the difference is greater than one, then the tones are in different critical bands.

II. EXPERIMENT I

A. Procedure

A modified Békésy tracking procedure was used to evaluate the amount of ILR as function of the frequency separation. The experiment was divided in two conditions: the baseline and the inducer condition. In the baseline condition, a test tone and a comparison tone were presented during each trial. In the inducer condition, an inducer was presented prior to the test tone and the comparison tone. The stimuli were all pure tones with equivalent rectangular durations of 200 ms (i.e., they have the same energy as a rectangular temporal envelope tone). The tones had 6.67-ms Hanning-window rises and falls and 195-ms steady-state. Tones were presented monaurally to the right ear for all listeners. The inducer was a 2500 Hz tone presented at 80 dB SPL. The test tone was presented at a fixed level of 70 dB SPL. It started 1250 ms after the inducer in the inducer condition and 1450 ms after the previous trial in the baseline condition. The frequency of the test tone varied from 800 Hz to 6.2 kHz with a 0.1 ERBn increase in frequency at each trial. The comparison tone, which started 550 ms after the test tone, was presented at a fixed frequency of 500 Hz. The level of the comparison

tone started at 72 dB SPL and varied according to the judgment made in the previous trial. The starting level was selected in order to approximately set the comparison tone to the same loudness as the test tone at the beginning of the session (ANSI, 2004). A 1600-ms silence was left between the end of the comparison tone and the beginning of the next trial, resulting in trials totaling 4 seconds in length.

The task of the listener was to judge which of the test or the comparison tone was louder, while ignoring the inducer (if presented). If the listener judged that the test tone was louder than the comparison tone, the level of the comparison was increased by 1.5 dB in the following trial. Otherwise, the comparison-tone level was decreased by 1.5 dB. In order to keep the presentation timing carefully controlled, the next trial started automatically 1600 ms after the comparison tone. If the listener did not give any answer during this time, the judgment made on the previous trial was used to set the new level of the comparison tone. The session stopped when the frequency of the test tone reached 6.2 kHz. The total number of trials was 171, which lasted 11 min and 24 sec.

The baseline condition was identical to the inducer condition, except that the inducer was attenuated to be inaudible. Each condition was repeated three times. This experiment was run on three different days. Each day, one baseline and one inducer condition were run. The baseline condition was always first in order to reduce the ILR effect on this condition. Listeners took a fifteen-minute break between the sessions to prevent undesired ILR carry-over effects (Epstein and Gifford, 2006)

B. Listeners

Ten listeners (six females and four males) participated in this experiment. They ranged in age from 20-45 years old with a mean of 27 years. All listeners had audiometric thresholds less than 20 dB HL (ANSI, 2004) at octave frequencies from 250 to 8000 Hz in both ears when measured clinically and medical histories consistent with normal hearing. Listeners were members of the laboratory or volunteers, and all of them had previous experience making equal-loudness judgments. Laboratory non-

members were paid \$10 an hour for their participation.

C. Apparatus

Listeners were seated in a sound-attenuating booth. A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli, recorded the listener's responses, and executed the adaptive procedure. The sample rate was 48 kHz. The output of the 16-bit D/A converter was attenuated, low-pass filtered, attenuated again, and led to a headphone amplifier, which fed one earphone of the Sony MDR-V6 headset.

D. Results

Individual data

Figure 1 shows the individual results for the six sessions. The three thick black lines represent the level of the comparison as function of time (or frequency of the test tone) in the baseline condition. For clarity's sake, it is worth noting that these lines indicate how the loudness of a tone with a fixed level varies with frequency. These lines are different from equal-loudness contours, which represent the level of tones with fixed loudnesses as function of frequency. The small difference between these lines indicates that the amount of intra-subject variation is quite small (*i.e.*, the listeners are generally consistent.) The average standard deviation between these lines ranges from 2 for listener L3 to 7.9 for listener L10.

Figure 1 about here

The three light lines represent the level of the comparison tone for inducer condition. Listeners also were generally consistent in the inducer condition. The average standard deviation ranges from 1.4 for listener L3 to 5.5 for listener L1. However, two main inconsistencies are apparent. Listener L1 shows an abnormal increase after 3000 Hz. In a post-experiment interview, she recognized falling asleep in the middle of this session. During this period, the lack of response was interpreted by the procedure as a continuous repetition of the last response, which led to the constant increase in level. One session of Listener L10 is highly inconsistent

with his other sessions. No clear explanation was found for this behavior.

