DYNAMIC CUES IN BINAURAL PERCEPTION

Richard M. Stern and Stephen J. Bachorski

Department of Electrical Engineering and Biomedical Engineering Program
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213 USA

1. INTRODUCTION

The binaural system is exquisitely sensitive in its ability to resolve very small interaural differences of arrival time for signals presented to the two ears. It is at first surprising, then, that the system is relatively poor at processing temporal changes in these interaural time differences. In recent years there has been an increased interest in the perception of binaural stimuli with time varying interaural time delays (ITDs), interaural intensity differences (IIDs), and interaural correlation. In this paper we review some of the results of psychoacoustical experiments using stimuli with time-varying interaural differences, and we present some new theoretical predictions to describe some of these data.

2. PERCEPTION OF STIMULI WITH TIME-VARYING INTERAURAL DIFFERENCES

In this brief review we will consider only experiments with stimuli using constantly changing interaural differences, presented through headphones. This excludes, for example, the literature on the precedence effect, as well as experiments using moving loudspeakers as sound sources.

First insights into the dynamics of the binaural system came largely from subjective reports of the perceptual images of stimuli presented with time-varying ITDs. For example, the presentation of two monaural pure tones with slightly different frequencies, one to each ear, establishes an interaural time difference that varies linearly as a function of time. Licklider, et al. (1950), Perrott and Musicant (1977), and others have noted that the presentation of two tones with frequencies below about 1 kHz causes the perception of a "rotating tone" that smoothly traverses the head in a quasi-sinusoidal function of time. As the IFD is increased, the dominant perception becomes one of a change in binaural loudness. Still greater IFDs cause the perception of a roughness in the sound and, ultimately, two difference smooth monaural tones.

Time-varying ITDs and IIDs are also established when two *binaural* tones of slightly different frequencies are presented simultaneously. It is easy to show that the instantaneous ITDs and IIDs produced by these stimuli [which we refer to as $\tau(t)$ and $\alpha(t)$], are complex periodic waveforms, with a shape that depends on the ITD, IID, and overall amplitude of each of the components. The fundamental frequency of $\tau(t)$ and $\alpha(t)$ is equal to the difference between the frequencies of the component tones.

Pairs of binaural tones have been used as stimuli in the detection experiments of McFadden et al. (1972), and in recent unpublished interaural time discrimination experiments by Kaiser and Stern. McFadden et al. measured N_mS_m and N_0S_π detection using 400-Hz tonal maskers, varying the frequency of the tonal target as an experimental parameter. They found that the masking level differences (MLDs) for the two conditions were a nonmonotonic function of target-to-masker frequency separation, and that the greatest improvement in detection performance provided by the binaural system occurs when target and masker differ both by about 10 to 15 Hz. As McFadden et al. point out, it is probable that detection of the target in the N_0S_π configuration is affected by perception of motion of the brief target-masker complex, at least for small target-masker frequency separations.

Kaiser and Stern measured interaural time jnds for 500-Hz tonal targets in the presence of tonal maskers as a function of target-to-masker ratio and masker frequency, with the stimuli presented in the N_0S_0 and $N_\pi S_0$ configurations. They observed very little difference between jnds obtained in the presence of the N_0 versus the N_π maskers. In contrast, Cohen (1981), Ito et al. (1982), and Stern et al. (1983), all of whom measured interaural time and amplitude jnds in the presence of broadband maskers, observed that time jnds in the presence of the N_0 maskers were much smaller. These comparisons are interesting because if we assume that the peripheral auditory system includes a narrowband filtering operation, the probability distributions of $\tau(t)$ and $\alpha(t)$ for these experiments are identical for the tonal and broadband maskers. We believe that these two sets of jnds are different in form because $\tau(t)$ and $\alpha(t)$ for the combined target and masker vary much more slowly with the filtered broadband maskers than with the tonal maskers.

