Comparison of decision criteria for interaural correlation discrimination

Helge Lüddemann, Helmut Riedel, and Birger Kollmeier

Medizinische Physik, Fakultät V, Institut für Physik, Carl von OssietzkyUniversität Oldenburg, D-26111 Oldenburg, Germany

This contribution presents psychometric functions for interaural correlation (IAC) discriminability with respect to nine reference correlations (+1, 0, –1 and six intermediate values), measured in a novel 2-pair-2-AFC paradigm. Data were analyzed in various ways, to demonstrate that the normalized cross correlation coefficient (ρ) is generally inadequate for a description of psychoacoustical data by means of normal signal detection theory (SDT). Besides this rather destructive outcome, a Thurstone model case V [Thurstone, Psychol. Rev. **34**, 273-286 (1927)] was fitted to the joint set of data from all pairwise comparisons, to determine a decision variable which is fully compatible with equal variance Gaussian SDT (allows for prediction of discrimination rates and thresholds independent of the reference correlation). The optimal decision variable is well approximated by the dB-scaled ratio of energies in the correlated (N_0) and anticorrelated (N_π) signal components, but only if the effects of a noisy periphery are considered.

INTRODUCTION

Binaural listening can facilitate the identification of auditory objects and improve speech intelligibility in complex acoustical situations. However, it is well known that temporal and spectral fluctuations of binaural cues have a tremendous impact on the auditory system's ability to localize sound sources and thereby separate relevant signals and concurrent auditory events, e.g., reverb or ambient noise.

The sound field generated by a single or dominant source of little spatial extend in front of the head is usually characterized by a higher degree of interaural similarity than diffusely lateralized ambient noise and superimposed reflections. Accordingly, many studies support the idea that the interaural correlation (IAC) plays a key role for the temporal and spectral weighting of binaural information during spatial scene analysis (e.g., Faller and Merimaa, 2004), because as an integrative and thus robust measure for the similarity of the two ear signals it may be used to assess both the reliability and relevance of interaural disparities in complex acoustical situations.

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

The normalized cross correlation coefficient ρ: definition and percept

The IAC is usually specified in terms of the cross correlation coefficient ρ, i.e., as the normalized scalar product of the two ear signals $l(t)$ and $r(t)$:

$$\rho = \frac{\int l(t) \cdot r(t)\, dt}{\sqrt{\int l^2(t)\, dt} \cdot \sqrt{\int r^2(t)\, dt}} \qquad \text{(Eq. 1)}$$

The positive range of ρ is often associated with the perceptual continuum between compact (ρ=+1) and diffuse (ρ=0) spatial listening impressions. The just noticeable difference (ρ-JND) between a reference correlation ρ_{ref} and a deviant IAC ρ_{dev} is about 10 times smaller for ρ_{ref}=+1 than for ρ_{ref}=0 (Pollack and Trittipoe, 1959a; Gabriel and Colburn, 1981; Culling *et al.*, 2001; Boehnke *et al.*, 2002; Lüddemann *et al.*, 2009). It has further been demonstrated that ρ-JNDs for ρ_{ref}=+1, 0 and –1 are characterized by the same parameter range asymmetries as thresholds for the detectability of binaural gaps, i.e., brief IAC deviations interrupting an ongoing reference noise (Boehnke *et al.*, 2002; Lüddemann *et al.*, 2007).

Conceptual problems and open questions

Since IAC thresholds critically depend on ρ_{ref}, the amount of correlation change in terms of the normalized ρ can apparently not be used as a unique (zero-variance) measure for the salience of binaural parameter variations in terms of perceptual units. Consequently, although ρ has a tremendous effect on binaural performance, it is disputable if the normalized ρ, *per se*, is the appropriate variable for a linear weighting of binaural cues. It is further unclear if and how the absolute values and signs of ρ and $\Delta\rho$ interact and if the usability of ρ as a decision variable in the framework of signal detection theory (SDT) may be taken for granted.

Despite these facts, comparatively little is known about the perceptual distance between arbitrary values of IAC, and it is not yet well understood by which aspects of the physical stimulus the respective discrimination or detection probability is actually determined.