The amount of ILR can be measured by comparing the baseline and the inducer condition. Figure 1 shows that the inducer condition has the general tendency to be lower than the baseline. The average level of the inducer condition is 3.4 dB below the average of level of the baseline condition. In order to estimate if this difference was significant, a two-way repeated-measure ANOVA was run for each listener, with the frequency ($df=170$) and conditions ($df=1$) as main factors. For every subject, the effect of frequencies was significant ($p<0.00001$). For seven subjects out of ten (L2, L3, L4, L5, L7, L8 and L9), the interaction between the frequency and the condition was highly significant ($p<0.00001$). This result shows that 1] the amount of ILR is significant; 2] it is dependent on the frequency of the test tone. For the three other subjects, L1, L6 and L10, no significant effect was found for the condition or its interactions. Therefore, they showed no significant ILR, and they were excluded from the rest of the analysis. In previous ILR experiments, it was not uncommon to find listeners that do not show any ILR (Epstein, 2007).

Figure 2 shows the amount of ILR for each listener, estimated as the average of the three inducer conditions subtracted from the average of the baseline conditions. The results were fit with an eighth-order polynomial. The shaded area represents the average plus and minus the standard deviation at each frequency. All listeners showed a similar pattern of results. The amount of ILR starts at zero because every trial started at the same level, then it increases slowly to reach a maximum between 2.5 and 3 kHz, and then it decreases to a level close to zero at the highest frequencies tested. The maximum amount of ILR ranges from 6.5 for L3 to 21 for L9. Surprisingly, the maximum ILR occurred for most of the listener between 2.7 and 3.3 kHz, and not at the frequency of the inducer (2.5 kHz).

Figure 2 about here

Average Data

The last panel in Figure 2 shows the average amount of ILR as function of frequency for the 7 listeners that showed a significant amount of ILR. The average shows three striking results. First, the frequency scope of the effect of an inducer at 2.5 kHz is much wider than a critical band. The amount of ILR is greater than 5 dB between 1782 and 5300 Hz. This bandwidth corresponds to 9.4 critical bands (ERBn). Second, the effect is asymmetrical. The amount of ILR rises from 5 dB to the maximum in 3.4 ERBn then it decreases from the maximum to 5 dB in 5 ERBn. Therefore a high frequency tone is much more likely to be affected by a lower frequency tone, than the opposite. Third, the maximum amount of ILR is 12.9 dB and it occurs when the frequency of the test tone is 3 kHz. This is surprising because this frequency is higher than that of the inducer. Many explanations could account for this last striking result, but first it is necessary to determine whether this effect resulted from a bias introduced by the methodology because the frequency of the test tone was always increasing with each presentation. In order to test this possibility, a second experiment was run.

III. EXPERIMENT II

A. Procedure

This experiment was designed to reproduce previous ILR experiments (Mapes-Riordan and Yost, 1999; Arieh and Marks, 2003), while keeping the same parameters for the inducer and the comparison tone as in the present Experiment I. Listeners were asked to adjust the level of a comparison tone to match the loudness of a fixed frequency test tone in a two-interval, two-alternative forced-choice paradigm with four interleaved tracks. In the baseline condition, only the test and the comparison tones were presented. In the inducer condition, a series of twelve inducers was presented prior to the first trial. Then, before each test tone, one inducer was presented. The time separations between the tones were the same as in Experiment I except that subsequent trials did not automatically begin. Each new trial began 110 ms after the listener's response. Each condition was divided into

four blocks that differed only by the frequency of the test tone. The same types of stimuli as in Experiment I were used in Experiment II. The inducer and the comparison tone were exactly the same, only the test tone differed. Instead of having the frequency of the test tone increase at every trial, the frequency was fixed inside a block. Four frequencies were used: 2000, 2500, 3000 and 4000 Hz, corresponding to separations of -1.9, 0, 1.6 and 4 ERBn from the 2500-Hz inducer. The level of the comparison was adjusted according to a simple up-down method, with a step size of 5 dB until the second reversal, and then 2 dB for seven reversals until the track ended. For each trial, the track was selected randomly among those that had not yet ended. Equal loudness for each track was calculated as the average of the last four reversals. Each condition was repeated three times. The average of the twelve estimations per block was calculated as the final adjusted level. The amount of ILR was calculated as the difference between the baseline and inducer condition for the block frequency. Each condition was separated by at least a fifteen-minute break to prevent undesired ILR carry-over effects (Epstein and Gifford, 2006). The apparatus was the same as used in Experiment I.

