ON THE DIFFERENCE BETWEEN CROSS-CORRELATION AND EC-BASED BINAURAL MODELS

PACS REFERENCE: 43.66.Pn, 43.66.Ba, 43.66.Dc

Breebaart, Jeroen; van de Par, Steven; Kohlrausch, Armin
Philips Research Laboratories Eindhoven
Prof. Holstlaan 4
5656 AA Eindhoven
The Netherlands
Tel: +31 40 2744236
Fax: +31 40 2744675
E-mail: jeroen.breebaart@philips.com

ABSTRACT

The majority of binaural models are comprised of a peripheral transduction stage, a stage of binaural interaction and finally a decision stage. For the binaural interaction stage generally two approaches have been followed. The first approach relies on the cross-correlation (i.e., a measure for the similarity) of corresponding left-right signals stemming from the peripheral transducer. The second method is based on Durlach's EC-theory (i.e., on the difference between left and right signals) (Durlach, 1963). Although in the past it has been argued that these two approaches are linear functions of each other, we will show that this statement is incorrect for finite-length stimuli and that the predictive scope of both approaches is not equal.

1. INTRODUCTION

In the past decades, two general types of binaural interaction have been discussed. The first method relies on *coincidences* of spikes stemming from the left and right ears as a function of a relative delay. Jeffress (1948) introduced this concept as coincidence detectors that receive information from the left and right ears with different delay lines. The coincidence detectors are usually modeled as cross-correlation units and hence the complete system of delay lines and coincidence detectors can be elegantly modeled using the cross-correlation function. In binaural detection experiments, it is assumed that the (change in the) cross-correlation function is used as a decision variable. Considerable neurophysiological evidence has been published to suggest that such a system actually exists in the auditory system. The second method is based on *differences* in the signals at the ears. This concept is the basic idea of Durlach's EC (Equalization-Cancellation) theory. It is assumed that the auditory system first attempts to equalize (the E step) the waveforms stemming from both ears using a limited repertoire of transformations (a variable relative delay and a variable amplitude scaling). In the original model, this transformation led to an increase in the signal-to-masker ratio in binaural unmasking conditions and hence a binaural masking level difference (BMLD). More recently, it was shown that the residue of the EC process is a good predictor for the audibility of changes in binaural cues (Breebaart et al, 2001). Although somewhat limited, there is physiological support that EC-like neurons exist in the auditory system (Joris & Yin, 1995).

2. BINAURAL DETECTION BASED ON THE CROSS-CORRELATION

We will focus here on a standard BMLD condition in which a noise masker having a certain expected (normalized) interaural cross-correlation ρ_{ref} presented with a tonal signal that is presented interaurally out-of-phase (i.e., an NρSπ condition). Furthermore, we assume that the product of bandwidth and duration of the masker is small, and hence the masked thresholds are determined by stimulus uncertainty (Breebaart & Kohlrausch, 2001). For these conditions, equations can be derived that quantify the uncertainty in different decision variables that are often used in binaural models. If the uncertainty in the decision variable is compared to the change in that decision variable due to the addition of a signal we can predict the amount of informational masking, i.e., the amount of masking that results from stimulus uncertainty. Hence it is assumed that the decision variables are not subject to internal noise (i.e., stemming from internal processing stages).

The left and right waveforms $L(t)$ and $R(t)$ for an NρSπ condition are usually generated by mixing two independent noises $N_1(t)$ and $N_2(t)$ and the signal $S(t)$ according to

$$\begin{aligned} L(t) &= \sqrt{\frac{1+\rho_{ref}}{2}}N_1(t) + \sqrt{\frac{1-\rho_{ref}}{2}}N_2(t) + S(t) \\ R(t) &= \sqrt{\frac{1+\rho_{ref}}{2}}N_1(t) - \sqrt{\frac{1-\rho_{ref}}{2}}N_2(t) - S(t) \end{aligned}.$$

The *unnormalized* cross-correlation of these waveforms, $R_{LR}(\tau)$, is given by

$$R_{LR}(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} L(t-\tau/2)R(t+\tau/2)dt,$$

where T denotes the integration window. $R_{LR}(\tau)$ can be written for T approaching infinity as

$$R_{LR}(\tau) = \rho_{ref}R_{11}(\tau) - R_{SS}(\tau),$$

where R_{11} and R_{SS} denote the (unnormalized) autocorrelation functions of masking noise N_1[1] and the signal, respectively.