Another severe, yet unsolved problem is why IAC discriminability degrades in the presence of an additional level imbalance (Pollack and Trittipoe, 1959b), although any interaural level difference (ILD) is equalized by the denominator terms in Eq. (1) so that the cross correlation coefficient ρ is unaffected by ILDs. In addition, van de Par *et al.* (2001) argued that an explicit normalization of interaural levels, *per se*, is highly unlikely to actually occur as part of binaural processing, because the auditory system were required to compensate for level fluctuations in an implausibly accurate manner to achieve the measured binaural performance.

The "effective N_0/N_π-ratio" in dB$(N_0/N_\pi)_{SNR}$

The interaural correlation ρ of a dichotic stimulus $s(t)$ is fully determined by the mixing ratio between its correlated (N_0) and anticorrelated (N_π) components

$$s(t) = \sqrt{\frac{1+\rho}{2}} \cdot N_0 + \sqrt{\frac{1-\rho}{2}} \cdot N_\pi \qquad \text{(Eq. 2)}$$

When expressed in dB, this ratio is equivalent to the Fisher Z-transform of ρ:

$$\tilde{\rho}_{0/\pi}(\rho) = 10 \cdot \log \frac{1+\rho}{1-\rho} \qquad \text{(Eq. 3)}$$

This dB(N_0/N_π) ratio does not only change the metrics of the parameter space, it also normalizes the variance of sample correlations. The **"effective dB(N_0/N_π) transform"**

$$\tilde{\rho}_{\mathsf{SNR}}(\rho) = 10 \cdot \log \frac{1+\rho+\mathsf{SNR}^{-2}}{1-\rho+\mathsf{SNR}^{-2}} \qquad \text{(Eq. 4)}$$

contains an additional parameter to model the effect of a noisy periphery (i.e., reduced signal to noise ratio (SNR), specified as the ratio of RMS amplitudes of signal and peripheral noise) on binaural performance, and to assure finite scale values for $\rho=\pm1$. The relation between normalized IAC and dB(N_0/N_π) transformed values is illustrated in the left panel of Fig. 2.

The dB(N_0/N_π) ratio can also be calculated directly from the left and right signals ($l \pm r$, either stimulus waveforms or outputs of a noisy periphery model), to avoid the previously mentioned issues regarding the disputable normalization (see previous section):

$$\tilde{\rho}_{\mathsf{direct}}(\rho) = 10 \cdot \log \frac{\int (l+r)^2 dt}{\int (l-r)^2 dt} \qquad \text{(Eq. 5)}$$

EXPERIMENTS AND METHODS

Psychometric functions for the discriminability of signals with different interaural correlations were measured in a forced-choice paradigm with two pairs of narrow band signals (each 500 ms duration, 1 ERB centered at 500 Hz, 65 dB SPL) in every trial (2-pair-2-AFC). Subjects had to identify the pair in which the two signals had different correlations (a signal with reference IAC ρ_{ref} followed by a signal with deviant IAC ρ_{dev}). Both stimuli in the control pair had an IAC of ρ_{ref}. All stimuli were generated in Matlab, using the technique described by Culling *et al.* (2001). For all 94 combinations of ρ_{ref} and ρ_{dev}, a total number of 275 trials was acquired, including data from six normal hearing subjects, three male and three female (aged between 25 and 32).

The 94 combinations of reference and deviant IAC were selected after pilot testing so that psychometric functions (1) can be measured with respect to 9 reference correlations, including ρ_{ref}=+1, 0, –1 and 6 intermediate values, (2) have all roughly the same number of sampling points, (3) cover the whole range of possible discrimination rates except for the vicinity of 100% and (4) have a maximal intersect of absolute IAC levels, $\rho_{\text{dev}} = \rho_{\text{ref}} + \Delta\rho$.

In contrast to a 2-signal-2-AFC paradigm, the 2-pair-2-AFC paradigm avoids any misconception or preference of particular sound attributes or decision criteria (e.g., selection of the signal with more compact/broader percept or with higher/lower IAC). It further minimizes any subjective bias and ensures a chance level of 50% in the case of indistinguishable ρ_{ref} and ρ_{dev}. Nevertheless, the hit rates for the selection (preference) of the correct pair within the 2-pair-2-AFC paradigm may be interpreted as probabilities for the preference of $|d(\rho_{\text{dev}})-d(\rho_{\text{ref}})+\eta_1|$ over $|d(\rho_{\text{ref}})-d(\rho_{\text{ref}})+\eta_2|$ according to Thurstone's law of comparative judgement (η_i: random error of the decision variable d). Hence, the experimental design used here allows one to fit a Thurstone model case V to pairwise-comparison data which is essentially based on a discrimination task. The fitting procedure is explained later, in the respective section of the discussion.