C. Listeners

Nine listeners (five females, four males) participated in the second experiment. They ranged in age from 20-30 years old with a mean of 25 years. All listeners had audiometric thresholds less than 20 dB HL (ANSI, 2004) at octave frequencies from 250 to 8000 Hz in both ears when measured clinically and medical histories were consistent with normal hearing. Listeners were members of the laboratory or volunteers, and all of them had previous experience making equal-loudness judgments. Laboratory non-members were paid \$10 an hour for their participation.

Among the nine listeners, seven have participated in the Experiment I.

D. Results

Individual data

Figure 3 shows the individual responses for each listener. Listeners that participated in Experiment I

are designated by the same number in Experiment II. Listeners that did not participate are designated by a number greater than 10. The lines represent the average adjusted level of the comparison tone as function of the frequency of the test tone. The bars represent the standard deviation of the twelve tracks. The level for the inducer condition is always lower or the same (except for the L2 at 4 kHz, L3 at 2 kHz and L11 at 2 kHz), indicating a possible presence of ILR. The standard deviation ranges from 1.2 for L9 at 2.5 kHz to 7.9 for L13 at 2 kHz in the baseline condition and from 0.9 for L2 at 2.5 kHz to 10.41 for L11 in the inducer condition. The standard deviation is on average smaller than Experiment I, which should be expected because every condition was repeated twelve times in contrast to the 3 repetitions in Experiment I.

Figure 3 about here

A set of ANOVAs, similar to those run in Experiment I was conducted. In this experiment, frequency ($df=3$, $p<0.05$) was significant for all listeners. Additionally, all listeners also showed a significant effect of the condition or the interaction between the condition and the frequency. Three listeners, L4, L7 and L13, showed only significant effect of the condition ($df=1$, $p<0.05$), a listener, L3, showed only an effect of the interaction ($df=3$, $p<0.05$) and all of the other listeners showed a significant effect for both. It is worth noting that L10, who did not show any significant ILR in Experiment I, showed significant ILR in Experiment II. The consistency of L10's data with other listeners in Experiment II and the lack of consistency of L10's data with other listeners in Experiment I, suggests that L10 may have had some difficulty properly performing the task used in Experiment I.

The amount of ILR, shown in Figure 4, was calculated by subtracting the average level of the comparison tone in the baseline condition from the level of comparison tone in the inducer condition.

The standard deviation of the amount of ILR was calculated by taking the square root of the variance average of the two conditions.

Figure 4 about here

The maximum amount of ILR ranges from 4.9 for L7 to 13 for L2, consistent with Experiment I. However, this time, for six listeners out of nine (L2, L3, L9, L10, L12 and L13), the maximum occurs when the frequency of the test tone is the same as the frequency of the inducer. For two listeners, L4 and L7, the amount of ILR is about the same for 2.5- and 3-kHz test tones. For one listener, L11, the maximum amount of ILR occurs at 3 kHz

A T-test was performed on each condition and each listener to evaluate if the effect of ILR was significant. Results are represented in Figure 4. When the test showed no significant effect ($p > 0.01$) the amount of ILR is represented by an open symbol, and when it is significant by a filled symbol. ILR is significant for every listener at 2.5 kHz, for 7 listeners at 3 kHz, for 6 at 4 kHz and only for 3 at 2 kHz.

Average Data

The last panel in Figure 4 shows the average amount of ILR as function of frequency. Averaged results show a maximum of ILR at 2.5 kHz with an amount of reduction of 7.8 dB. The amount is within the range found in the literature. A two-way ANOVA was run for every frequency separately, with the condition ($df=1$) and the number of repetition ($df=11$) as main factors. As showed in the figure by an open symbol the effect of ILR was not significant at 2 kHz ($p > 0.01$), but was significant at 2.5 kHz ($p < 0.0001$), 3 kHz ($p = 0.0009$) and 4 kHz ($p = 0.0049$). This result shows that in average an inducer at 2.5 kHz has no significant effect of tone at frequency at least below 2 kHz (1.9 ERBn), but can have an effect on a tone with a frequency at least below 4 kHz (4 ERBn). This result confirms two of the results of Experiment I: the effect is asymmetrical, and is much wider than a critical

band. But it repeats the third conclusion: maximum amount of ILR does not appear at a frequency higher than the inducer.

