Table I. Detection thresholds, in dB SPL, of correlated (C) and uncorrelated (U) binaural noise-band targets in the presence of binaural tonal maskers.

Tonal Masker	Noise Target	TT	AH	Average
500 Hz, 90 dB	400-600 Hz, U	54.1	56.3	55.2
500 Hz, 90 dB	400-600 Hz, C	55.9	56.0	56.0
500 Hz, ΔI:4 dB 92 dB(L), 88dB(R)	100-700 Hz, U	45.0	41.8	43.4
500 Hz, ΔI:4 dB 92 dB(L), 88 dB(R)	100-700 Hz, C	45.8	41.2	43.5
680 Hz, 90 dB	100-700 Hz, U	41.2	35.2	38.2
680 Hz, 90 dB	100-700 Hz, C	43.2	38.8	41.0
4 kHz, 90 dB	3.7-4.3 kHz, U	50.9	48.8	49.8
4 kHz, 90 dB	3.7-4.3 kHz, C	52.5	46.3	49.4
50-Hz Sq. Wave	400-600 Hz, U	68.1	64.3	66.2
50-Hz Sq. Wave	400-600 Hz, C	67.0	67.6	67.3

Our 'worst-case stimulus' was a 680-Hz tone masking a band of noise with energy between 100 and 700 Hz. This type of quantization noise would be difficult to hide even in a monaural situation because most of the energy of the signal occurs in frequencies below the frequency of the masker. As one can see in Table I, the measured detection thresholds are lower than those obtained for any of the other masking conditions. The amount of binaural masking release, however, is on average limited to about 3 dB.

All of these observations are consistent with those of McFadden et al. (1972). They measured the detectability of tonal signals masked by sinusoids as a function of the difference in frequency between the target and the 500-Hz masker. Using the same interaural configurations as we did, they too found about 3 dB of release from masking when the frequency of the target was well below the frequency of the masker. In our 'worst-case stimulus' it is quite likely that the target is detected on the basis of frequencies at the lower end of the pass band, where the target is spectrally most remote from the masker.

Of course, in the case of different types of maskers and signals, considerably more release from masking could occur. It seems possible that spectrally rich sounds might provide enough monaural masking to support a large binaural release. Data obtained with a 50-Hz square wave as a masker permitted an evaluation of binaural release from masking with a complex signal having a rich, yet discrete, spectrum. As shown in Table I, that condition led to an average of only 1 dB binaural masking release.

Finally an attempt was made to compare the absolute threshold levels of Table I with data for comparable situations in the literature. Data on masking of noise by tones are very rare. The only relevant study we found to be the above-mentioned paper by Schroeder at al. (1979), which describes masking levels of noise bands of one critical bandwidth by a pure tone of variable frequency. Their general finding was that, if the masking tone was centered in the target band, the detection threshold was 24 dB below the masker level. For the 8th entry of Table I, where the target is about one critical band wide, this would imply a threshold of $90 - 24 = 66$ dB. The threshold of 49.4 dB which we found is well below this level. In fact, an attempt to reproduce the results of Schroeder et al. using the experimental paradigm and equipment of our study and the first author (AH) as subject resulted in thresholds that were 10–15 dB too low. Moreover, it was found that thresholds were critically dependent on the low-frequency cutoff slope of the noise target. An increase from 48 to 96 dB per octave caused the thresholds to increase by 11 dB. This is not surprising, since the low-frequency slope of the masking pattern of a pure tone is much steeper than 48 dB per octave. It all shows that threshold levels of noise bands masked by pure tones only have a meaning in the absolute sense if the target's spectral shape is exactly controlled and specified.

In summary, several measurements were made concerning the detectability of interaurally correlated and uncorrelated noise. These noises were chosen to mimic quantization noise resulting from bit-reduced coding of stereophonic signals. We have consistently found very small differences in thresholds between correlated and uncorrelated noises. To the degree that these data are generalizable, this means that only about one-half bit more, on average, has to used to code stereophonic material as compared with monophonic material.

References

Brandenburg, K. (1988). High quality sound coding at 2.5 bits/sample. *AES Preprint* **2582**.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust.Soc. Am.* **49**, 467–477.

McFadden, D., Russel, W. E., and Pulliam, K. A. (1972). Monaural and binaural masking patterns for a low-frequency tone. *J. Acoust. Soc. Am.* **51**, 534–543.