RESULTS AND DISCUSSION

Psychometric functions and IAC discrimination thresholds (ρ-JND)

The psychometric functions in Fig. 1 (left) are characterized by very different shapes and slopes, depending on both ρ_{ref} and the sign of the IAC deviation. These dependencies are also apparent in the thresholds for the just noticeable IAC increase or decrease (Fig. 1, right): ρ-JNDs are best for ρ_{ref}=+1 (0.015), slightly larger for ρ_{ref}=–1 (0.042) and largest in the vicinity of zero (0.65/–0.80 for positive/negative deviations). Hence, the ρ-JND strongly depends on $|\rho_{\text{ref}}|$. But also the sign of the IAC deviation and the issue if the task is performed in the positive/negative range of ρ have (interacting) effects on the ρ-JND. The worst ρ-JND is presumably not found at ρ_{ref}=0, but instead at small positive/negative ρ_{ref} for negative/positive IAC deviations.

The dependence of the ρ-JND on ρ_{ref} observed in the present work is in qualitative agreement with previous literature (e.g., Pollack and Trittipoe 1959a; Boehnke *et al.*, 2002). The quantitative differences to their data can be understood as an effect of stimulus bandwidth (Gabriel and Colburn, 1981).

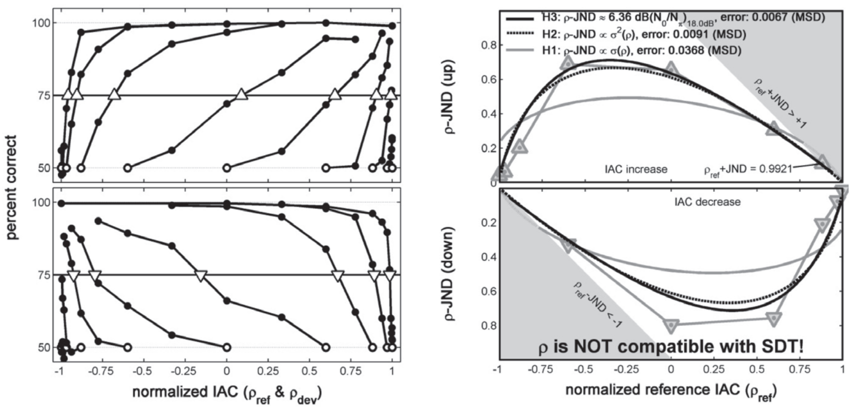

Fig. 1: Left: Psychometric functions for the discriminability of narrowband noise with different IAC (500 ms, 1 ERB centered at 500 Hz). Reference correlations are depicted as empty dots at chance level. (Upper panel: $\rho_{\text{dev}} > \rho_{\text{ref}}$, i.e., IAC increase. lower panel: $\rho_{\text{dev}} < \rho_{\text{ref}}$, i.e., IAC decrease.) **Right**: ρ-JNDs as a function of ρ_{ref}: JNDs for an increase (upper panel, $\rho_{\text{dev}} > \rho_{\text{ref}}$) and for a decrease (lower panel, $\rho_{\text{dev}} < \rho_{\text{ref}}$) of IAC were derived from the 75%-level of the psychometric functions in the respective panels to the left. Lines represent the three models which were fitted to the data, fit functions are described in the text.

Is ρ a valid decision variable? Or is the ρ-JND a constant dB(N_0/N_π) step?

Normal SDT predicts that two stimuli, represented by two decision variables d_1 and d_2 on the perceptual continuum, can be discriminated if the difference $|d_1–d_2|$ exceeds a constant multiple of the respective standard deviation $\sigma(d_1–d_2) = [\sigma^2(d_1)+\sigma^2(d_2)]^{1/2}$. According to Gabriel and Colburn (1981), the actual correlation in temporal/spectral parts of a signal with an overall IAC of ρ is a random variable with a variance σ^2 approximately proportional to $(1–\rho^2)$. Hence, if ρ were a valid decision variable by means of Gaussian SDT (i.e., d proportional to ρ), the measured ρ-JNDs should solve the equation (α_σ: fit parameter, const. for all ρ_{ref}):

$$\rho\text{JND} = \alpha_\sigma \cdot \sqrt{\left(1-\rho_{\text{ref}}^2\right) + \left(1-(\rho_{\text{ref}} \pm \rho\text{JND})^2\right)} \qquad \text{(Eq. 6)}$$