E. Discussion

The results of both experiments demonstrate that ILR is a more pervasive effect, relative to frequency, than previously believed. Specifically, an inducer may cause ILR on another tone that is at least 4 ERBn above in frequency. The pattern shows a wider spread of ILR toward higher frequencies than lower frequencies. This spread is similar to the spread of excitation on the basilar membrane (Zwicker and Fastl, 1990; Moore and Glasberg, 1987). Figure 5 shows an excitation pattern according to the ANSI standard loudness model (ANSI, 2007) for a 2.5 kHz tone at 80 dB SPL. This similarity suggests that the amount of ILR is dependent on the amount of excitation on the basilar membrane. However, because of the contralateral effects of ILR, it has been argued that at least some of the effects of ILR are attributable to a central mechanism (Nieder *et al.*, 2007). Therefore, more data are needed to determine whether this correlation is a result of the same physiological mechanism.

The two experiments show some dissimilarity. In Experiment I, when the tracking procedure was used, the maximum amount of ILR occurred when the test tone frequency was higher than that of the inducer. As this finding was not replicated in Experiment II, it is likely that this finding was due to some shift in the ILR pattern that occurred as a result of the sweep rate of the test tone. It is clear that some delay between the subject's response and the compensatory level adjustment should be seen on the frequency scale, however, it is not clear to what degree the present results reflect only this factor. It would be necessary to study the effects of sweep rate on the ILR pattern in order to better understand this relationship. In addition, the frequency width of the effect of ILR is greater in Experiment I than Experiment II. Again, this may simply result from the data delay caused by the sweep or some other inherent difference between

the procedures, similar to results seen in sweeping auditory-threshold measurements (Miskiewicz *et al.*, 1994).

The finding that ILR is more pervasive across frequency than previously believed provides another guideline for the careful design of experiments in order to prevent undesired ILR effects.

Additionally, it may help explain some of the variability seen in loudness-matching and masking experiments that include tones that are relatively close in frequency.

These results suggest that a number of experiments, including, for example, those commonly used to measure equal-loudness contours may have been affected by ILR. Most of the experiments deriving contours used loudness-matching paradigms that required listeners to match the loudnesses of tones at different frequencies to a reference tone at 1 kHz (see Suzuki and Takeshima, 2004 for review). The results from the present experiment suggest that the reference tone may have reduced the loudness of matched tones nearby in frequency. Therefore, the shape of the equal-loudness contours might depend on the frequency of the reference tone.

ACKNOWLEDGMENTS

The authors thank Mary Florentine and Bertram Scharf for their support and helpful comments, and Simona Ioffe for collecting the data. This research was supported by NIH/NIDCD Grants No. R01DC02241 and the Swiss National Science Foundation.

REFERENCES

ANSI (2004). "American National Standard Specification for Audiometers," *Journal ANSI S3.6-2004*.

ANSI (2007). "American National Standard Procedure for the Computation of Loudness of Steady Sounds," *Journal ANSI S3.4-2007*.

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., and Marks, L.E. (2001). "Recalibration of loudness: sensory vs. decisional processes," in *Fechner Day 2001* edited by Sommerfeld, E., Kompass, R. and Lachmann, T. (Pabst, Berlin).

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

Braida, L.D., and Durlach, N.I. (1972). "Intensity perception. II. resolution in one-interval paradigms," *J. Acoust. Soc. Am.* **51**, 483-502.

Buus, S., Florentine, M., Poulsen, T., and Müsch, H. (1999). "On the relation between peripheral compression and the growth of loudness," *J. Acoust. Soc. Am.* **105**, 1022.

Epstein, M. (2007). "An introduction to induced loudness reduction," *J. Acoust. Soc. Am.* **122**, EL74-80.

Epstein, M., and Florentine, M. (2005). "A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions," *J. Acoust. Soc. Am.* **118**, 907 - 913.

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.

Florentine, M., and Epstein, M. (2006). "To honor Stevens and repeal his law (for the auditory system)," in *Fechner Day* edited by D. E. Kornbrot, R. M. Msetfi and A. W. MacRae (International Society for Psychophysics, St. Albans, England).

