informa
healthcare

Original Ar icle

International Journal of Audiology 2006; 45(Supplement 1):S34–S44

hearing

Binaural processing in normal hearing activities is based on the ability of listeners to use the information provided by the differences between the signals at the two ears. The most prominent differences are the interaural time difference and the interaural level difference, both of which depend on frequency. This paper describes the stages by which these differences are estimated by the physiological structures of the auditory system, summarizes the sensitivity of the human listener to these differences, and reviews the nature of the interaural differences in realistic environments.

El procesamiento binaural durante actividades auditivas habituales está basado en la habilidad de los oyentes para usar la información proporcionada por las diferencias entre las señales en ambos oídos. Las diferencias más prominentes son de tiempo y de intensidad interaural, ambas dependientes de la frecuencia. Este trabajo describe las etapas por las cuales estas diferencias son estimadas por las estructuras fisiológicas del sistema auditivo; resume la sensibilidad del oyente humano a éstas diferencias y revisa la naturaleza de las diferencias interaurales en ambientes reales.

The fact that the signals at the two ears are different leads to a number of advantages in listening binaurally. Some of these advantages are related to the simple ability to exploit the location of the better placed ear. Other advantages are based on the ability to extract information from the differences between the signals. This extraction involves the analysis of the interaural differences in the waveforms, i.e., binaural processing, which is the focus of this paper. These differences are usually described in terms of the interaural time difference (ITD), the interaural level difference (ILD), the interaural cross-correlation coefficient (ICC), and the energy in the difference in the waveforms. Each of these quantities are generally considered as a function of frequency, specifically computed from bandpass filtered versions of the input stimuli. These parameters are not independent, many are related to the interaural cross-correlation function (ICF), and they are frequently used to understand, characterize, and model binaural phenomena.

The initial processing of auditory inputs is summarized briefly here in terms of signal processing operations. Acoustic inputs are filtered to generate to a set of narrowband waveforms that are transduced to generate neural firing patterns on the auditory nerve. These patterns are further processed in the brainstem nuclei, resulting in neural patterns that are sensitive to the interaural differences in the narrowband filtered stimuli. Thus, the result of the peripheral and brainstem processing is a neural representation that is equivalent to a sequence of estimates of interaural time and level differences for each frequency band. This is the basic input to the binaural processor.

These initial stages of binaural processing are relatively well understood and provide the input for central processing that interprets the interaural-difference information to generate perceptions that facilitate interpretation of the acoustic world. The interpretation of this information involves a combination of bottom-up and top-down processing. In order to understand more about this processing, it is useful to start with simple environments containing single sources with minimal reverberation and then address the more complex environments that involve multiple, simultaneous sources and significant reverberation. Real environments, which generally include multiple sources and reflections, are very complicated acoustically and raise challenges for listeners, for scientists attempting to understand and model hearing in such environments, and for

ISSN 1499-2027 print/ISSN 1708-8186 online
DOI: 10.1080/14992020600782642

H. Steven Colburn
Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215.
E-mail colburn@bu.edu

engineers attempting to develop artificial processing systems to assist listeners in such environments. Although these challenges are substantial even when the listener has normal hearing, they often become crucial when the listener suffers from a significant hearing impairment.

The cochlea filters sound into relatively narrow (one-sixth to one-third octave) frequency bands, creating a parallel representation of the input broken down into mutiple frequency channels. The resulting frequency-based representation of sound affects all subsequent stages of neural auditory processing and auditory perception, including binaural processing and binaural hearing. As a result, understanding binaural hearing depends upon understanding the nature of narrowband signals. The effective bandwidth of auditory peripheral filtering increases roughly proportionally with the center frequency of the channel, so that low-frequency channels have a narrow bandwidth (in Hz) compared to high-frequency channels.

The impact of peripheral bandpass filtering of acoustic stimuli is profound, particularly because filters are relatively narrow so that waveforms have well-defined envelopes. Basically, the acoustic input is filtered to give narrowband stimuli with several important properties. First, the fine structure (the rapid oscillations at the center frequency of the narrowband filter) and the envelopes of the signal (corresponding to the time-varying amplitude of the oscillations) can be specified almost separately as two distinct time functions. Second, the phase of the fine structure varies with time. This phase variation corresponds to small variations in the length of individual cycles of the fine structure in the waveforms (i.e., in the deviations around the average cycle length). Although this phase variation is a general property of narrowband waveforms, it is less evident than the envelope and the fine structure oscillations. Third, because the signals are narrowband, the envelopes and the phase-variations are relatively smooth functions of time.

