A MODEL OF THE REPRESENTATION OF TIMBRE

SL Denham
Plymouth Institute of Neuroscience,
University of Plymouth, Plymouth, PL4 8AA, UK
sue@pion.ac.uk

ABSTRACT

Understanding how sounds are represented and recognised in cortex remains a challenging problem. Complex sounds may be distinguished perceptually in a number of ways, such as pitch, loudness, duration, and timbre. Of these aspects, timbre is in some ways the most significant since it is that essential quality of a sound that allows it to be recognised despite considerable variation in other features. Although timbre is not well defined, it seems that the spectral shape and temporal envelope of a sound, particularly at onset, are its most perceptually salient features [5].

Arguably, the currently most successful means of representing timbre are the Mel frequency cepstral coefficients used in automatic speech recognition. The cepstral coefficients are obtained from the discrete cosine transform of the log of the power spectrum, and efficiently represent the spectral shape of a sound segment. In general, the delta, and double delta coefficients which capture the local spectral change are also found to be useful. Cepstral processing is clearly very similar to the representation of visual signals by means of spatial frequency filtering [3]. There are however a number of examples of human perception which challenge this global representation scheme, including robustness to rather large temporal asynchronies [1], and narrow band filtering [4], and it is these aspects which this work addresses.

Inspired by the visual analogy of drifting gratings, experiments using wide band noises in which the spacing and movement of spectral peaks are parameterised, have shown that the spectrotemporal response fields (STRFs) of cells in primary auditory cortex encode local scale and asymmetry both in the spectral and temporal domain [2,6,7]. Therefore, it has been suggested that the convolution of tonotopically organised stimulus evoked activity with STRFs spanning a range of preferred spectral and temporal scales could result in a useful representation of the timbre of the stimulus [8].

A computational model of timbre representation based on these ideas is presented. The properties of the model are demonstrated through simulations and perceptual experiments.

Keywords: complex sounds, timbre, spectrotemporal receptive fields.

References

[1] Arai T, Greenberg S (1998). IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Seatle, pp. 933-936.

[2] Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001). *J Neurophysiol.* **85**:1220-1234.

[3] De Valois, RL, De Valois KK (1988). *Spatial Vision*. Oxford University Press.

[4] Healy E W, Warren R M (2003). *J. Acoust. Soc. Am.* **113**: 1676-1688.

[5] Iverson P (1995). *J. Exp. Psych.: Human Perception and Performance* **21**: 751-763.

[6] Miller LM, Escabi MA, Read HL, Schreiner CE (2002). *J Neurophysiol.* **87**: 516-527.

[7] Schreiner CE, Calhoun BM (1994). *Aud. Neurosci.* **1**: 39-61.

[8] Versnel H, Shamma SA (998). *J. Acoust. Soc. Am.* **103**:2502-2514.