Hellman, R.P., and Zwislocki, J.J. (1963).

"Monaural loudness function at 1000 cps and interaural summation," *J. Acoust. Soc. Am.* **35**, 856-865.

Mapes-Riordan, D., and Yost, W.A. (1997).

"Loudness recalibration as a function of recalibration and comparison tone level," *J. Acoust. Soc. Am.* **101**, 3170.

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. (1992). "The slippery context effect in psychophysics: Intensive, extensive, and qualitative continua," *Percept. Psychophys.* **51**, 187-198.

- Marks, L.E. (1994). "'Recalibrating' the auditory system: the perception of loudness," *J Exp Psychol Hum Percept Perform* **20**, 382-396.
- Marks, L.E., and Warner, E. (1991). "Slippery context effect and critical bands," *J Exp Psychol Hum Percept Perform* **17**, 986-996.
- Miskiewicz, A., Buus, S., and Florentine, M. (1994). "Auditory facilitation: Procedural or sensory effect?," *J. Acoust. Soc. Am.* **96**, 1429-1434.
- Moore, B.C.J., and Glasberg, B.R. (1987). "Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns," *Hear. Res.* **28**, 209-225.
- Nieder, B., Buus, S., Cazals, Y., and Scharf, B. (2007). "Loudness reduction induced by a contralateral tone," *J. Acoust. Soc. Am.* **122**, 35-37.
- 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.
- Parker, S., and Schneider, B. (1994). "The stimulus range effect: Evidence for top-down control of sensory intensity in audition," *Percept. Psychophys.* **56**, 1-11.
- Poulton, E.C. (1989). *Bias in Quantifying Judgments*, (Erlbaum, Hillsdale, NJ).
- Schneider, B., and Parker, S. (1990). "Does stimulus context affect loudness or only loudness judgments," *Percept. Psychophys.* **48**, 409-418.
- 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.
- Ward, L.M., Armstrong, J., and Golestani, N. (1996). "Intensity resolution and subjective magnitude in psychophysical scaling," *Percept Psychophys* **58**, 793-801.
- Zwicker, E., and Fastl, H. (1990). *Psychoacoustics - Facts and Models*, (Springer-Verlag, Berlin).

FIGURE CAPTIONS

- Figure 1 – The level of the comparison tone as a function of the frequency/time matched to the loudness of a 2500-Hz test tone at 60 dB SPL in each of the three repetitions of the baseline (dark lines) and the inducer (light lines) conditions for 10 listeners in Experiment I. The error bars show the standard deviation.
- Figure 2 The amount of ILR caused by a 2500-Hz inducer as a function of frequency for each of the seven listeners who exhibit some amount of ILR in Experiment I and the average of those listeners. The shaded area represents the average plus and minus the standard deviation (individual listeners) or plus and minus one standard error of the mean (average) at each frequency.
- Figure 3 – The level of the comparison tone as a function of the frequency matched to the loudness of a 2500-Hz test tone at 60 dB SPL in the baseline (dark lines) and the inducer (light lines) conditions for 9 listeners in Experiment II. The error bars show the standard deviation.
- Figure 4 - The amount of ILR caused by a 2500-Hz inducer as a function of frequency for each of the nine listeners in Experiment II and the average of those listeners. The error bars show the standard deviation (individual listeners) or the standard error (average). The filled symbols represent a significant amount of ILR according to a T-test (individual listeners) ($p < 0.01$). or ANOVA (average) ($p < 0.01$).
- Figure 5 – Excitation pattern of a 2.5 kHz tone presented through headphones at 80 dB SPL according to the ANSI standard loudness model (ANSI, 2007).