Schroeder, M. R., Atal, B. S., and Hall, J. L. (1979). Optimizing digital speech coders by exploiting masking properties of the human ear. *J. Acoust. Soc. Am.* **66**, 1647–1652.

Trahiotis, C., Bernstein, L. R., Buell, T. N., and Spektor, Z. (1990). On the use of adaptive procedures in binaural experiments. *J. Acoust. Soc. Am.* **87**, 1359–1361.

Yost, W. A. (1988). The masking-level difference and overall masker level: Restating the internal noise hypothesis. *J. Acoust. Soc. Am.* **83**, 1517–1521.

Further Bit Rate Reduction Through Binaural Processing

A. J. M. Houtsma, C. Trahiotis*, R. N. J. Veldhuis,

Institute for Perception Research (IPO), P.O. Box 513, 5600 MB Eindhoven, The Netherlands

R. van der Waal

Philips Research Laboratory, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

PACS no. 43.66.Pn, 43.38.Vk, 43.38.Wl

Received 23 May 1995,
accepted 9 January 1996.

* Permanent address: University of Connecticut Health Center, Farmington CT, USA

1. Introduction

In a separate study (Houtsma et al., 1996) we examined how much allowance must be made for possible binaural masking release when designing codecs for stereo sound at low bit rates. It was shown that binaural release effects are limited to no more than 3 dB, which implies an extra half of a bit on average when comparing two-channel digital sound with one-channel mono sound. Other aspects of binaural hearing, however, may enable an even further reducion in average bit rates, by capitalizing on knowledge concerning how and how well the binaural auditory system processes sound. Of course, any coding schemes based on mechanisms and empirical data drawn from the scientific literature on binaural hearing must provide benefits without leading to other types of errors that are detectable monaurally.

It appears likely that principles derived from the literature on binaural hearing could be used to reduce bit rates determined by monaurally-based coding still further. Our reasoning was based on several reports in the literature that discuss the relative ability of interaural differences of time and intensity to move an intracranial image from midline toward one ear. These interaural time and intensity differences provide the information listeners use to discern the origins of sounds external to them, as well as the relative positions of sounds presented through earphones. The following findings are particularly relevant. Firstly, interaural temporal differences conveyed by the fine structure of low-frequency signals (below about 1500 Hz) are very dominant (Bilsen and Raatgever, 1973) and powerful determiners of where sounds will be heard (Blauert, 1983; Bernstein and Trahiotis, 1985;; Wightman and Kistler, 1992). Secondly, interaural time delays within the envelope of narrow-band, high-frequency signals are quite ineffective in determining the perceived location of sounds. This is true even though envelope-based time delays are easily detectable (Blauert 1983; Trahiotis and Bernstein, 1990; Henning,1974; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1976). Thirdly, the effectiveness of interaural intensity differences to determine lateralization is independent of frequency (Yost, 1981). This is the case even though naturally occurring intensity differences measured across a human head are quite small at frequencies below 500 Hz (Shaw and Teranishi, 1968). Lastly, low-frequency sounds do not only dominate judgments about location. They also interfere with the listener's use of interaural temporal information conveyed by the envelopes of signals with frequencies at or above 2 kHz (e.g. Zurek, 1985; Blauert and Divenyi, 1988; Bernstein and Trahiotis, 1992).

When combined, these observations suggested that it might be possible to reduce bit rates for stereophonic signals considerably by only coding information below about 1.4 kHz separately for the left and right channels. We hypothesized that, above that frequency, only one channel's information could be encoded without the spatial qualities of the signals being greatly affected. Of course, such a coding scheme would have to be used adaptively. No such single-channel high-frequency coding will occur if, at any time, insufficient amounts of low-frequency information are present.