The maximum rates of variation in both the envelope and phase increase with the bandwidth of the signal. When a broadband signal is presented to a listener, the effective bandwidth of the signals represented in each frequency channel is determined solely by the bandwidth of the peripheral filter, so the maximal rates of envelope and phase variation depend on the center frequency of the channel being considered. In particular, because the bandwidth (in Hz) is broader for high-frequency channels compared to low-frequency channels, the variations in the envelope and phase are generally more rapid in the high-frequency channels.

If the input is broadband relative to the auditory peripheral filters, the bandwidths of the filtered signals are determined by the auditory periphery. However, if the input signal has a narrow bandwidth relative to the auditory peripheral filter responsive to that input, the input signal bandwidth will determine the bandwidth of the signal in the auditory periphery, and the envelope variaion will be slower. For instance, for tonal stimuli, the envelope and the phase in the peripheral representation are constant (since the stimulus bandwidth is zero) except for transients at the beginning and end of the stimulus. Similarly, for purely amplitude-modulated waveforms, there is no phase variation in the peripheral responses of the auditory system, and for purely frequency- or phase-modulated waveforms, there is no envelope variation.

The firing patterns of the primary auditory nerve, i.e., the eighth or cochlear nerve, can be modeled by rectifying and low-pass filtering the narrowband filter outputs and using the result as the rate of a random firing generator. At all frequencies, the firing rate of the auditory nerve varies with the envelope (i.e., essentially with the short-term amplitude of the signal). The exact timing of individual spikes depends on the center frequency of the corresponding filter. For low center frequencies, the neural responses are highly synchronized to the fine structure. In contrast, at high center frequencies, neural responses do not follow the fine structure because of temporal limitations in the biophysical properties of cells in the auditory periphery. Instead, in high frequency channels, the sometimes rapid fluctuations in the signal envelope cause synchronous bursts of neural spikes that track the envelope shape. The cut-off frequency at which the fine-structure timing information begins to decline varies with species, and is near 800 Hz in the cat (although some synchronization to the fine structure is seen above 4 or 5 kHz) and 8 kHz in the barn owl. For human listeners, the lack of perceptual sensitivity to interaural phase (and to any ongoing fine structure delays) above about 1.5 kHz is often modeled as a consequence of the loss of synchronization as frequency increases. The fact that listeners can discriminate ongoing time delay for narrowband high-frequency waveforms is consistent with the use of synchronization to the envelopes of high-frequency waveforms.

Both physiological and perceptual empirical results support the idea that binaural comparisons are made only between frequency-matched, narrowband inputs from the left and right ears. This postulate, that interaural differences are computed from comparisons of inputs from the same frequency bands in each ear, is supported by both psychophysical and physiological data. Specifically, when narrowband waveforms with different center frequencies are used in interaural time difference (ITD) sensitivity experiments, performance rapidly degrades if the center frequencies of the inputs become separated by more than the commonly assumed bandwidths of the peripheral auditory filters (c.f., Nuetzel and Hafter, 1981). Physiological experiments show that binaurally sensitive neurons are driven by ipsilateral and contralateral inputs that generally have the same frequency tuning (Guinan et al, 1972; Goldberg and Brown, 1969; Boudreau and Tsuchitani, 1970; Yin and Chan, 1990), again consistent with frequency-band-by-frequency-band binaural comparisons.

Given such results, it is reasonable that research focuses on

IPDs are not available (due to loss of fine-structure sensitivity);

create images in the center of the head (e.g., with a more intense,

expected, approximately the same as the interaural differences of the stronger signal by itself (as determined by the location of the dominant source relative to the head). If the signal levels are comparable at a single ear, the combined signal can have almost any amplitude from the sum of the amplitudes to the difference. This leads to an extremely large range of possible interaural level differences. Similarly, the resultant interaural phase difference can be outside the range of the interaural phases of the individual source waveforms. When the two sources are variable, or are fixed sources with random noise waveforms, the situation is even more complex, and the interaural differences vary randomly. A similarly complex pattern of interaural differences are generated in reverberant rooms, as discussed further below.