The experimental data in Fig. 1 (right), however, rather suggest that the ρ-JND is modeled far better as a multiple of the variance [$\sigma^2(\rho_{ref})+\sigma^2(\rho_{dev})$], i.e.,

$$\rho\mathsf{JND} = \alpha_{\sigma^2} \cdot \left[\left(1-\rho_{\mathsf{ref}}^2\right) + \left(1-(\rho_{\mathsf{ref}} \pm \rho\mathsf{JND})^2\right)\right] \qquad \text{(Eq. 7)}$$

This also explains why the metrics of the parameter space is modeled quite well as a constant step size on the dB(N_0/N_π) scale, because the local derivative of the Z-transform is inversely proportional to $\sigma^2=(1-\rho^2)$. An internal SNR of 18 dB is assumed to assure finite differences on the dB(N_0/N_π) scale for $\rho_{ref}=\pm1$, i.e.,

$$\tilde{\rho}_{\mathsf{SNR}}(\rho_{\mathsf{ref}}) \pm \alpha_{\mathsf{dB}} = \tilde{\rho}_{\mathsf{SNR}}(\rho_{\mathsf{ref}} \pm \rho\mathsf{JND}) \qquad \text{(Eq. 8)}$$

A comparison of the three analytical models for the prediction of ρ-JNDs for arbitrary ρ_{ref} can be seen in Fig 1, right. The fits suggest that, for an effective SNR of 18 dB, the ρ-JND corresponds to 6.4 dB$(N_0/N_\pi)_{18dB}$. The standard model (Eq. 6), in contrast, cannot properly fit the data. Hence, it is implausible that the auditory system uses the normalized ρ itself as a decision variable which allows for a perceptually adequate description of IAC differences by means of Gaussian signal detection theory (SDT).

Psychometric data as functions of the "effective dB(N_0/N_π) ratio"

As shown in Fig. 2, the usual dB(N_0/N_π) transform lets the psychometric functions for intermediate ρ appear "more normal" than in Fig. 1, left. However, without taking into account the effects of internal noise, psychometric functions seem to "saturate" for $\rho_{dev}\rightarrow\pm1$. If the data are, in contrast, plotted as functions of the "effective dB(N_0/N_π) ratio", assuming an internal SNR of 18 dB, all psychometric functions have a similar shape as if they were shifted (for different ρ_{ref}) or mirrored (for IAC increase/decrease) copies of one unique Gaussian cumulative density function (CDF). Hence, in the domain of the effective dB(N_0/N_π) transformed IAC, equal variance Gaussian SDT is applicable. This suggests that the ρ-JND corresponds to a constant difference by means of dB$(N_0/N_\pi)_{18dB}$. In the domain of normalized IAC, in contrast, SDT is not applicable (see previous section and Fig. 1).

Perceptual distance between arbitrary correlations on a Thurstone scale

The strong dependence of IAC discrimination thresholds on $|\rho_{ref}|$, sign(ρ_{ref}) and sign($\rho_{dev}-\rho_{ref}$) illustrates that the normalized cross correlation coefficient ρ, *per se*, does obviously not reflect the "metrics" of the perceptual continuum. An even bigger problem, however, is that the normalized ρ, because of its statistical properties, cannot be used as an adequate decision variable for the description of IAC sensitivity by means of Gaussian SDT at all (see previous sections).

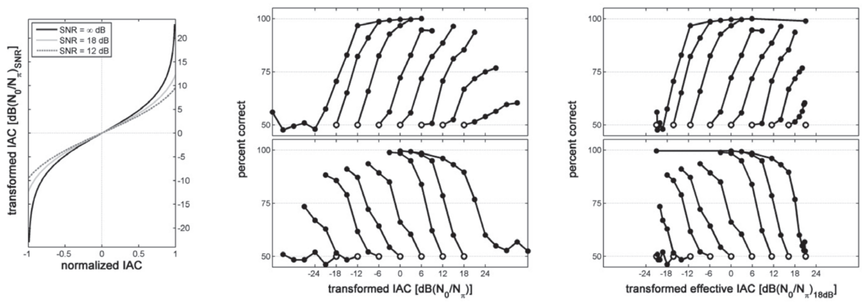

Fig. 2: Left: Relation between normalized IAC and the respective dB(N_0/N_π) transformed values, for different SNR-parameters (see Eq. 3 and Eq. 4). Middle: same data as in Fig. 1 (left), but with the stimulus parameter expressed in dB(N_0/N_π) without correction terms for reduced SNR (usual Z-transform according to the black line in the left panel). **Right**: same data as in Fig. 1 (left), but with the stimulus parameter in terms of the "effective dB(N_0/N_π) ratio", assuming that processing errors before binaural feature extraction can be modeled by an SNR of 18 dB (modified Z-transform according to the grey line in the left panel).