More recently, Grantham has studied various aspects of the dynamics of the binaural system in a series of experiments using forced-choice paradigms. For example, Grantham and Wightman (1977) measured the ability to discriminate broadband stimuli with sinusoidally varying ITDs from spectrally-matched diotic stimuli. Their results indicate that the binaural system is extremely "sluggish" in its response to the time-varying stimuli, as the discrimination performance becomes progressively worse as the frequency of the ITD is increased above about 5 Hz. In a later experiment Grantham and Wightman (1979) measured the detectability of short-duration tonal targets in the presence of narrowband maskers with sinusoidally varying interaural correlation, and they found that the binaural system is similarly slow to respond to temporal variations in the interaural correlation. From these results Grantham and Wightman estimated the effective integration time of the binaural system to be between approximately 45 and 140 ms for 500-Hz stimuli. Blauert, in an earlier (1972) study reported similar phenomena, but he used an experimental method that required the listeners to adopt a subjective criterion in forming their jnd.

3. BINAURAL MODELS FOR TIME-VARYING STIMULI

All of the above results are consistent with the general hypothesis that the binaural system can perfectly track the perceptual images of binaural stimuli with very slowly changing ITD, IID, or interaural correlation, but that the system is very slow to respond when the rate of change of these stimulus parameters exceeds a few Hz. However, the exact mechanism in the binaural system that limits the perception of temporally fluctuating stimuli has not yet been identified.

Most of the models that have been developed to describe and to predict the subjective perception of binaural stimuli are composed of similar structural components, which include peripheral bandpass filtering of the signals to the two ears, mechanisms for the estimation of ITD and IID, and a mechanism for the formation of spatial percepts from the estimated ITD and IID (cf. Colburn and Durlach, 1978). The slow temporal response that "blurs" the perception of time-varying ITDs and IIDs could occur after the level of, the peripheral filtering, the extraction of ITD and IID, or the percept-formation mechanism, or any combination of these levels of the system.

a) The Running Crosscorrelation Model

The earliest binaural model that could explicitly describe the perception of stimuli with time varying ITDs is that of Sayers and Cherry (1957). In the simplest formulation of this model, interaural timing information of the stimulus is obtained via a short-term or "running" crosscorrelation operation of the form

$$R_s(t,\tau) = \int_{-\infty}^{t} s_L(\alpha)\, s_R(\alpha-\tau)\, W(t-\alpha)\, d\alpha \qquad (1)$$

where $s_L(t)$ and $s_R(t)$ are the inputs to the two ears, and $W(t)$ is a short-duration pulse that serves to cause more recent events to be more heavily weighted in the computation of the crosscorrelation function. Sayers and Cherry used the temporal window $W(t) = e^{-kt}$ in their calculations, and proposed that 6 ms was a reasonable value for the integration time constant (Cherry, 1961). Sayers and Cherry also proposed a method by which the inputs to the crosscorrelator could be weighted by the IID of the stimulus, and a mechanism for judging the subjective laterality of the signals. They also proposed, in a subsequent elaboration of the model, a peripheral frequency analysis that could be performed by autocorrelation of the signals to the two ears.

The running crosscorrelation model implicitly assumes that the lowpass operation is part of the ITD estimator, since the expression for $R_s(t,\tau)$ is mathematically equivalent to the output of a causal lowpass filter with impulse response e^{-kt} when the signal $s_L(t)s_R(t-\tau)$ is the input. Since the expanded Sayers and Cherry model also includes a time-intensity conversion that takes place before the running crosscorrelation operation, temporal variations in IID would be lowpass filtered as well. Unfortunately, Sayers and Cherry never compared the predictions of their model to experimental results using stimuli with time-varying interaural differences. Blauert and Cobben (1978) did apply a similar model to transient stimuli that produced the precedence effect, using a value of 5 ms for the integration time.

b) *Extension of the Position-Variable Model to Describe Time Varying Binaural Phenomena*

The position-variable model (Colburn, 1973; Stern and Colburn, 1978) was developed to describe the lateralization, discrimination, and detection of binaural stimuli in terms of firing patterns of the auditory nerves. In this model information related to the ITD is contained within a timing function, the IID information is contained within an intensity function, and a position variable $\hat{P}$ is generated by computing the centroid (or center of mass) of the product of the timing and intensity functions. The firing times of the auditory nerve are modeled as sample functions of nonhomogeneous Poisson processes. The timing function, which is related to the interaural crosscorrelation of the stimuli, is obtained from the outputs of hypothetical units which record coincidences of this auditory-nerve activity, after a deterministic interaural delay. Specifically, the timing function is defined to be the total number of counts of a network of such units as a function of their interaural delay. This function may be regarded as a quantification of earlier theories suggesting crosscorrelation mechanisms (Jeffress, 1948; Sayers and Cherry, 1957; Licklider, 1959). The intensity function is assumed to be a Gaussian-shaped pulse of constant width, with a location that depends on the IID of the stimulus.

