# THE EFFECT OF INDUCED LOUDNESS REDUCTION ON LOUDNESS MATCHING: THE ADJUSTMENT ERROR

Bertram Scharf
*Department of Psychology, Northeastern University, Boston, MA 02115*
*scharf@neu.edu*

## Abstract

*In the matching of loudness by the method of adjustment, one sound—the variable—is varied in intensity and the other—the standard—is fixed. Listeners usually judge a sound as softer when it is the variable than when it is the standard. They set sound A to a higher level to match sound B when A is the variable than when B is the variable. This difference or adjustment error may be as large as 30 dB but is usually under 5 dB. A large part if not all of the error appears to result from induced loudness reduction, or ILR. ILR is the loudness decline imposed by a stronger tone on a weaker one that follows within a few seconds. Its magnitude and temporal characteristics, including formation and disappearance, are compatible with the characteristics of the experimental conditions that give rise to the adjustment error.*

Towards the end of his notable career as psychophysicist, experimental psychologist, philosopher, S. S. Stevens wrote that "the matching operation becomes the key" to measurement (Stevens, 1975, p. 46). In particular, all psychophysical measurements are matches, between numbers and sensations, between sensations from different modalities, between perceived magnitudes of different stimuli on the same modality. In every case, questions arise as to just how well the matches represent the "true" underlying experiences. Do the ratios among assigned numbers in magnitude estimation reflect the relations between the sensations that the numbers represent? Is it possible to equate reliably sensation magnitudes from different modalities? Likewise, within a given modality for a given stimulus with its many perceptual attributes can one be isolated from all the others to serve as the basis for a meaningful match? Positive answers to these questions come primarily from converging operations, which show that different procedures lead to congruent answers. For example, magnitude estimation and cross-modality matching yield similar approximations of the relation between sensation magnitude and stimulus magnitude.

Sensory matching within a modality is usually concerned with determining how various stimulus variables influence the relative positions of stimuli along some perceptual dimension. In hearing, loudness matching has been the subject of an enormous literature to determine how loudness depends on the frequency of tones, on the bandwidth of complex sounds, on duration, on the presence of other sounds, and on how these stimulus variables interact with each other and with intensity. The classical psychophysical methods of adjustment, limits, constant stimuli, and assorted variations have and are still used to determine what it takes to make two sounds equal in loudness. Currently, adaptive procedures take advantage of the rapidity and power of computers to optimize the measurements, combining the best of the classical methods to make them more reliable and more precise. [However, as one recent report indicated, "excessive measurement time" and listener dissatisfaction may make the method of adjustment a better choice (Mauermann, Long, and Kollmeier, 2004).] The present paper focuses on the method of adjustment and a seeming error that afflicts it. My analysis, which should help us to understand some of the anomalies and discrepancies among loudness measurements based on the method of adjustment, is also relevant to the popular adaptive procedures.

I first define the *adjustment error*, a term suggested by Stevens and Guirao (1967). I then review induced loudness reduction or ILR, which is likely to be the principle source of the error.

## Adjustment Error

In the method of adjustment, one stimulus—the *variable*—is varied in stimulus magnitude and the other—the *standard*—is fixed. The observer's task is to equate the two stimuli along some specified perceptual dimension. If the dimension is loudness and the stimulus variable is intensity, listeners usually judge a sound as softer when it is the variable than when it is the standard. They set sound A to a higher level to match sound B when A is the variable than when B is the variable. The usual way to deal with this discrepancy, which is the adjustment error, is to have both sounds serve in turn as standard and as variable and then to average the results. As shown later, this solution may be suitable for the matching of one tone with another at a different frequency or with a different duration, but not for the matching of tones with complex tones or noise. Moreover, the adjustment error can be very large, especially when the two sounds being matched are very different in frequency or quality. For example, Hellman (2001) reported an error of nearly 30-dB error for a loudness match between 5-kHz and 16-kHz tones, and Zwicker (1958) reported 12 dB for a match between a 1-kHz tone and broadband noise.