Instead, Figs. 1 (right) and 2 (right) suggest that the above-mentioned flaws of ρ can, for the most part, be resolved if IAC differences are measured in terms of the effective dB(N_0/N_π) ratio. However, the suggestion that the effective dB(N_0/N_π) ratio were the appropriate decision variable in IAC discrimination tasks is hard to approve, and its suitability (and testability) might suffer from the restricted selection of heavily constrained models which all essentially represent competing hypotheses for different parameter space metrics. After all, such a hypothesis-driven (instead of data-driven) approach cannot clarify the question if there exists an even better scale transformation than the modified Z-transform (Eq. 4).

If there is any way to describe IAC discrimination by means of standard normal SDT with a single decision variable T being monotonically related to the external stimulus parameter ρ, an optimal lookup table for the scale transformation T: $\rho \rightarrow T(\rho)$ can be found by fitting a Thurstone model case V to the data. Note that, since the Thurstone model represents $T(\rho)$ as a lookup table, any further assumptions regarding the functional character of $T(\rho)$ are explicitly avoided. Instead, an iterative fitting algorithm was implemented to simultaneously adjust the differences between any two scale values, $\Delta T = T(\rho + \Delta\rho) - T(\rho)$, or: $\Delta T = T(\rho_{dev}) - T(\rho_{ref})$, until the Thurstone model could predict the percentage of correctly discriminated stimuli by a Gaussian CDF with unique shape for all ρ_{ref} and ρ_{dev}, i.e., $p_{correct} = \Phi_\sigma(|\Delta T|)$ with the global scale parameter σ being independent of ρ (see Fig. 3, left).

Hence, the Thurstone scale values T(ρ) which are shown in Fig. 3 (right) directly reflect the metrics of the perceptual continuum. The global scale parameter σ was adjusted so that one upward/downward JND in the domain of normalized ρ (Fig.

3, right, x-axis) always corresponds to an increment/decrement by one unit in the transformed domain (Fig. 3, right , y-axis). In other words: one unit on the Thurstone scale (ΔT=+/–1) always corresponds to the upward/downward ρ-JND, regardless of ρ_{ref}:

$$T(\rho_{\text{ref}}) \pm 1 \equiv T(\rho_{\text{ref}} \pm \text{JND}) \qquad \text{(Eq. 9)}$$

The fitted line in Fig. 3 (right) illustrates that the Thurstone scale values (dots) are roughly proportional to the effective dB(N_0/N_π)-transformed IAC after noisy monaural preprocessing with an SNR of 18 dB. The proportionality factor of the fit function indicates that one unit on the Thurstone scale corresponds to 5.73 dB$(N_0/N_\pi)_{18\text{dB}}$. However, in contrast to the dB(N_0/N_π) transform, the Thurstone scale is not exactly symmetric in the sense $T(\rho) = -T(-\rho)$. Instead, positive correlations are mapped to scale values between 0 and 4.3, whereas the negative range of ρ is represented by only 3.3 units. This indicates that IAC sensitivity is generally better in the positive range or ρ than in the negative range.

