Loudness Scaling

Thomas Brand

Medical Physics, University of Oldenburg

Introduction

Categorical loudness scaling determines the loudness in the whole auditory dynamic range in terms of categories like 'inaudible', 'very soft', 'soft', 'medium' etc. as a function of the stimulus level. This makes categorical loudness scaling interesting for the diagnosis of recruitment (pathological reduction of the auditory dynamic range) and for the fitting of hearing aids with dynamic compression. An overview can be found in Kollmeier (1997).

Categorical loudness scaling was invented by Heller (1985) who used a two step scaling procedure. In the first step, the subject scaled the loudness of the stimulus roughly using a verbal scale with five categories. Afterwards, the stimulus was presented again at the same level and the subject used a subscale with 10 fine subdivisions. This very precise procedure can still be regarded as the 'Gold Standard' of loudness scaling, especially for research purposes. However, other less time consuming procedures were used in audiology as well. A simplified version of Heller's procedure was invented by Hellbrück and Moser (1985). Their procedure used one rating step and a scale with more than 50 response alternatives, consisting of five rough verbal categories and 10 subdivisions per category - all displayed on one scale. This procedure became very popular in Germany under the name 'Würzburger Hörfeld'. Allen et al. (1990) presented the 'Loudness Growth in ½-Octave Bands' (LGOB) procedure which used a much simpler scale consisting of seven response alternatives. This procedure also became very popular and was also used in commercial hearing aid fitting. An alternative procedure using 11 response alternatives was presented by Hohmann and Kollmeier (1995). The popularity of categorical loudness scaling strongly decreased when Elberling (1999) formulated fundamental criticism in categorical loudness scaling procedures and concluded that they should not be used at all. Elberling's critical arguments are:

1. Different methods produce different loudness functions that can not be compared.

2. Accuracy of categorical loudness scaling is bad.

3. Normal-hearing listeners show large differences in loudness functions.

4. Input/output functions of hearing aids based on loudness scaling are unclear.

5. Most hearing-impaired listeners can be fitted using a simple fine tuning.

Brand and Hohmann (2002) presented an adaptive procedure for categorical loudness scaling called 'ACALOS'. This procedure tried to minimize the measuring time as well as bias effects using an optimized model loudness function, fitting procedure, and stimulus placement. The development of this procedure is described below. In 2006 the new standard ISO 16832 (2006) 'Acoustics – Loudness scaling by means of categories" was released (see also Kinkel, 2007, in this volume). This standard sets conditions for reliable measurements in order to reduce differences between results of different categorical loudness scaling procedures and refers to the ACALOS procedure as reference procedure.

This article discusses Elberling's arguments and tries to show that some of them are not valid if an adequate procedure is used and how the remaining problems might be dealt with.

Adaptive Categorical Loudness Scaling (ACALOS)

The adaptive categorical loudness scaling procedure (ACALOS) (Brand, 2000; Brand and Hohmann, 2002) was developed in three steps which are described briefly below. In the first step, a statistical model was derived from loudness scaling data. This model served for Monte-Carlo simulations of the measurement process. In the second step, an adequate model loudness function was derived. In the third step, an adaptive procedure was developed, optimized using Monte-Carlo simulations, and evaluated using measurements with human listeners.

Statistical model

Categorical loudness scaling procedures differ in many aspects, such as presentation levels, response scales, and model loudness functions. Different procedures are therefore difficult to compare with respect to accuracy, validity, and efficiency. In order to predict these properties for arbitrary categorical loudness scaling procedures, a statistical model was introduced and evaluated (Brand, 2000). The model estimates the accuracy of loudness functions and is based on the reproducibility of single categorical loudness ratings, which was empirically derived from data of normal-hearing and hearing-impaired listeners. An `optimistic' and a `pessimistic' estimate of the response characteristic were derived that should account for the differences between listeners. Repeated measurements with 8 normal-hearing and 8 hearing-impaired listeners were in good consistence with the model predictions. The measured accuracy was slightly worse than predicted by the optimistic estimate but much better than predicted by the pessimistic estimate. The model was used to calculate the influence of fitting procedure, track length, and number of response alternatives on loudness function estimates. If the outlier rate is 7 % or higher, a robust fitting procedure, e.g. a maximum likelihood fit with Lorentzian merit function is recommended rather than a least-squares fit. Minimum numbers of about 10 response alternatives and about 15 trials are needed to yield suitable loudness function estimates for practical purposes.