Figure 1

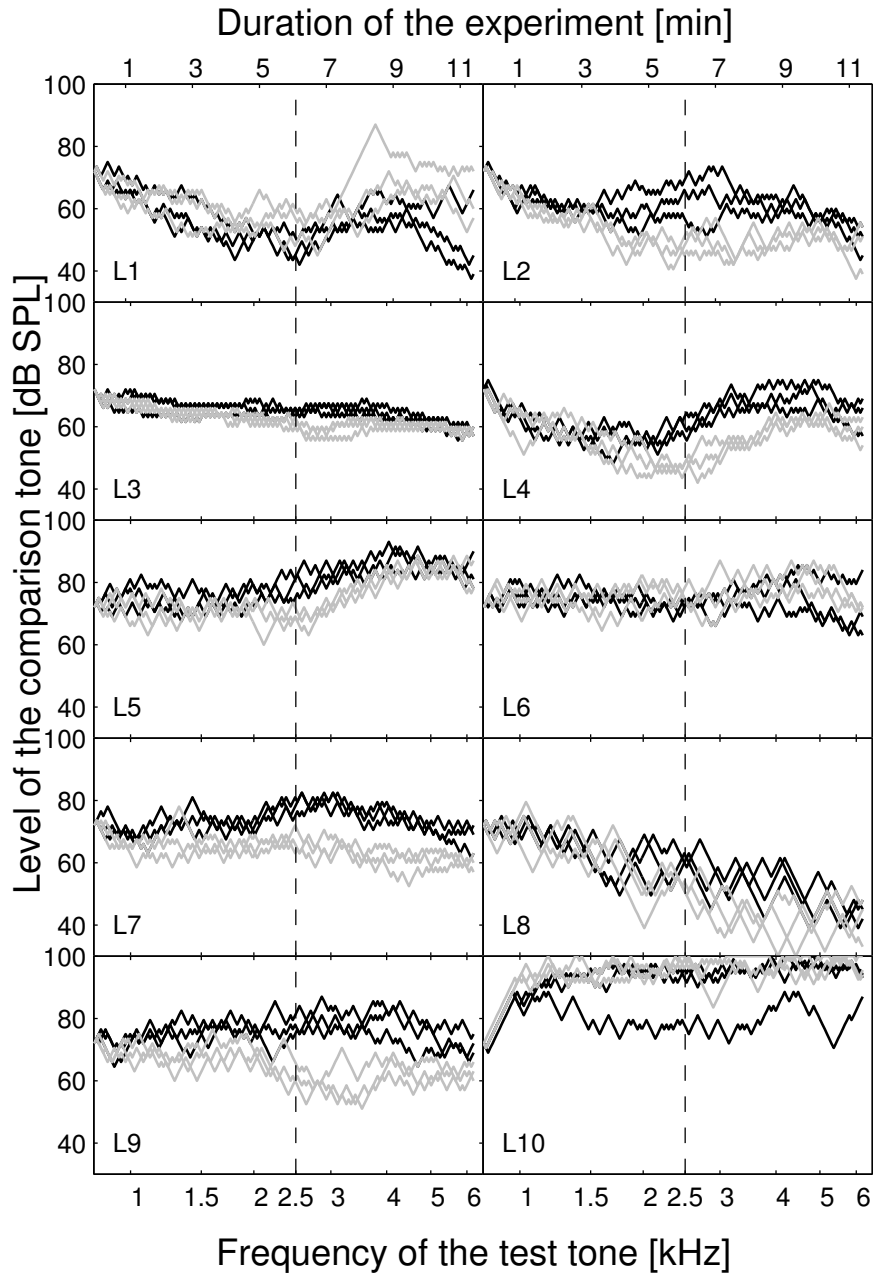


Figure 2