## Induced Loudness Reduction

As to induced loudness reduction, it takes place when a stronger tone is followed by a similar but weaker tone at least 500 ms later. (We do not know whether sounds, other than tones, such as multi-tone complexes and noise are also subject to ILR.) After 1 or 2 min of repeated exposure, the loudness of the weaker tone may be as much as halved (for a recent review, see Wagner and Scharf, 2006). Nieder, Buus, and Scharf (submitted), on the basis of new measurements and some in the literature (e.g. Nieder et al., 2003), concluded that the amount of ILR increases with SPL up to 70 or 80 dB above which it increases no further and may even decrease. This maximum amount of ILR is approximately 10 dB but varies greatly among listeners. With respect to the dependence of ILR on time, several reports have shown that it develops rapidly as stronger and weaker tones at the same frequency are repeated in succession over a minute or so, after which it appears to approach asymptote (Wagner and Scharf, 2006). Recovery from ILR, i.e. the return of the loudness of the weaker tone to its value before the introduction of the stronger tone, requires at least 2 min (Arieh, Kelly, and Marks, 2005; Wagner and Scharf, 2006; Epstein and Gifford, in press).

In their study, Wagner and Scharf (2006) presented an inducer tone at 80 dB SPL followed by a weaker test tone at 70 dB. The 200-ms tones were separated by 1800 ms. By the method of successive magnitude estimation, 12 listeners assigned a number to represent the loudness of the weaker tone every 5 s. Figure 1, adapted with permission from Wagner and Scharf (2006), illustrates the development of and recovery from ILR for a 500-Hz tone. The geometric mean estimation is plotted as a function of the time from the first trial. The initial estimation with no inducer averaged 4.3. A few seconds later, the first estimation with a preceding inducer was down to 3.5. After 3 min the mean estimation was 2.6, a decline in loudness of 40% which appears to be close to asymptote. Recovery after 1 min without an inducer was far from complete. Results were similar at other frequencies up to 8000 Hz, the highest tested.