In the original formulation of the position-variable model, it was assumed that all coincidences occurring during the presentation interval of the stimulus tone contribute equally to the timing function. Stern and Colburn (1978) represent the number of coincidences occurring over a stimulus interval by the random function $L_m(\tau)$, where τ is the characteristic interaural delay parameter. It can be shown that if the duration of the coincidence window, T_W, is sufficiently short, the expected value of $L_m(\tau)$ is approximately

$$E[L_m(\tau)] = T_W \int_0^{T_S} r_L(t)\, r_R(t-\tau)\, dt \tag{2}$$

where T_S is the duration of the stimulus. The functions $r_L(t)$ and $r_R(t)$ are the rate functions of the Poisson processes characterizing the auditory-nerve response at a given characteristic frequency to the signals presented to the two ears.

This model can easily be modified to produce a time-varying representation of interaural timing information by allowing more recent coincidences to be given greater weight in the formation of the timing function. The expected value of this new time-varying function, which we refer to as $R_m(t,\tau)$ is

$$E[R_m(t,\tau)] = T_W \int_{-\infty}^{t} r_L(\alpha)\, r_R(\alpha-\tau)\, W(t-\alpha)\, d\alpha \tag{3}$$

where the weighting function $W(t)$ in our calculations is of the form $W(t) = e^{-kt}$ for t greater than zero. This expression is obviously very similar to the original running crosscorrelation function proposed by Sayers and Cherry (Eq.1), except that in the present model we are, in effect, crosscorrelating auditory-nerve rate functions rather than the acoustical stimuli themselves. If T_W is sufficiently small, the occurrences of coincidences of auditory-nerve activity can also be modeled as Poisson processes. It is then easy to verify that the function $R_m(t,\tau)$ can be modeled as a filtered Poisson process (Parzen, 1962). Because of this, the variance of $R_m(t,\tau)$ resulting from the variability of the auditory-nerve patterns is

$$\mathrm{Var}[R_m(t,\tau)] = T_W \int_{-\infty}^{t} r_L(\alpha)\ r_R(\alpha-\tau)\ W^2(t-\alpha)\ d\alpha \qquad (4)$$

Given $E[R_m(t,\tau)]$ and $\mathrm{Var}[R_m(t,\tau)]$, time-varying predictions for the mean and variance of the subjective lateral position, $\hat{P}$, may be obtained using the assumptions stated in Stern and Colburn (1978). Specifically, we obtain the expected value and variance of $\hat{P}(t)$ from the equations

$$E[\hat{P}(t)] = \frac{\int_{-\infty}^{\infty} \tau L_I(\tau)\ p(\tau)\ E[R_m(t,\tau)d\tau}{\int_{-\infty}^{\infty} L_I(\tau)\ p(\tau)\ E[R_m(t,\tau)]d\tau} \qquad (5)$$

and

$$\mathrm{Var}[\hat{P}(t)] = \frac{\int_{-\infty}^{\infty} \tau^2\ L_I^{\,2}(\tau)\ p(\tau)\ \mathrm{Var}[R_m(t,\tau)]d\tau}{\{\int_{-\infty}^{\infty} L_I(\tau)\ p(\tau)\ E[R_m(t,\tau)]d\tau\}^2} \qquad (6)$$

where $p(\tau)$ is a function describing the relative number of fiber pairs as a function of their characteristic interaural delay, and $L_I(\tau)$ is the weighting function reflecting the stimulus IID, as discussed in Stern and Colburn (1978).