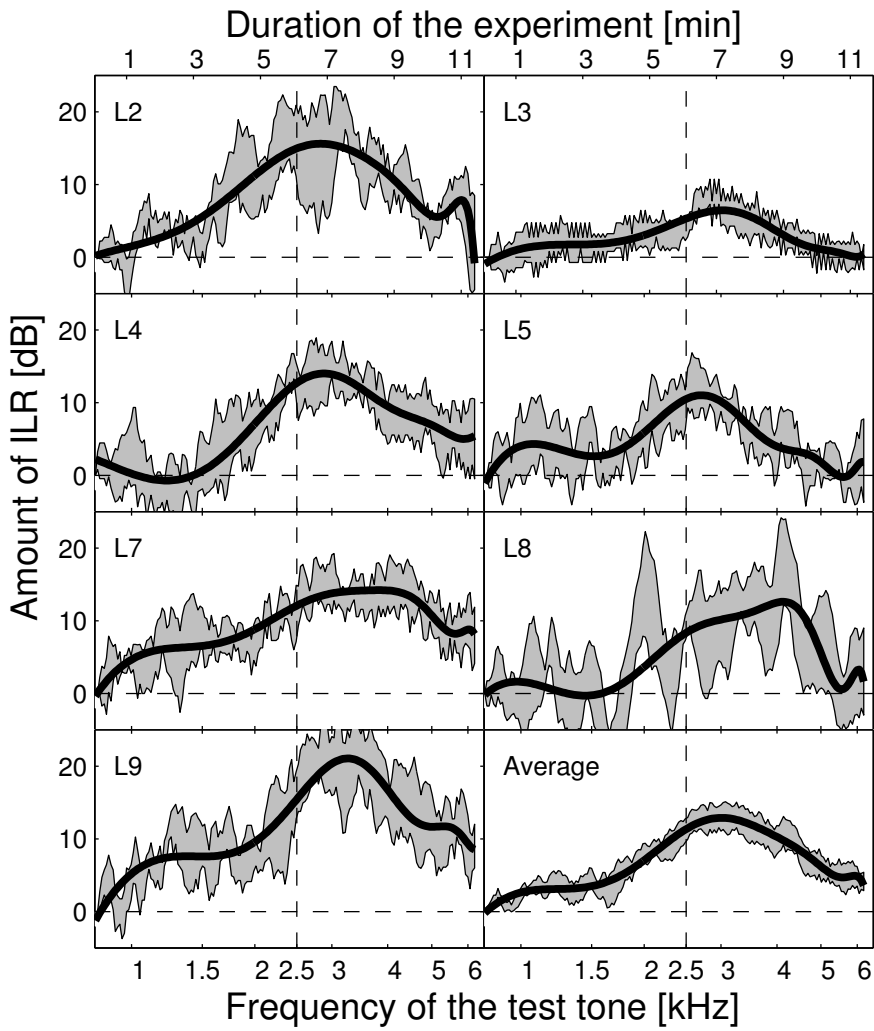


Figure 3

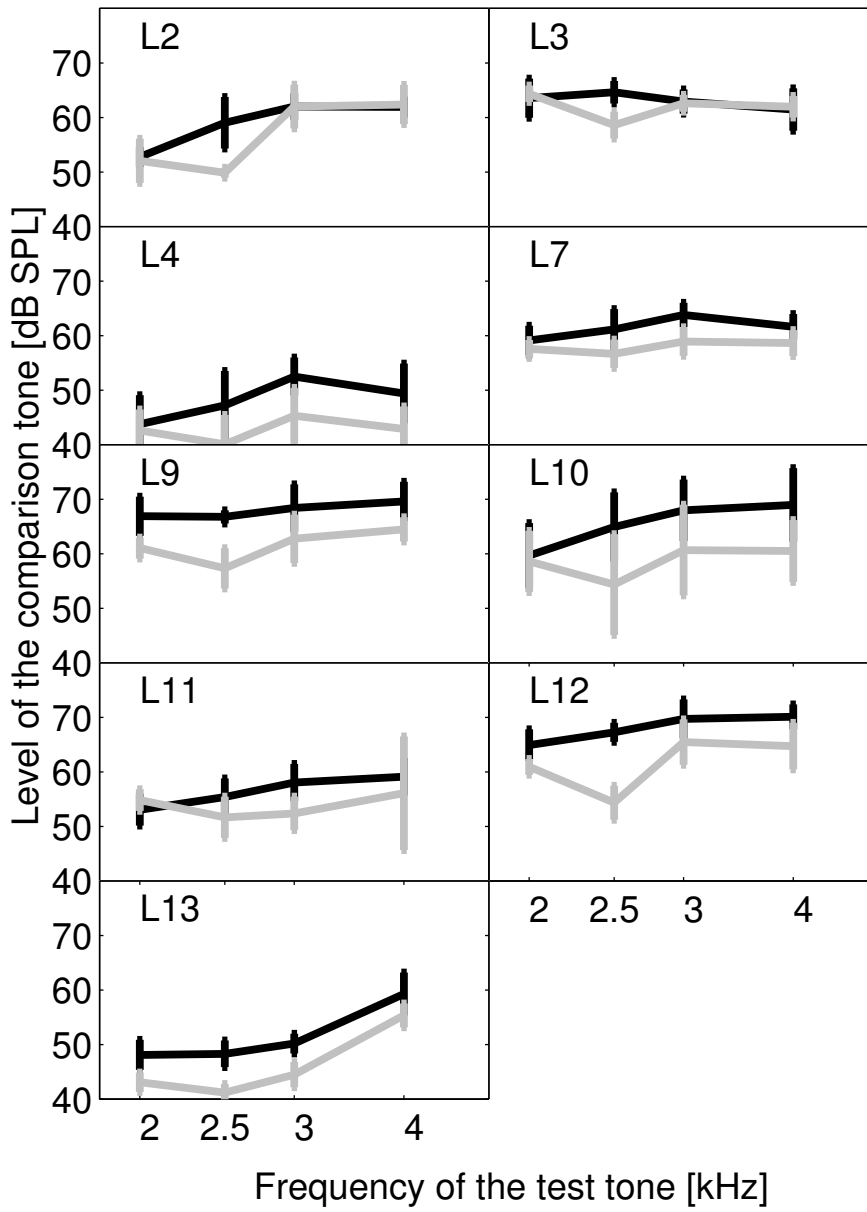


