Hybrid Auditory Masking Models

Matthew Raphael Flax, Jesse S. Jin
School of Computer Science & Engineering, University of New South Wales
flatmax@cse.unsw.edu.au

Abstract

Hybrid masking models are defined to be the separation of the perceptual and non-perceptual elements of auditory processing. This separation allows masking models to be defined for arbitrary cochlea. Using this separation, hybrid masking models may serve people with normal or impaired hearing, people of different ages and even the wider mammalian kingdom. As perceptual models of the cochlea improve, the potential exists for masking models to improve by directly applying more accurate cochlea descriptions.

1 Introduction

Masking models are used to highlight perceptually redundant information in auditory media streams. Techniques are currently frequency threshold based.

Masking models are traditionally based on frequency spreading functions (FSFs). FSFs are defined individually in the original masking model papers [1, 2, 3, 4]. The FSFs are defined by curve fitting biological data gathered from cochlea analysis. Another community who derive curves indirectly from the nature of the cochlea are the designers of PFBs. Both FIR [5, 6, 7, 8] and IIR [9, 10, 11, 12] PFBs have been developed. PFBs are constantly improving match with the biological nature of the cochlea and with progress of research, more improvements are expected. PFBs may also be constructed to match the tuning curves of hearing impaired people. Efforts have also begun to map the tuning curves of the wider mammalian kingdom [13]. A technique exists for the derivation of FSFs from the output of PFBs, it is defined in [8] and found to have constant computational complexity in [14].

2 Hybrid Masking Models

As outlined in Figure 1, hybrid models replace the frequency spreading functions with a PFB and a FSF operator. The FSF operator converts the output of a PFB to an FSF. A masking curve is derived from the FSFs using spreading function combination methods.

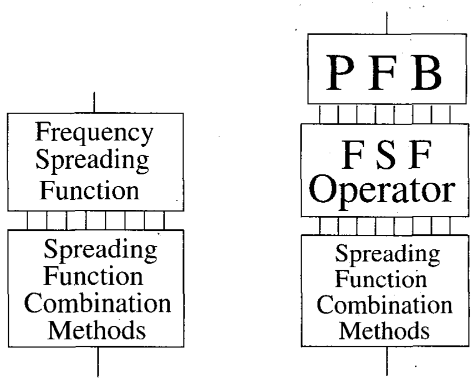

Figure 1: Traditional masking model (left) and hybrid masking model (right)

The FSFs are typically level dependent functions and have no ripple in their slope. FSFs have a peak at the masker frequency, both lower and higher frequencies ramp to lower magnitudes. Those FSFs which are level dependent, have a decrease in high frequency slope with rising masker level.

FSFs derived from PFBs on the other hand are also level dependent and ramp to lower magnitudes either side of the masker frequency. Again there is a rising high frequency slope with rising masker level. PFB based FSFs tend to have rippled slopes, which is in contrast to the more mechanical non rippled FSF slopes of the original masking models. This it is believed is a more natural model of the masking process in the cochlea.

Figure 2 depicts the FSFs of [1, 2, 3, 4, 15]. The level dependent FSFs show clearly a decrease in slope with increase in masker level with no ripple.

Figure 3 depicts an FSF derived from a PFB. Of note is the ripple in the slope and the decrease in slope with increase in masker level. FSFs derived from a PFB must be calculated for every frame of audio. Due to the constant computational complexity, this is simply a memory shift and may be accomplished in the background on capable digital signal processing microprocessors (background DMA).