Besides the unnormalized cross-correlation, the normalized cross-correlation function, $\rho_{LR}(\tau)$, is also an often used decision variable in binaural models and is given by

$$\rho_{LR}(\tau) = \frac{\int_{-T/2}^{T/2} L(t-\tau/2)R(t+\tau/2)dt}{\sqrt{\int_{-T/2}^{T/2} L^2(t-\tau/2)dt \int_{-T/2}^{T/2} R^2(t+\tau/2)dt}}.$$

It can easily be shown that as T approaches infinity, $\rho_{LR}(\tau)$ can be simplified to

[1] Since T approaches infinity, the autocorrelation functions R_{11} and R_{22} of the two independent noises are equal.

$$\rho_{LR}(\tau) = \frac{\rho_{ref} R_{11}(\tau) - R_{SS}(\tau)}{R_{11}(0) + R_{SS}(0)}.$$

The above equations are valid for an infinite integration time. In most psychoacoustic measurements, however, stimulus durations of a few hundred milliseconds are typically used. This has important implications for the equations shown above. First, the autocorrelation functions of the noise and signal are replaced by *estimates* of these functions based on the short intervals. Said differently, the autocorrelation values are stochastic variables that differ from token to token. Second, the correlation between the individual signal components (the two noises and the signal) is almost never exactly zero. For the unnormalized cross-correlation R_{LR}, the decision variable can be approximated by (neglecting correlations between individual components)

$$R_{LR}(\tau) = \frac{1+\rho_{ref}}{2}\hat{R}_{11}(\tau) - \frac{1-\rho_{ref}}{2}\hat{R}_{22}(\tau) - \hat{R}_{SS}(\tau)$$

where the hats indicate that estimates of the auto- and crosscorrelation functions are obtained from single finite intervals. In order to estimate the variance of the decision variable ($R_{LR}(\tau)$) we assume the following: $R_{11}(\tau)$ and $R_{22}(\tau)$ follow an approximate normal distribution with an approximate variance of (Rice, 1959)

$$\sigma^2\left\{\hat{R}_{11}(\tau)\right\} = \sigma^2\left\{\hat{R}_{22}(\tau)\right\} \approx \frac{R^2{}_{11}(0)}{BT},$$

where B denotes the bandwidth of the noise and T denotes the duration. The resulting variance of $R_{LR}(\tau)$ due to variance of the noise autocorrelation functions is then given by

$$\sigma^2\left\{R_{LR}(\tau)\right\} \approx \frac{(1+\rho_{ref}^2)R^2{}_{11}(0)}{2BT}.$$

Furthermore, the change in the decision variable ΔR_{LR} due to addition of the signal is given by

$$\Delta R_{LR}(\tau) = -\hat{R}_{SS}(\tau).$$

The detectability index d', assuming best detection for τ=0 is then given by

$$d' = \frac{\Delta R_{LR}(0)}{\sigma\left\{R_{LR}(0)\right\}},$$

which simplifies to

$$d' = \frac{R_{SS}(0)\sqrt{2BT}}{R_{11}(0)\sqrt{1+\rho_{ref}^2}} = \frac{S}{N}\frac{\sqrt{2BT}}{\sqrt{1+\rho_{ref}^2}}.$$

At threshold, the value of d' is typically around +1 although its exact value depends on the experimental procedure. Given a certain value for d', we can predict the masked threshold (in dB) for the signal as a function of the bandwidth and duration of the noise following

$$\frac{S}{N} = 10\log_{10}\left(d'\sqrt{\frac{1+\rho_{ref}^2}{2BT}}\right).$$

While the actual proof is beyond the scope of this paper, we can show in a similar way that the predicted signal-to-masker ratio (in dB) using the *normalized* cross-correlation as the decision variable can be approximated by

$$\frac{S}{N} = 10\log_{10}\left(\frac{d'(1-\rho_{ref})(1+\rho_{ref})}{\sqrt{2BT}}\right).$$