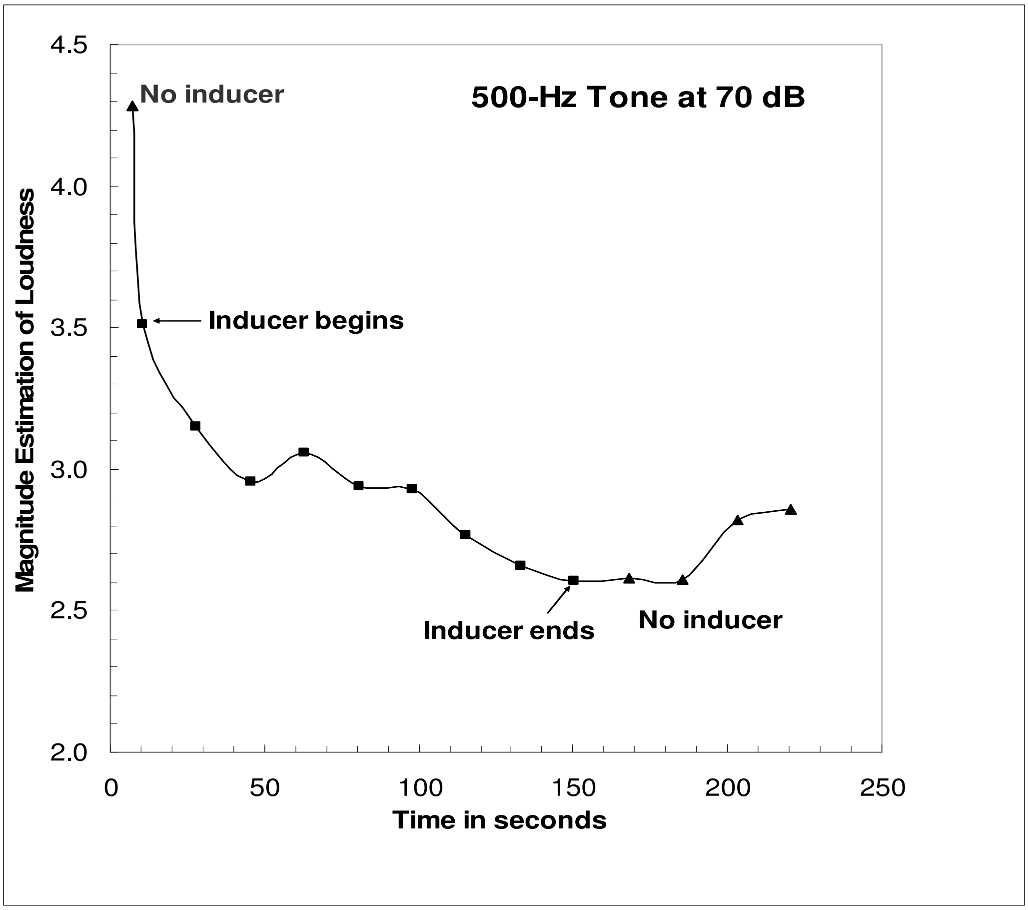

Fig. 1. Magnitude estimation as a function of time from the first exposure to the 70-dB tone.

## Induced Loudness Reduction and the Adjustment Error

These results are directly relevant to the adjustment error because they show that ILR has time to develop within the usual period required to complete a loudness match by the method of adjustment. During this time, the sound that is varied will be set alternately to higher and lower levels. Set at higher levels, the variable acts to reduce the loudness of its weaker self on the next and subsequent trials. Whenever the variable is stronger on one trial, it will contribute to ILR on weaker tones in later trials. Hence during the course of a single match, the level of the variable will be set higher and higher to compensate for its reduced loudness. A kind of vicious cycle is set in motion. But why don’t stronger exemplars of the variable also reduce the loudness of the standard sound? They do not, provided the tones being matched are very different in frequency. For ILR to take place, the stronger tone must be in the same critical band as the weaker tone (e.g. Marks and Warner, 1991). For this reason, as noted above, the adjustment error is greatest when the sounds to be matched in loudness are very different in frequency (and/or quality). In the case where the two tones are close in frequency, whenever the variable tone is made stronger than the standard tone, it will reduce the loudness of both the standard and a subsequent variable tone that happens to be weaker. As a result both the standard and variable will decline in loudness during the course of a loudness match, thereby minimizing the adjustment error. A similar analysis applies to loudness matches based on a 2AFC adaptive procedure to which I return later.

A striking example of the adjustment error is provided by Zwicker (1958) whose 20 listeners used a tracking procedure to match the loudness of a 1000-Hz tone to that of a broadband noise over a 110-dB range of sound pressure levels. In tracking loudness, the listener controls the direction of a continuous change in the level of the variable sound. The instruction was to make the variable clearly louder and clearly softer than the standard, reducing the difference until the two sounds were approximately equal in loudness. Both the tone and noise served as the standard, probably in a mixed order. Table I shows at each SPL of the tone the difference between the levels of the tone and noise required for equal loudness. At all but the highest level, the tone had to be more intense to equal the noise in loudness. However, the difference was nearly always greater when the tone was varied than when the noise was varied. This difference is the adjustment error. The error is greatest at moderate SPLs. Guirao and Stevens (1967) reported similar matches at four levels from 30 to 90 dB. They used the classical method of adjustment with the listener controlling sound intensity by means of a sone potentiometer. Although they reported smaller differences between tone and noise and smaller adjustment errors than did Zwicker (1958), the magnitudes tended to vary with sound level in similar fashion. The decrease at lower levels mimics ILR which also decreases as level decreases below about 60 dB (Nieder et al., submitted). The decrease in the in the adjustment error at the higher levels is not like ILR which seems to remain constant above 80 dB or so. Two factors could result in a smaller error at higher than at moderate levels. First, as Stevens (1956) pointed out, at very high levels listeners are likely to avoid setting the variable still higher which would make the sound even more unpleasant. Second, since the matches were made in the same session, sounds at moderate levels would have been preceded often by matches at higher levels. Matches at the highest level would never have been preceded by matches at still higher levels. Given the long time required for recovery from ILR (e.g. Epstein and Gifford, in press), some residual ILR would affect matches at moderate levels. Although both standard and variable ought to be affected by previous matches at higher levels, the effect could be greater on the variable because the standard