By altering the PFB to match the nature of the mam-

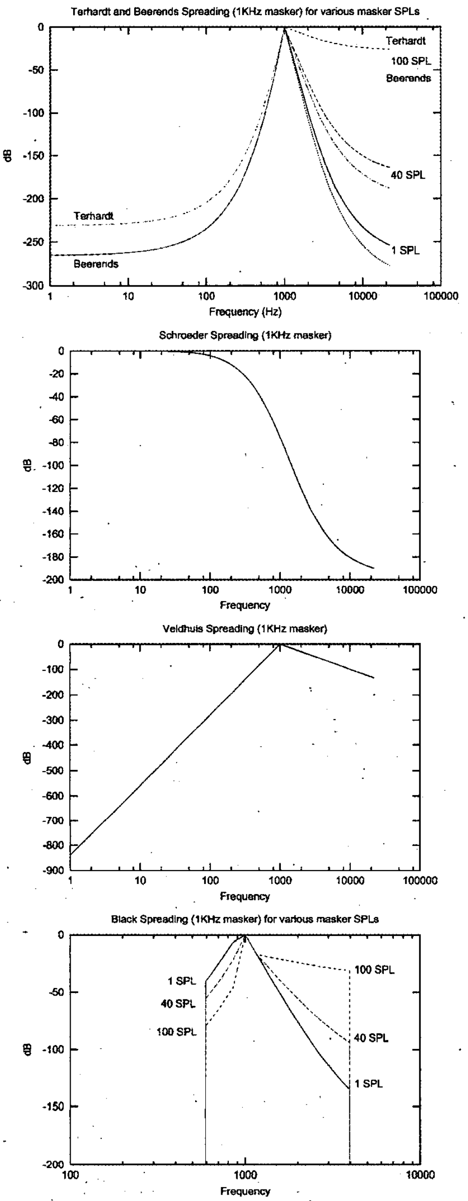

Figure 2: Terhardt and Beerends et.al. FSFs (top left), Schroeder et.al. FSF (top right), Veldhuis et.al. FSF (bottom. left) and Black et.al FSF (bottom. right).

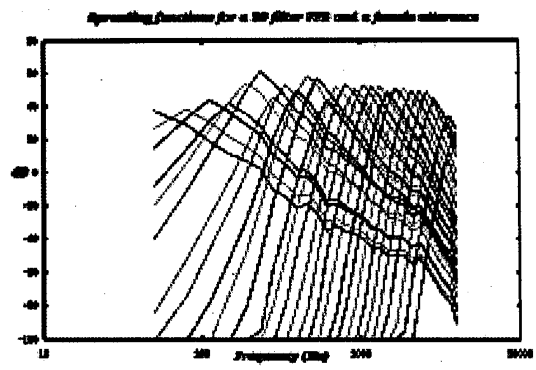

Figure 3: PFB based FSF for a frame of speech.

mal under analysis, it is possible to derive FSFs for an arbitrary mammal.

Spreading function combination methods may be roughly split into two stages. The first stage is the derivation of a spread masking threshold. The second stage is the derivation of the masking threshold, which is the output of a masking model.

A spread masking threshold is derived from the FSFs using operations of addition (Black et.al.), subtraction (Terhardt), multiplication (Veldhuis et.al.) and convolution (Schroeder et.al.).

In all cases the spread masking threshold is combined using addition. Schroeder et.al. alters this using a sensitivity function. [15] adapts the Terhardt masking model to include the masker's spread masking threshold. Again addition is used in this case, but in a compressed domain.

Terhardt and Beerends et.al. are the only combination methods which are level dependent and require no extra perceptual parameters. This maintains the aim of perceptual segregation between the derivation of the FSF and FSF combination stages.

The results of a Terhardt based hybrid masking model is depicted in Figure 4. The rough curve is the power spectrum. The smooth curve is the masking threshold. Power spectral peaks such as the formants are left unmasked, they reside above the masking threshold. Frequency bins which reside below the masking threshold are deemed redundant in the perception of the audio media. In general redundant frequencies tend to be located in power spectrum local minima.

The potential to improve the quality of traditional masking models using PFB based FSFs has been suggested in [14].