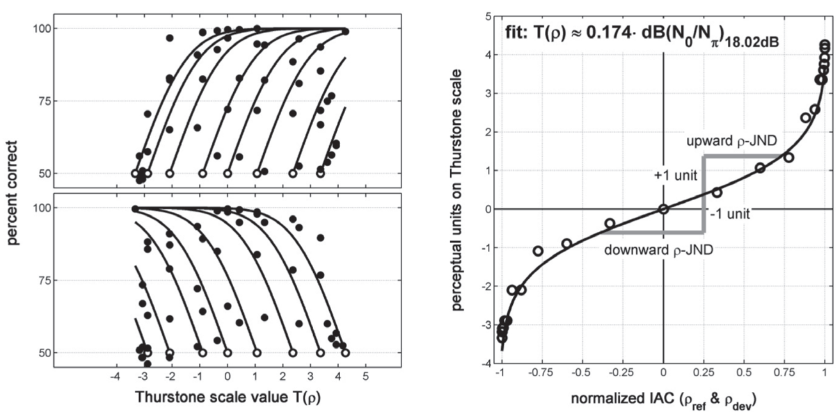

Fig. 3: Left: Final state of the Thurstone scale fitting procedure: lines are model functions, dots are data. **Right**: The metrics of the perceptual continuum can be depicted as a nonlinear transform of the physical parameter ρ into perceptual quantities $T(\rho)$. $T(\rho)$ is well approximated by a multiple of the dB(N_0/N_π) transformed effective IAC. The fit parameters suggest that the internal SNR amounts to 18 dB and that one perceptual unit (JND) corresponds to 5.7 dB$(N_0/N_\pi)_{18\text{dB}}$ (inverse value of the fit parameter 0.174).

SUMMARY

The shape and the slope of psychometric functions critically depend on the reference IAC ρ_{ref} and are different for IAC increase/decrease. The discrepancy between the standard deviation of sample correlations and measured ρ-JNDs suggests that the normalized ρ is no valid decision variable by means of standard SDT. A Thurstone model was used to investigate the relation between the physical stimulus parameter ρ and its related decision variable $T(\rho)$. In good approximation, $T(\rho)$ is proportional to the dB-scaled ratio of energies in the correlated (N_0) and anticorrelated (N_π) signal components after noisy peripheral preprocessing. In contrast to the normalized ρ, the proposed measure is compatible with equal variance Gaussian SDT, i.e., it allows one to predict the percentage of correct discriminations independent of $|\rho_{\mathrm{ref}}|$, sign(ρ_{ref}) or sign($\rho_{\mathrm{dev}}-\rho_{\mathrm{ref}}$). The ρ-JND corresponds to a constant difference of about 6 dB$(N_0/N_\pi)_{18\mathrm{dB}}$ in the transformed domain. Because it is also sensitive to the ILD (cf. Eq. 5), we conclude that the dB(N_0/N_π) transform reflects the perceptually adequate metrics for the weighting of concurrent auditory events in coherence-based algorithms and models.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, project KO 942/19-1).

REFERENCES

Boehnke, S. E., Hall, S. E., and Marquardt T. (**2002**). "Detection of static and dynamic changes in interaural correlation," J. Acoust. Soc. Am. **112**, 1617-1626.

Culling, J. F., Colburn, H. S., and Spurchise, M. (**2001**). "Interaural correlation sensitivity," J. Acoust. Soc. Am. **110**, 1020-1029.

Faller, C., and Merimaa, J. (**2004**). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075-3089.

Gabriel, K. J. and Colburn, H. S. (**1981**). "Interaural correlation discrimination: I. bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394-1401.

Lüddemann, H., Riedel, H., and Kollmeier, B. (**2007**). "Logarithmic scaling of interaural cross correlation: a model based on evidence from psychophysics and EEG," In *Hearing: from sensory processing to perception – 14th International Symposium on Hearing*, edited by B. Kollmeier *et al.* (Springer, Berlin), pp. 379-388.

Lüddemann, H., Riedel, H., and Kollmeier, B. (**2009**). "Electrophysiological and psychophysical asymmetries in sensitivity to interaural correlation steps," Hear. Res. **256**, 39-57.

Pollack, I., and Trittipoe, W. J. (**1959a**). "Binaural listening and interaural noise cross correlation," J. Acoust. Soc. Am. **31**, 1250-1252.

Pollack, I., and Trittipoe, W. J. (**1959b**). "Interaural noise correlation: examination of variables," J. Acoust. Soc. Am. **31**, 1616-1618.

Thurstone, L. L. (**1927**). “A Law of comparative judgment,” Psychol. Rev. **34**, 273-286.

van de Par, S., Trahiotis, C., and Bernstein, L. R. (**2001**). “A consideration of the normalization that is typically included in correlation-based models of binaural detection,” J. Acoust. Soc. Am. **109**, 830-833.

van der Heijden, M., and Trahiotis, C. (**1997**). “A new way to account for binaural detection as a function of interaural noise correlation,” J. Acoust. Soc. Am. **101**, 1019-1022.