Model function

The second step in the development of the procedure was the selection of an adequate model loudness function. The number of the free parameters of the model function should be as low as possible in order to allow stable fitting even to a relative small number of loudness ratings. On the other hand, the number of free parameters has to be large enough in order to enable the loudness function to be fitted to the different shapes that occur in different listeners. The criteria that were investigated were the bias and the reproducibility (intra-individual standard deviation) of the fit. The bias was defined as the mean difference between the single measurement loudness functions fitted to single loudness scaling tracks (i.e. approximately 15 loudness ratings for one stimulus at different levels) and the individual reference loudness function of the specific listener and stimulus. The individual reference loudness functions were derived by calculating the median response levels for the different loudness categories across all measurements with a specific listener and stimulus. Repeated loudness scaling measurements were performed with 10 normal-hearing and 10 hearing-impaired listeners. Both the constant stimuli version (Hohmann and Kollmeier, 1995) and the adaptive version (Brand and Hohmann, 2002; see below) were applied. The following model functions were systematically fitted to these data: A linear function according to Fechner's law, a linear function with an offset for low levels (Hohmann and Kollmeier, 1995), two linear functions with different slope values, connected in the level related to the loudness (‘medium’) (‘broken stick’), ‘broken stick” with smoothing of the transition area, a modified Fechner function according to Nowak (Heller *et al.*, 1997), a further modified Fechner function, a function presented by Nowak (1990), an exponential function according to Stevens (1956), and polynomials of second and third order. Examples of these model functions can be seen in Fig. 1.

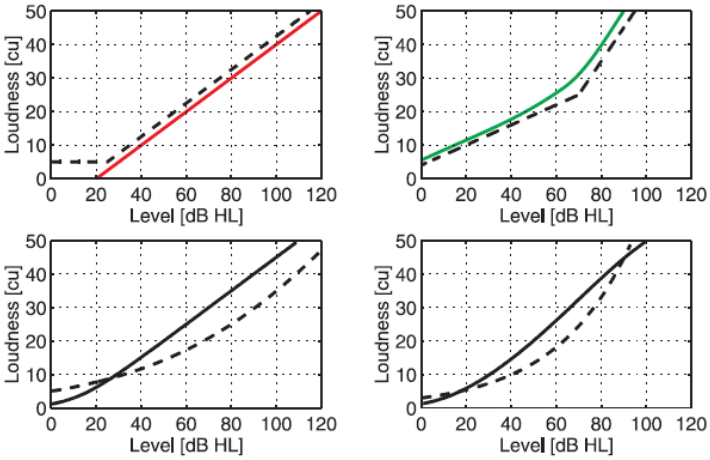

Figure 1: Examples for 8 model functions. Upper left panel: Linear function (Fechner's law) (solid) and linear function with offset (dashed). Upper right panel: ‘broken stick’ with and without smoothing. Lower left panel: Modified Fechner function (Heller et al., 1997) (solid) and further modified Fechner function (dashed). Lower right panel: Function by Nowak (1990) (solid) and function by Stevens (1956) (dashed) (from Brand, 2000).

Mean bias and intra-individual standard deviations of these model loudness functions are shown in Fig. 2. The model function which yielded the lowest bias and the lowest intra-individual standard deviation in response level estimates was the smoothed ‘broken stick’ function. It consists of two straight lines which are connected at the ‘medium'-level and which are smoothed in the transition area. This model function has three free parameters, namely the ‘medium'-level and the two slope values of the two straight lines.