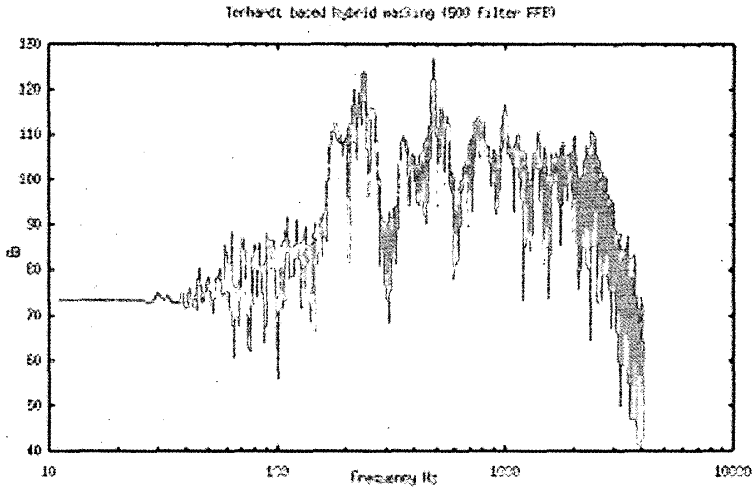

Figure 4: Masking curve derived using a Terhardt based hybrid masking model with a 500 filter PFB. The frame of speech is from a female utterance.

3 Conclusion

Hybrid masking models are defined as the separation of the perceptual masking model elements from the general elements. This is accomplished by using a FSF operator to derive a FSF directly from a PFB. By matching the PFB to the desired cochlea, an accurate masking model may be constructed for arbitrary cochlea. Accurate masking models are hence possible for people, young or old, with normal or impaired hearing, as well as other mammals.

As perceptual models of the cochlea improve, the potential exists for masking models to improve and hybrid masking models are a method for the application of such improvements.

4 Acknowledgements

This work is in co-ordination with a visual framework for audio analysis in the Visual Information Processing Laboratory, School of Computer Science & Engineering, University of New South Wales.

Thanks to Adjunct Professor H.W. Holmes and Dr. E.Ambikairajah from the Speech group, School of Electrical Engineering & Telecommunications, University of New South Wales for intellectual support and healthy discussion.

References

[1] "Calculating Virtual Pitch", Terhardt E., Hearing Research, Vol. 1, p. 155-182, 1979

[2] "Optimizing digital speech coders by exploiting masking properties of the human ear", Schroeder M.R., Atal B.S., Hall J.L., Journal of the Acoustical Society of America, Dec., Vol. 66(6), pp. 1647-1652, 1979

[3] "Subband coding of digital audio signals without loss of quality", Veldhuis R.N.J., Breeuwer M., Van Der Waal R., ICASSP, vol.3, pp 2009 -2012, 1989

[4] "Computationally efficient wavelet packet coding of wide-band stereo audio signals", Black M., Zeytinoglu M., ICASSP, Vol. 5, pp: 3075 -3078, 1995

[5] "Derivation of auditory filter shapes from notched-noise data", Glasberg B.R., Moore B.C.J. , Hearing Research, Vol. 47, pages 103-138, 1990

[6] "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns", Moore B.C.J., Glasberg B.R., Hearing Research, vol.28, no.2-3, pp.209-25, 1987

[7] "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness" Moore B.C.J., Glasberg B.R. and Baer T., Journal of the Audio Engineering Society, vol. 45, no. 4, April, pp.224-40, 1997

[8] "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns", Moore B.C.J., Glasberg B.R., The Journal of the Acoustical Society of America, Vol. 74, pp. 750-753, 1983

[9] "An Efficient implementation of the Patterson-Holdsworth auditory Filter Bank", Slaney M., Apple Computer Technical Report #35, 1993

[10] "An Optimal Auditory Filter", Irino T., IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New York, USA, p.198-201, 1995

[11] "A time-varying analysis/synthesis auditory filterbank using the gammachirp", Irino T., Unoki M., 1998 IEEE International Conference On Acoustics, Speech and Signal Processing (ICASSP98) , VI ,pp. 3653-3656, Seattle, Washington, May 12-15

[12] "An analysis/Synthesis Auditory FilterBank Based on an IIR Implementation of the Gammachirp", Irino T., Unoki M., JASJ accepted May, 1999

[13] "Comparing octaves, frequency ranges, and cochlea-map curvature across species", Greenwood D.D., Hearing Research, Vol. 94, Iss. 1-2, 1996

[14] "Improved Auditory Masking Models", Flax M., Ambikairajah E., and Holmes W.H. and Jin J.S., 8'th Australian International Conference on Speech Science & Technology, Dec. 2000

[15] "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation", Beerends J.G., Stemerdink J.A., Journal of the Audio Engineering Society, Vol. 40, No. 12, Dec., pp 963-978, 1992