3. BINAURAL DETECTION BASED ON THE EC RESIDUAL

The EC residual *E* is the amount of signal power that remains after equalization and cancellation transformations. The EC transformation includes a relative time delay τ, a level adjustment α and computes the remaining difference signal power $E(\tau,\alpha)$:

$$E(\tau,\alpha) = \frac{1}{T}\int\left(\sqrt{\alpha}L(t-\tau/2) - \sqrt{\frac{1}{\alpha}}R(t+\tau/2)\right)^2 dt,$$

which equals

$$E(\tau,\alpha) = \alpha R_{LL}(0) + \frac{1}{\alpha}R_{RR}(0) - 2R_{LR}(\tau).$$

Here we see the analogy with the cross-correlation: the EC residual is a linear combination of the powers of the left and right signals and the (unnormalized) cross-correlation. This result has led to suggestions that binaural models based on the EC theory and the cross-correlation are equivalent (e.g., have the same predictive power) since their decision variables are linear functions of one another (Colburn & Durlach, 1978; Green, 1992). This statement may be true for infinitely-long integration windows, but it is not generally true for noise tokens typically used in BMLD experiments. This statement will be demonstrated below.

For a partially-correlated noise masker (i.e., $N\rho S\pi$), the decision variable for α=1 and τ=0 becomes

$$E(0,1) = 2(1-\rho_{ref})\hat{R}_{22}(0) + 4\hat{R}_{SS}(0).$$

If a similar analysis as for the correlation is performed, the detection threshold as a function of the variables *d'*, *B* and *T* using the EC residual as decision variable can be predicted according to

$$\frac{S}{N} = 10\log_{10}\left(\frac{d'(1-\rho_{ref})}{2\sqrt{BT}}\right).$$

Interestingly, the three different decision variables share a common product of bandwidth (*B*) and noise duration (*T*) as denominator. This term results from the power fluctuations of the individual noise sources. However, each decision variable depends differently on the (reference) correlation ρ_{ref}. These differences will be discussed in the next section.

4. COMPARISON OF PREDICTED THRESHOLDS

In this section, the predicted thresholds based on informational masking according to the three different models for an NρSπ condition are discussed. Figure 1 shows predicted thresholds[2] for the three decision variables presented above combined with experimental data. The correlation ρ_{ref} was varied between 0.05 and 0.99. For all predictions, d' was set to +1.26 (according to a two-down, one-up adaptive procedure and a 3IFC task). For the left and right panels, the product of bandwidth and masker duration equals 3 and 30, respectively. It was shown previously (Breebaart & Kohlrausch, 2001), that for these values, behavior thresholds are mainly determined by stimulus variability.

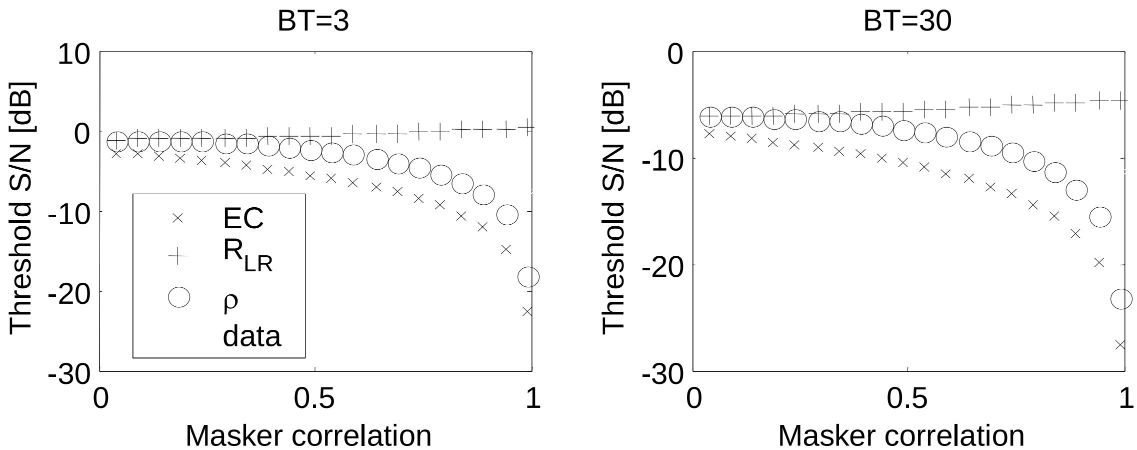

Figure 1. Predicted thresholds as a function of the masker correlation for three different decision variables: the EC residual (crosses), the unnormalized cross-correlation (plus signs) and the normalized cross-correlation (circles). The left panel corresponds to BT=3, the right panel to BT=30. The downward triangles are experimental data adapted from Breebaart & Kohlrausch (2001).