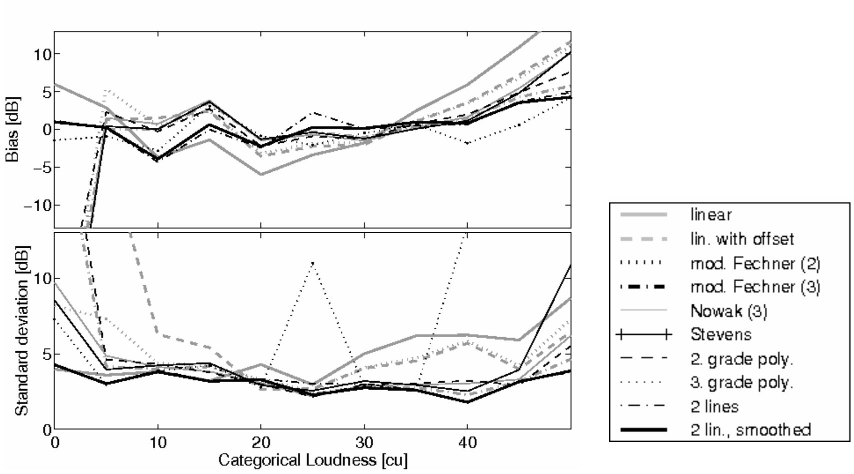

Figure 2: Mean bias (upper panel) and intra-individual standard deviation (lower panel) for the different model functions for hearing-impaired listeners and the adaptive procedure.

Adaptive procedure

The adaptive procedure consists of two phases, which is not obvious to the subject, because he/she rates loudness in both phases. The auditory dynamic range of the subject is roughly estimated in the first phase. In the second phase, more data are collected and the dynamic range, in which the stimuli are presented, is re-estimated twice. Fig. 3 gives a sketch of the adaptive procedure with two iterations. A more detailed description can be found in Brand and Hohmann (2002).

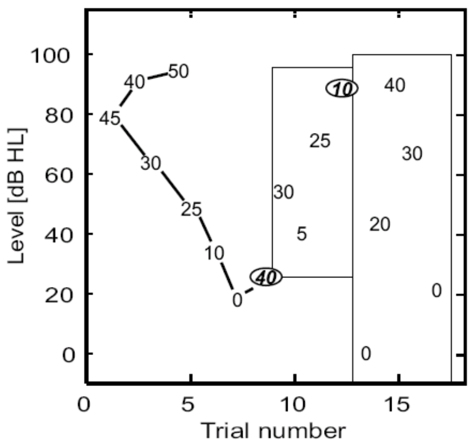

Figure 3: Example of a run produced by the adaptive procedure. The responses are indicated with numbers between 0 (‘inaudible') and 50 (‘too loud'). The numbers that are marked with ellipses indicate obvious outliers. The abscissa indicates the trial number. The ordinate indicates the presentation level. Those presentation levels which belong to the same iteration of the adaptive procedure are combined by rectangles. The upper and lower limits of the rectangles correspond to the limits of the estimated auditory dynamic range per iteration (from Brand and Hohmann, 2002).

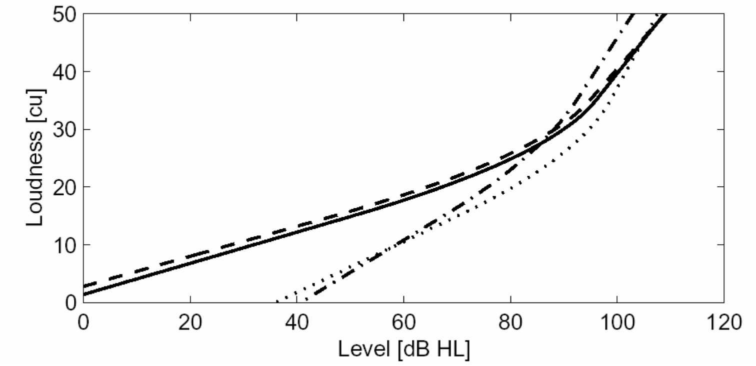

Figure 4: Loudness functions with the median parameters. Normal-hearing subjects with adaptive procedure (solid), normal-hearing subjects with constant stimuli procedure (dashed), subjects with hearing impairment with adaptive procedure (dotted), subjects with hearing impairment with constant stimuli procedure (dash-dotted) (from Brand and Hohmann, 2002).

The adaptive procedure was evaluated using 10 normal-hearing and 10 hearing-impaired subjects. All subjects performed 10 tracks with both, the adaptive procedure and the constant stimuli procedure of Hohmann and Kollmeier (1995).

Figure 4 shows loodness functions for 1 kHz narrow band noises. The parameters (medium level, lower slope, upper slope) of these loudness functions are the median values for both constant and adaptive procedure, respectively.

Figure 5 shows the mean intra-individual standard deviation of the L_x estimate (the 'categorical loudness level' L_x denotes the level which is related to the loudness categories x). The adaptive procedure yielded 20 to 50 % smaller standard deviations than the constant stimuli procedure for both normal-hearing and hearing-impaired listeners.