Figure 4

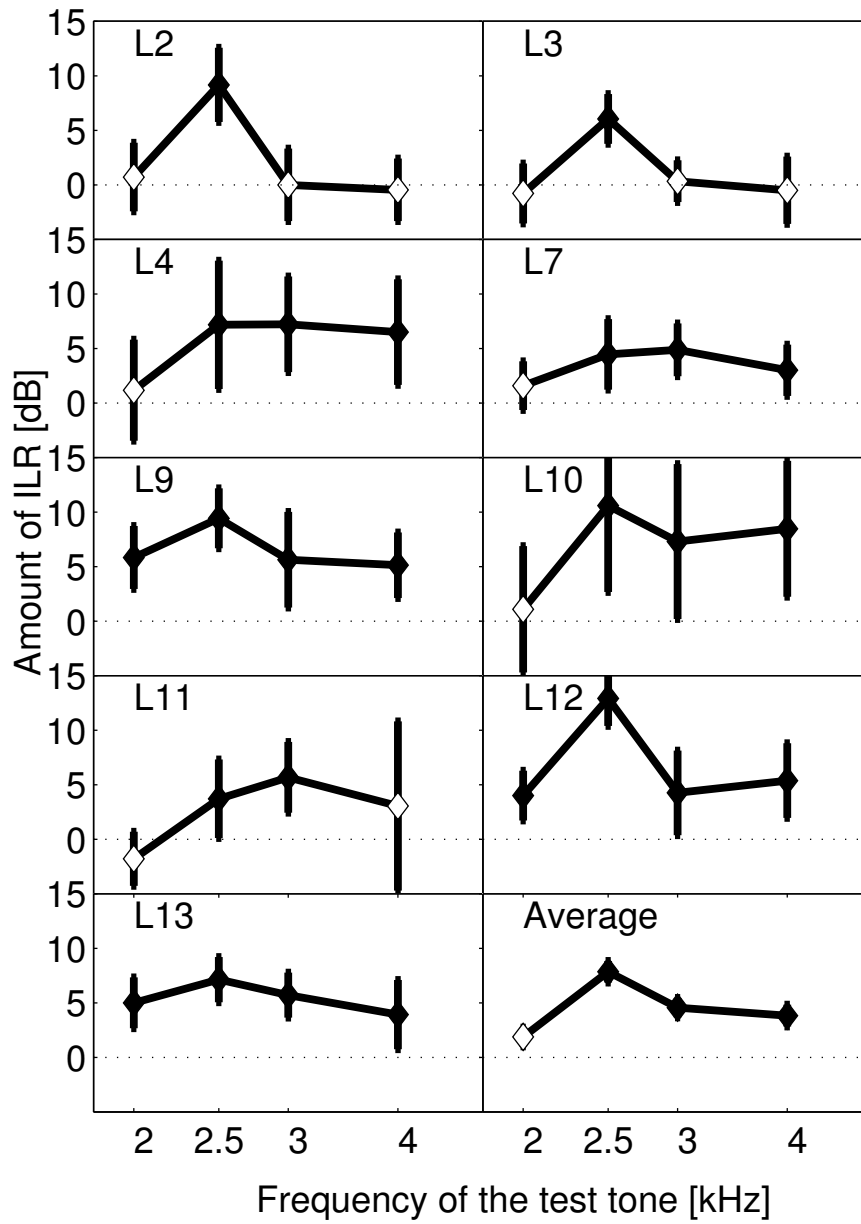


Figure 5