2. Listening experiment

Several ways of implementing such a code are conceivable. One way is to code and transmit only one channel of high-frequency information and present it to both the left and the right channel. A second approach is to preserve interaural intensity differences at high frequencies by sending along scalars for each channel and sub-band. Because such scalars could cause aliasing difficulties across sub-bands, we chose to implement and evaluate the first scheme. We passed the left and right outputs of a CD player through a pair of low-pass and a pair of high- pass filters, each with a cut-off frequency of 1.4 kHz. We then sent the low-frequency portions of the left and right signals to the left and right loudspeakers. One of the high-pass signals was eliminated, and the other high-pass signal was sent to both the left and the right loudspeaker at equal levels. We then compared the original stereophonic CD signal with the newly made test signal. Sometimes we listened to ongoing music in a room resembling a living room using high-quality commercial electronics and loudspeakers, while manipulating a hand-held switch to change between the two conditions. Other times, we listened in a special sound studio equipped with professional electronics, loudspeakers and headphones. In those cases we used short fragments of digitized musical signals, listened to each coded fragment, and compared it with the original and with other coded fragments. Typically two to five very highly trained listeners judged the quality of the sounds under both listening conditions, and made comments during the process of listening.

3. Results

Our observations can be summarized as follows. As expected, the sharing of high-frequency information between the left and right channels was almost uniformly judged to be acceptable and not unpleasant. Of course, we could detect differences from the original material at the moment of switching for some passages in ongoing music. However, if one missed the actual moment of switching, it was often difficult or impossible to judge whether the original or transformed signals were being heard. All of our experiences were in line with our expectations based on results reported in the literature.

4. Conclusions

Because such coding represents a considerable reduction in the number of bits at only a minimal loss in stereo sound quality, it seems that it may be useful in other than the very best quality systems and presentations. That is, even though the transformed information can often be discriminated from the original, it is quite pleasant and has no monaurally detectable flaws. At present we view such spatially-based binaural bit rate reductions as a desirable feature that could be used to record digitally for a longer period of time on the same amount of tape. For those purposes, such coding could be incorporated as an option in tape recorders and other instruments, just like the various types of noise reduction that are now available in analog recorders.

In summary, our observations indicate that digital coding based on properties of binaural hearing could allow further meaningful reductions in bit rate in many practical environments.

Acknowledgement

Research reported in this and its companion paper was done while C. Trahiotis was a Van Houten Fellow at the Institute for Perception Research. Travel was made possible through NIH grant 5RO1 DC00234-13. Stimulating discussions with members of the Hearing and Speech group at IPO and of the Digital Signal Processing group of Philips Research are gratefully acknowledged. We thank the Medical School of the University of Connecticut for enabling the second author to collaborate on this research.

References

Bernstein, L. R. and Trahiotis, C. (1985). Lateralization of sinusoidally amplitude-modulated tones: Effects of spectral locus and temporal variation. *J. Acoust. Soc. Am.* **78**, 514–523.

Bernstein, L. R. and Trahiotis, C. (1992). Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies. *J. Acoust. Soc. Am.* **91**, 306–316.

Bilsen, F. A. and Raatgever, J. (1973). Spectral dominance in binaural lateralization. *Acustica* **28**, 131–132.

Blauert, J. (1983). *Spatial Hearing*. MIT, Cambridge, MA.

Blauert, J. and Divenyi, P. L. (1988). Spectral selectivity in binaural contralateral inhibition. *Acustica* **66**, 267–274.

Henning, G. B. (1974). Detectability of interaural delay in high-frequency complex waveforms. *J. Acoust. Soc. Am.* **55**, 84–90.

Houtsma, A. J. M., Trahiotis, C., Veldhuis, R. N. J., and van der Waal, R. (1996). Bit rate reduction and binaural masking release in digital coding of stereo sound. *Acustica* **this issue**.

McFadden, D. and Pasanen, E. G. (1976). Lateralization at high frequencies based on interaural time differences. *J. Acoust. Soc. Am.* **59**, 634–639.

Nuetzel, J. M. and Hafter, E. R. (1976). Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration. *J. Acoust. Soc. Am.* **60**, 1339–1346.

Shaw, E. A. G. and Teranishi, R. (1968). Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *J. Acoust. Soc. Am.* **44**, 240–249.

Trahiotis, C. and Bernstein, L. R. (1990). Detectability of interaural delays over select spectral regions: Effects of flanking noise. *J. Acoust. Soc. Am.* **87**, 810–813.

Wightman, F. L. and Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* **91**, 1648–1661.

Yost, W. A. (1981). Lateral position of sinusoids presented with interaural intensive and temporal differences. *J. Acoust. Soc. Am.* **70**, 397–409.

Zurek, P. M. (1985). Spectral dominance in sensitivity to interaural delay for broadband stimuli. *J. Acoust. Soc. Am.* **78**, S18.