| | **SPL of tone minus SPL of noise** | | |
|---|---|---|---|
| **SPL Tone** | **Noise varied** | **Tone varied** | **Adjustment Error** |
| 20 | 0 | 6 | **6** |
| 30 | 3 | 5 | **2** |
| 40 | 8 | --- | **---** |
| 50 | 10 | 23 | **13** |
| 60 | 11 | 23 | **12** |
| 70 | 15 | 21 | **6** |
| 80 | 15 | 20 | **5** |
| 90 | 12 | 18 | **6** |
| 100 | 13 | 17 | **4** |
| 110 | 16 | 16 | **0** |

*Table I. Adjustment error as a function of the SPL of the tone. The tone levels are exact for the case when it was the standard and are approximate when it was the variable. For example, when the tone was the standard fixed at 60 dB SPL, it was judged equal in loudness to a noise adjusted on average to 49 dB. When the noise was the standard at 40 dB, the tone was adjusted to 63 dB for equal loudness. Thus at around the same loudness level, the difference required for equal loudness was 11 dB in one case and 23 dB in the other yielding an adjustment error of 12 dB. (All values in the table are in dB and are approximated from data points in Fig. 7 of Zwicker, 1958.)*

would have more time to recover during the course of a match. More important for matches between tone and noise, it may be that noise undergoes less ILR than tones (cf. Nieder et al., submitted).

A number of papers report loudness matches between multi-tone complexes and a pure tone (e.g. Zwicker, Flottorp, and Stevens, 1957; Scharf, 1962). The adjustment errors are mostly under 3 dB, much smaller than Zwicker (1958) reported for a tone and broadband noise, but like his, the errors are greatest at moderate levels and decrease at the highest levels. In the one study (Scharf, 1976) in which one of the components of a three-tone complex had the same frequency as the comparison tone, the error was unusually small—averaging less than 1 dB—and did not vary with level. Perhaps the common component resulted in ILR between the comparison tone and the complex so it did not matter which one was varied.

A study by Scharf (1969) gives a clear picture of how the frequency difference between standard and variable affects the size of the adjustment error. A binaural tone, same frequency to both ears, was matched in loudness to a dichotic tone, different frequency to each ear. As the frequency separation between the ears increased so did the size of the adjustment error, going from about 1 to 2 dB for frequency separations smaller than a critical band to over 5 dB at wider separations. These results are in accordance with the dependence of ILR on frequency separation between a stronger and weaker tone; once the separation exceeds roughly the critical band, the amount of ILR begins to decrease (Marks and Warner, 1991). Moreover, once again the adjustment error was greatest at moderate levels.

Data for pure-tone matching, in which both tones were adjusted, are harder to come by. Hellman (2001) did report such data for many pairs of frequencies between 1 and 16 kHz. The graphed data for seven pairs show that the loudness of the member of the pair that is adjusted is nearly always underestimated. For the most part, the adjustment error with these pure tones decreased at the highest levels just as for complex tones and broadband noise, but the decrease at the lowest levels was not as evident. The error tended to become greater as the frequency separation between the matched tones increased. With the two tones at 3.15 and 5 kHz, the smallest separation graphed, the error was at most 2 dB; with the tones at 3.15 and 16 kHz, the largest separation, the error exceeded 10 dB and even reached 30 dB.