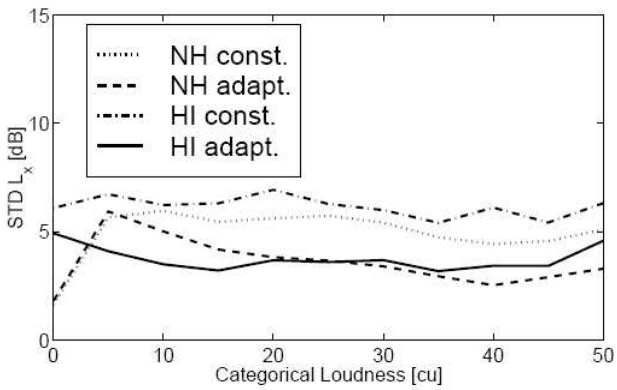

Figure 5: Intra-individual standard deviations of L_x estimates for normal-hearing (NH) and subjects with hearing impairment (HI) with the adaptive and the constant stimuli procedure (from Brand and Hohmann, 2002).

Consequences for hearing aid gain

The use of a model loudness function that is able to parameterise the bending of the loudness function in combination with the improved accuracy of the adaptive procedure (see above) reveals second order effects like temporal loudness summation (Garnier, 1999) or the level-dependency of spectral loudness summation using categorical loudness scaling procedures: Loudness functions of narrowband and broadband stimuli were measured with 8 normal–hearing and 8 hearing–impaired listeners using the ACALOS method (Brand and Hohmann, 2001). In the normal-hearing listeners, narrowband stimuli generally generated loudness functions whose slope increased with increasing level, whereas broadband stimuli generated more linear loudness functions. These differences can be explained by the level dependence of spectral loudness summation, which is known to be most prominent at moderate levels (Zwicker et al., 1957; Florentine and Zwicker, 1979). In hearing–impaired listeners, the narrowband loudness functions generally showed a more linear shape than in the normal–hearing listeners. A consequence of these findings might be that the optimal shape of the input/output curve of a hearing aid is affected not only by the signal power in the respective frequency channels but also by the bandwidth of the input signal. Figure 6 shows examples for such different shapes of loudness functions and resulting i/o cases for different band witdths. However, there were considerable differences between listeners in both groups regarding the individual shape and absolute position of the loudness functions. Therefore, no normative reference could be extracted that would allow for a quantification of the bandwidth-effect on an individual basis.

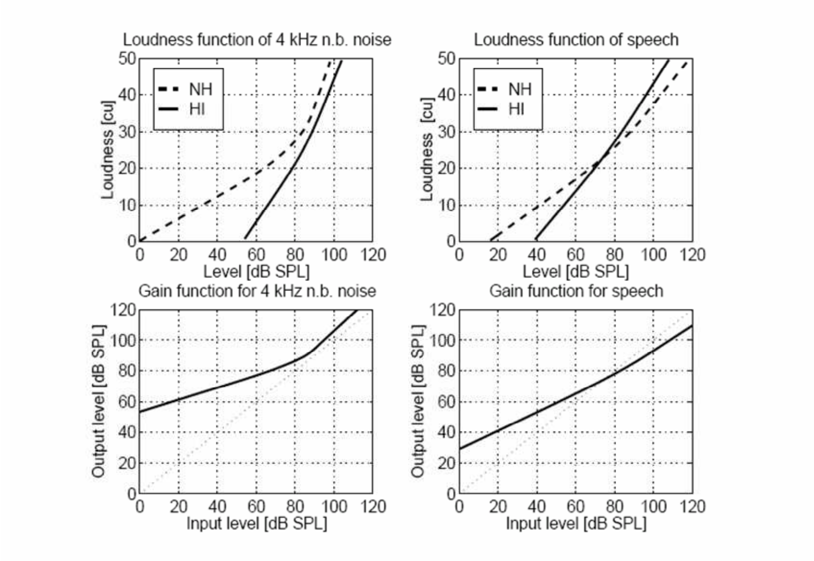

Figure 6: Examples for loudness functions and hearing aid gain functions: The upper left panel shows the loudness functions for the 4 kHz narrowband stimulus for a normal–hearing listener (dashed line) and a hearing–impaired listener (solid line). The hearing aid gain function that results if the loudness perception of the hearing–impaired listener should be restored to that of the normal–hearing listener is shown in the lower left panel. The upper right panel shows the loudness functions of the speech stimulus for the same subjects. The lower right panel shows the resulting gain function for the speech stimulus (from Brand and Hohmann, 2001).