As can be observed from Fig. 1, the predicted thresholds are different across decision variables. The unnormalized cross-correlation (plus signs) shows only a slight decrease with a decrease in correlation, while its values are mostly higher than the experimental data (downward triangles), especially for masker correlations above 0.9. Hence this is a quantitative support of the statement by van de Par et al. (2001), that the unnormalized cross-correlation suffers from level variance of the masker in a binaural detection experiment. The predictions for the EC theory and the normalized cross-correlation deviate by 4.5 dB for a masker correlation close to +1 and by about 2 dB for a masker correlation near zero, the latter resulting in higher predicted thresholds. Furthermore, an increase of the product of time and bandwidth of the masking noise results in a decrease in the predicted thresholds. This is qualitatively supported by the experimental data (downward triangles). However, none of the three decision variables describe the experimental data satisfactorily.

---

[2] To support the mathematical derivations for the thresholds, a simulation was performed to verify the equations. For each combination of parameters, 1000 masking noise samples were generated digitally and the decision variables were computed accordingly. Subsequently, masked-threshold predictions according to the variance in the simulated decision variables were compared with the theoretical derivatives. These values differed by less than 1 dB.

5. DISCUSSION

Derivations for thresholds based on informational masking in an NρSπ condition were given as a function of the duration, correlation and bandwidth of the masking noise. These thresholds are based on the uncertainty for three different decision variables. This uncertainty is based on the statistics of the noises that are used to generate the stimuli. This means that 'internal' noise of the auditory system is not incorporated in the simulations. Hence the predicted thresholds form the lower boundaries for detection performance given a certain decision variable.

The predictions for the unnormalized cross-correlation are considerably higher than the predictions for the normalized cross-correlation and the EC residual, especially for masker correlation values near +1. Moreover, no BMLD is observed for masker correlations near +1 and thresholds decrease somewhat with a decrease in correlation. This is in contrast to experimental data, which show low thresholds for a masker correlation near +1 and increase with decreasing correlation. Given the fact that the predictions are lower boundaries of the thresholds for a certain decision variable, it can be concluded that the unnormalized cross-correlation per se is not suitable as decision variable in binaural models.

Both the EC residual and the normalized cross-correlation show a correlation dependence of thresholds which qualitatively matches subject data. However, both models predict thresholds below the actual data. Furthermore, the EC model predicts lower thresholds than the normalized cross-correlation. This indicates that either these decision variables are not a suitable model for the detection process of the human auditory system or additional losses of information exist in the human auditory system.

6. CONCLUSIONS

It has been shown that three decision variables which are often used in binaural models (the cross-correlation, the normalized cross-correlation and the EC residual) predict different thresholds in NρSπ conditions with a small product of bandwidth and masker duration. Although some aspects of the data can be qualitatively predicted by these theories, none of the theories can quantitatively predict the data.

REFERENCES

Breebaart, J. and Kohlrausch, A. (2001). The influence of interaural stimulus uncertainty on binaural signal detection. *J. Acoust. Soc. Am.* 109, 331-345.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001). Binaural processing model based on contralateral inhibition. I. Model structure. *J. Acoust. Soc. Am.* 110, 1074-1088.

Colburn, H. S. and Durlach, N. I. (1978). Models of binaural interaction. In *Handbook of perception*, edited by E. Carterette and M. Friedman, Academic, New York.

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *J. Acoust. Soc. Am.* 35, 1206-1218.

Green, D. (1992). On the similarity of two theories of comodulation masking release. *J. Acoust. Soc. Am.* 91, 1769.

Jeffress, L. A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.* 41, 35-39.

Joris, P. X., and Yin, T. C. T. (1995). Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences. *J. Neurophysiol.* 73, 1043-1062.

Rice, S. O. (1959). *Selected papers on noise and stochastic processes*. Dover, New York.

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