In the discussion of simple environments, the only circumstance considered in which there was more than one sound source was the case in which the perception of a single target source was degraded by the presence of a single, much stronger masking source. In the preceding discussion, we considered the case of two nearly identical sources. From both practical and theoretical viewpoints, it is important to consider conditions in which more thane circuit is tuned to the 547-Hz frequency band and the processing is based on short-time estimates (four cycles of the center frequency). When there is minimal reverberation, as shown in the left column (anechoic), the ITD values are all relatively close to the ITD value corresponding to the direction of the source. The network response shows that in this case the ITDs are all near zero for the 0-degree (straight ahead) direction, near 0.5 ms for the 45-degree direction, and near 0.7 ms for the 90-degree direction. The middle column shows the effects of reverberation for a location near the middle of the classroom that was used for this simulation. It should be apparent that the distribution of ITD values over time show a relatively strong dispersion throughout the approximately one-second stimulus waveform. The values are concentrated near the anechoic values but there is substantial variability. The right column shows the effects of stronger reverberation, as calculated for a location near the corner of the room. In this case, the variability of the ITD values is still larger and approaches a uniform distribution.

The effects of reverberation on the interaural level difference can also be substantial. An example here is based on recordings made from the two ears of KEMAR in a room in which all six surfaces may be covered with materials having different sound absorption properties. Figure 3 (from Kidd et al, 2005a) shows the variation of the measured ILD versus frequency for three room conditions. The room conditions are foam-covered (''FOAM''; near anechoic), typical untreated IAC booth surfaces (''BARE''; small amount of reverberation) or plexiglas-covered (''PLEX''; high reverberation). Note that there is a substantial reduction in the ILD as the reverberation increases, especially at high frequencies where the anechoic ILDs are larger. This reduction in ILD is also accompanied by a decrease in the peak of the cross-correlation function and in the direct to reverberant ratio.

The impulse responses for these three room conditions are shown in Figure 4. The recordings were made with a single microphone suspended at the approximate location of the center of the subject's head. It is obvious that the amount of reverberant energy increases steadily as the room surfaces become more reflective. The direct-to-reverberant ratio (D/R) is also given for each recording. Note that the most reflective case had a negative D/R ratio in decibels indicating that there was more reflective energy than direct energy at the location of the head.

These observations illustrate that interaural differences are very complicated in reverberant environments. They vary substantially over time and depend in detail on the specific stimulus and the specific physical environment. It would be expected, therefore, that tests of functional hearing abilities in anechoic and reverberant environments may give different

65(t'ips91('anal34-416.118824(i)-4.29emask)D[(the)-339.464)''),(gy)0-5.

noise processed into sets of equally narrow frequency bands that also were mutually exclusive with the target sentence bands. This masker was intended to provide an estimate of the small amount of energetic masking present for the DBS masker. The other noise masker – same-band noise or SBN – is comprised of a set of narrow bands of noise that directly superimpose on the target bands, thus maximizing energetic masking.

Examples of the magnitude spectra for a target paired with samples of each of these maskers is shown in Figure 5.

The sentences used were from the Coordinate Response Measure (CRM) corpus (Bolia et al, 2000) which have the format, ‘‘Ready [callsign], go to [color] [number] now.’’ The task is to identify the color and number associated with a specified target callsign. The DBS masker always had a different callsign, color, and number than the target.

The measured values are speech reception thresholds (SRTs) computed as target-to-masker ratios corresponding to approximately 51% correct on the psychometric functions. The stimuli were played over loudspeakers in a sound-treated room. The location of the target loudspeaker was always straight ahead of almost 15 dB for the NH group is significant in that: 1) it demonstrates that a large release from masking due to spatial separation of sources may be obtained because of perceptual factors, rather than traditional binaural analysis, and 2) it confirms that spatial-perceptual cues may provide a large benefit to listeners with sensorineural hearing loss in complex, multi-source listening environments.

Effects of reverberation on spatial release from masking

small amount of masking to begin with as discussed above. Second, the amount of masking in the SBN condition shows

Freyman, R.L., Balakrishnan, U. & Helfer, K.S. 2004. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. , 115, 2246–2256.

Gabriel, K.J. & Colburn, H.S. 1981. Interaural correlation discrimination: I Bandwidth and level dependence. , 69, 1394–1401.

Goldberg, J.M. & Brown, P.B. 1969. Response of Binaural Neurons of Dog Superior Olivary Complex to Dichotic Tonal Stimuli: Some Physiological Mechanisms of Sound Localization. , 32, 613–636.