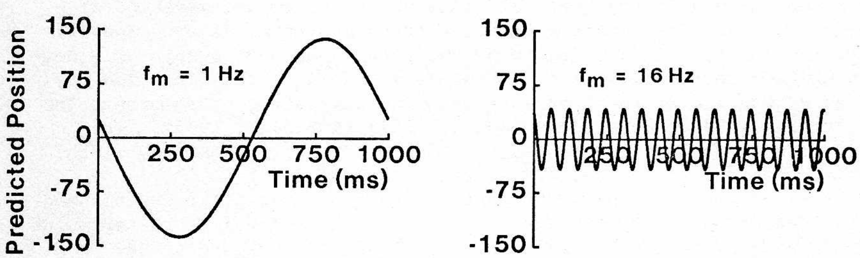

Fig 1. Sample position predictions of the expanded position-variable model for stimuli with sinusoidally-varying ITDs of 1 and 16 Hz

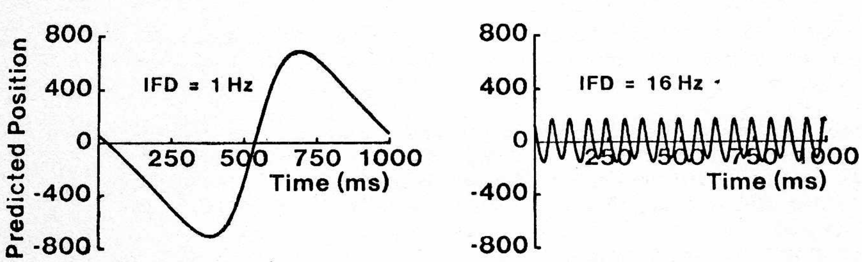

Fig 2. Position predictions for binaural stimuli with interaural frequency differences of 1 and 16 Hz

Figure 1 shows sample predictions for $E[\hat{P}(t)]$ for stimuli presented with sinusoidally varying ITDs of frequencies 1 and 16 Hz. Figure 2 shows similar predictions for a "rotating-tone" stimulus (producing a linearly increasing ITD) presented with IFDs of 1 and 16 Hz. Both sets of predictions were obtained using the same exponential temporal weighting used by Sayers and Cherry and Blauert and Cobben, with k equal to $(2\pi)(5)$ radians/s. This weighting function may be thought of as a lowpass filter with cutoff 5 Hz, or as a leaky integrator with a time constant of about 32 ms. We have found that the peak amplitude of excursion of $E[\hat{P}(t)]$ rolls off at about 6 dB/octave above 5 Hz for both sets of stimuli, which is consistent with the shape of $W(t)$. The units of the vertical axis are arbitrary, but consistent over the four sets of stimuli.

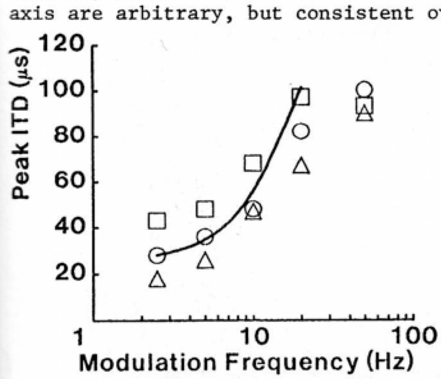

Fig. 3. Comparison of theoretical predictions (curve) and data (symbols) of Grantham and Wightman (1978) from an experiment measuring the smallest "amplitude" of a sinusoidally-varying ITD needed for discrimination from a diotic stimulus

In Figure 3 we compare predictions of the extended position-variable model to some of Grantham and Wightman's (1978) results describing the ability to discriminate targets with sinusoidally varying ITDs from diotic targets. These predictions were obtained by modelling the observed performance as that of an ideal receiver discriminating between two time-varying signals of the form $E[\hat{P}(t)]$ in the presence of an additive Gaussian noise process with average power equal to the average value of $Var[\hat{P}(t)]$. This is an oversimplified realization of the model, since it ignores the possibility that subjects may make use of temporal fluctuations in $Var[\hat{P}(t)]$ in forming their judgments. The theoretical predictions in Figure 3 are only relative, as the absolute predictions were freely adjusted to roughly describe the average of the data points at the modulation frequency of 2 Hz. Since the average variance of $\hat{P}(t)$ was found to be approximately constant over all ITDs and modulation frequencies of interest in the Grantham and Wightman experiment, predicted discrimination performance depends primarily on the "energy" of the function $E[\hat{P}(t)]$ for the stimuli with the sinusoidally varying ITDs. It is seen that, at least with the above assumptions, the model provides a reasonable description of the discrimination data for low modulation frequencies.