Underestimation of the loudness of a variable sound is not restricted to procedures that have the listener in control. Adaptive 2AFC procedures also involve changes in the level of one sound while a standard sound remains constant. Thus, the variable sound ought to produce ILR on itself with a consequent underestimation of its “normal” loudness. However, usually the two sounds serve both as standard and variable within the same series so that, if both sounds are subject to similar ILR, they would be reduced about equally in loudness with no adjustment error apparent. Nonetheless, Nieder et al. (2003) determined that when two sounds differ only in duration, the shorter one undergoes ILR from a longer sound but not vs. versa. On the basis of this finding, they could explain why Buus, Florentine, and Poulsen (1999) found a greater difference between short and long tones when the adaptive procedure was run in a series that included many more intense sounds than when run with weaker sounds. Even though the two tones were at the same frequency, the asymmetrical ILR resulted in a clear adjustment error (Buus et al., 1999, Fig. 1). Other indirect evidence for a role for ILR in adaptive procedures is provided by Nieder et al. (submitted), but the analysis is too long to be repeated here. Direct evidence seems not to be available.

## Conclusions

Induced loudness reduction appears to play an important role in the adjustment error in loudness matching. Comparable reductions in perceptual magnitude may play in role in modalities other than hearing because context effects reminiscent of ILR have been documented by Marks and Arieh (in press) in a number of other sensory modalities.

## Acknowledgments

I thank Rhona Hellman for helpful suggestions and John Suciu for help with the graph.

## References

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

Buus, S., Florentine, M., and Poulsen, T. (1997). Temporal integration of loudness, loudness discrimination, and the form of the loudness function. *J. Acoust. Soc. Am. 101*, 669-680.

Epstein, M. and Gifford, E. (in press). Measuring induced loudness reduction. *J. Acoust. Soc. Am.*

Hellman, R. Takeshima, H., Suzuki, Y., Ozawa, K., and Sone, T. (2001). Determination of equal-loudness relations at high frequencies. in *Fechner Day 2001*, edited by E. Sommerfeld, R. Kompass and T. Lachmann (Pabst, Berlin), pp. 225-230.

Marks, L. E. and Arieh, Y. (in press). Differential effects of stimulus context in sensory processing. *European Rev. of Applied Psychol.*

Marks, L. E., and Warner, E. (1991). Slippery context effect and critical bands. *J. Exp. Psychol. Hum. Percept. Perform. 17*, 986-996.

Mauermann, M., Long, G. R., and Kollmeier, B. (2004). Fine structure of hearing threshold and loudness perception. *J. Acoust. Soc. Am. 116*, 1066-1080.

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.

Nieder, B., Buus, S., and Scharf, B. (submitted). Dependence of induced loudness reduction on level. *J. Acoust. Soc. Am.*

Scharf, B. Loudness of complex sounds as a function of the number of components. *J. Acoust. Soc. Amer* ., 1959, *31*, 783-785.

Scharf, B. Loudness summation and spectrum shape. *J. Acoust. Soc. Amer.* 1962, *34*, 228-233.

Scharf, B. Dichotic summation of loudness. *J. Acoust. Soc. Amer.*, 1969, *45*, 1193-1205.

Scharf. B. Acoustic reflex, loudness summation, and the critical band. *J. Acoust. Soc. Am.* 1976, *60*, 753-755.

Stevens, S. S. (1956). Calculation of the loudness of complex noise. *J. Acoust. Soc. Am. 28*, 807-832.

Stevens, S. S. (1975). *Psychophysics*. John Wiley and Sons, New York.

Stevens, S. S. and Guirao, M. (1967). Loudness functions under inhibition. *Perception and Psychophysics 2*, 459-465.

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.

Zwicker, E. (1958). Ueber psychologische und methodische Grundlagen der Lautheit (About psychological and methodical principles of loudness), *Acustica 8*, Beiheft 1, 237-258.

Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). Critical bandwidth in loudness summation. *J. Acoust. Soc. Am. 57*, 548–557.