Discussion and outlook

With the background of the research of the last years, the arguments against categorical loudness scaling by Elberling (1999) are discussed as follows:

Argument 1: *Different scaling methods produce different loudness functions that can not be compared.* This problem exists in many other psychoacoustic procedures as well and can be solved by using a standardised procedure.

Argument 2: *Accuracy of categorical loudness scaling is bad.* The accuracy was improved considerably since Elberling's study (see above). The accuracy can be regarded as sufficient for diagnostic and hearing aid fitting.

Argument 3: *Normal-hearing listeners show large differences in loudness functions.* This problem exists in many other psychoacoustic procedures as well and can be solved indirectly: In normal-hearing listeners usually very similar loudness functions are found for the two ears and at different frequencies (Brand and Hohmann, 2001). Therefore, if a listener has normal hearing at some frequencies at one ear, the loudness functions in this normal area can be used as the individual reference loudness function for all frequencies and for both ears. In cases with no such normal area one has to refer to the median loudness functions of normal-hearing listeners.

Argument 4: *Input/output functions of hearing aids based on loudness scaling are unclear.* It is true that categorical loudness scaling does not predict directly how the input/output function of a hearing aid should be. Consequently, prescriptive rules based on loudness functions have to be used, comparable to the prescriptive rules for hearing aid fitting based on the pure tone audiogram. One possibility for such a prescriptive rule based on categorical loudness scaling would be the restoration of loudness perception. In this case, the individual target loudness function described under argument 3 has to be used. However, other rules that do not aim at loudness restoration but at maximizing speech intelligibility might be used as well.

Argument 5: *Most hearing-impaired listeners can be fitted using a simple fine tuning.* At the end of all of these prescriptive rules, fine tuning is still required. However, the calculations of Elberling assume that loudness functions are always linear which than results in lines input/output functions that can be fine tuned relatively easily because there are only two free parameters. However, since loudness functions are not linear (see above) more complex input/output functions are required, which can hardly be realized using fine tuning, because there are two many parameters. On the other hand, the shape of these more complex input/output functions can be derived from categorical loudness scaling. Furthermore, Elberling's calculations neglect that also the estimate of the hearing threshold has some uncertainty of approximately 5 dB which biases his comparison biases in favor to the fine tuning approach.

A problem occurs with the model loudness function described above when other procedures are used: The 'broken stick' parameterization is not adequate for other loudness scaling procedures. A comparison of different studies shows that the number of response alternatives influences the place where the slope of the loudness function increases. Allen *et al.* (1990) used 7 response alternatives and found a more rapid increase in ratings at the higher end of the scale above the category 'loud'. The ACALOS (Brand and Hohmann, 2002) procedure uses 11 response alternatives and the slope of the loudness function increases at the loudness 'medium'. Keider *et al.* (1999) used 23 response alternatives and found a smooth increase of slope between soft and medium. Heller (1985) used 51 and more response alternatives and found an increase of the slope around category 'soft'). Taken together, there is a clear tendency towards lower knee points of the fitted loudness function with increasing number of response alternatives.

If the ACALOS procedure is used for the diagnosis of recruitment, the slope of the loudness function at low loudness categories is the best criterion. In normal-hearing listeners the lower slope has a typical value of 0.3 cu/dB. If this value is increased to values larger than 0.6 cu/dB this is a relatively clear indicator for recruitment. 'cu' stands for 'categorical units' from 'inaudible' (0 cu) to 'too loud' (50 cu) there are 51 categorical units.

As described in this article, categorical loudness scaling procedures have been improved in the recent years. However, there are still remaining problems that should be addressed in the future:

One problem is that in some measurements not enough loudness ratings occur for the categories from 'loud' to 'too loud'. This is often caused by the fact that the maximum presentation level which has to be limited due to technical as well as safety reasons might be lower than the level required for these loudness ratings. Furthermore, in some listeners the loudness function is very steep for large loudness categories. Since the adaptive procedure was optimized to produce an even level distribution in the auditory dynamic range, the density of stimuli in the loudness domain is lower at high loudnesses than at low loudnesses. If the number of loudness ratings for high loudness categories is too low, the fit of the upper part of the loudness function is not stable and can not be used for diagnostics and hearing aid fitting. In such cases which occur approximately in 10% of the measurements the upper range may be estimated for example by estimating the uncomfortable level based on Pascoe (1988). Alternatively, more loudness scaling data have to be collected.