We presently believe that this type of model has the ability to characterize experimental phenomena incorporating dynamically changing ITDs, IIDs, and interaural correlations, although it is clear that many details remain to be worked out. For example, while the present calculations assume poor (running) temporal resolution only at the level of the interaural timing processor, it is also quite possible that the mechanisms producing the intensity functions and/or the mechanisms generating the subjective position percept are similarly limited in their ability to process binaural stimuli with dynamic interaural cues. We are currently attempting to gain greater insight into these phenomena and classes of models by comparing predictions and data for other discrimination experiments of Grantham and Wightman, McFadden et al., and Kaiser and Stern, under the assumption that performance in these experiments is limited by the variability of the auditory-nerve activity.

Acknowledgements. This research was partially supported by NIH grants 5 R01 NS14908 and 1 T32 GM07477. We also thank Stuart Meyer for assistance in some of the original theoretical calculations.

REFERENCES

Blauert, J. (1972). On the Lag of Lateralization Caused by Interaural Time and Intensity Differences. *Audiology* 11, 265-270.

Blauert, J. and Cobben, W. (1978). Some Consideration of Binaural Cross-correlation Analysis. *Acustica* 39, 96-103.

Cherry, C. (1961). Two Ears - but One World. In: *Sensory Communication.* (W.A. Rosenblith, ed.). pp. 99-117. Cambridge, MIT Press.

Cohen, M.F. (1981). Interaural Time Discrimination in Noise. *J. Acoust. Soc. Amer.* 70, 1289-1293.

Colburn, H.S. (1973). Theory of Binaural Interaction Based on Auditory-Nerve Data. I. General Strategy and Preliminary Results on Interaural Discrimination. *J. Acoust. Soc. Amer.* 54, 1458-1470.

Colburn, H.S., and Durlach, N.I. (1978). Models of Binaural Interaction. In: *Hearing*. Vol. IV of Handbook of Perception (E.C. Carterette and M.P. Friedman, eds.) New York, Academic Press.

Grantham, D.W. and Wightman, F.L. (1978). Detectability of Varying Interaural Temporal Differences. *J. Acoust. Soc. Amer.* 63: 511-523.

Grantham, D.W. and Wightman, F.L. (1979). Detectability of a Pulsed Tone in the Presence of a masker with Time-Varying Interaural Correlation. *J. Acoust. Soc. Amer.* 65, 1509-1517.

Ito, Y., Colburn, H.S., and Thompson, C.L. (1982). Masked Discrimination of Interaural Time Delays with Narrowband Signal. *J. Acoust. Soc. Amer.* 72, 1821-1826.

Jeffress, L.A. (1948). A Place Theory of Sound Localization. *J. Comp. Physiol. Psychol.* 41, 35-39.

Licklider, J.C.R. (1959). Three Auditory Theories. In: *Psychology: A Study of a Science, Study 1.* Vol. 1 (E.S. Koch, ed.). New York: McGraw-Hill.

Licklider, J.C.R., Webster, J.C., and Hedlun, J.M. (1950). On the Frequency Limits of Binaural Beats. *J. Acoust. Soc. Amer.* 22, 468-473.

McFadden, D.M., Russell, W.E., and Pullman, K.A. (1972). Monaural and Binaural Masking Patterns for a Low-Frequency Tone. *J. Acoust. Soc. Amer.* 51, 534-543.

Parzen, E. (1962). Stochastic Processes (Holden-Day, San Francisco).

Perrott, D.R., and Musicant, A.D. (1977). Rotating Tones and Binaural Beats. *J. Acoust. Soc. Amer.* 61, 1288-1292.

Sayers, B. McA., and Cherry, E.C. (1957). Mechanism of Binaural Fusion in the Hearing of Speech, *J. Acoust. Soc. Amer.* 29, 973-987.

Stern, R.M., Jr., and Colburn, H.S. (1978). Theory of Binaural Interaction Based on Auditory-Nerve Data. IV. A Model for Subjective Lateral Position. *J. Acoust. Soc. Amer.* 64, 127-140.

Stern, R.M., Jr., Slocum, J.E., and Phillips, M.S. (1983). Interaural Time and Amplitude Discrimination in Noise. *J. Acoust. Soc. Amer.* (in press).