Another problem is that categorical loudness scaling procedures sometimes have problems with listeners using hearing aids with dynamic compression. The model loudness functions usually used in the different procedures are optimized for unaided listeners. Since the hearing aid can profoundly modify the shape of the loudness function (e.g. by peak-clipping) it might be the case that the model function can not be fitted to the loudness ratings adequately. In such cases special loudness functions with more degrees of freedom have to be used.

Conclusions

1. The accuracy of categorical loudness scaling has been improved in the recent years.

2. Older procedures using linear model functions are not able to describe loudness functions adequately.

3. There is a new ISO standard on categorical loudness scaling.

4. For diagnostic reasons the lower slope of the loudness function is most important.

5. Even second order effects like spectral or temporal effects of loudness summation can be measured using categorical loudness scaling.

Acknowledgement

This article was funded by Audiologie Initiative Niedersachsen and by the EU Project HearCom.

Literature

Allen JB, Hall JL, Jeng PS (1990) Loudness growth in 1/2–octave bands (LGOB) – A procedure for the assessment of loudness. J Acoust Soc Am, 88:745–753.

Brand T (2000) Analysis and optimization of psychophysical procedures in audiology. Dissertation, Carl von Ossietzky Universität, Physics, Oldenburg.

Brand T, Hohmann V (2001) Effect of hearing loss, centre frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling. Audiology, 40(2), p. 92-103.

Brand & Hohmann (2002) An adaptive procedure for categorical loudness scaling. J Acoust Soc Am, 112:1597-1604.

Elberling C (1999). Loudness Scaling Revisited. J Am Acad Audiol, 10:248-260

Florentine M, Zwicker E (1979) A model of loudness summation applied to noise-induced hearing loss. Hear Res,1, 121-132.

Garnier S, Micheyl C, Berger-Vachon C, Collet L (1999) Effect of signal duration on categorical loudness scaling in normal and in hearing-impaired listeners. Audiology, 38, 196-201.

Hellbrück J, Moser LM (1985) Hörgeräte–Audiometrie: Ein computerunterstütztes psychologisches Verfahren zur Hörgeräteanpassung. Psycholog Beiträge, 27:494–508.

Heller O (1985) Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU). Psycholog Beiträge, 27:478–493.

Heller O, Boretzki M, Fichtl E, Knoblach W, May B, Nowak T, Stock A (1997) Projektbereich: Hilfen für Hörgeschädigte. Psychologisches Institut der Universität Würzburg.

Hohmann V, Kollmeier B (1995) Weiterentwicklung und klinischer Einsatz der Hörfeldskalierung. Audiologische Akustik, 34:48–59.

ISO 16832 (2006) Acoustics – Loudness scaling by means of categories.

Keidser G, Seymour J, Dillon H, Grant F, Byrne D (1999) An efficient, adaptive method of measuring loudness growth functions. Scandinavian Audiology, 28, 3-14.

Kinkel M (2007) The new ISO 16832 ‘Acoustics – Loudness scaling by means of categories’. 8th EFAS Congress/10th Congress of the German Society of Audiolgy, Heidelberg 2007 (on this CD).

Kollmeier B (1997) Hörflächenskalierung - Grundlagen und Anwendung der kategorialen Lautheitsskalierung für Hördiagnostik und Hörgeräteversorgung. In Buchreihe Audiologische Akustik, Median-Verlag, Heidelberg. ISBN 3-922766-26-9,

Nowak T (1990) Loudness scaling based on Fechner's idea. In: Proceedings of the sixth Annual Meeting of the International Society for Psychophysics, Ed. Müller F, Universität Würzburg.

Pascoe DP (1988) Clinical measurements of the auditory dynamic range and their relation to formulas for hearing aid gain. In Hearing Aid fitting, 13th Danavox Symposium. Ed. Jensen JH, 129-152.

Stevens SS (1956) The direct estimation of sensory magnitudes - loudness, Am J Psychol, 69, 1-25.

Zwicker E, Flottorp G, Stevens SS (1957) Critical bandwidth in loudness summation. J Acoust Soc Am, 9:548–5.