Hearing Research 333 (2016) 8–24

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Perception of stochastic envelopes by normal-hearing and cochlear-implant listeners

Philip A. Gomersall [a, b, c, *], Richard E. Turner [d], David M. Baguley [b, c], John M. Deeks [a], Hedwig E. Gockel [a], Robert P. Carlyon [a]

[a] MRC Cognition & Brain Sciences Unit, 15 Chaucer Rd., Cambridge, England, United Kingdom
[b] Addenbrookes NHS Trust, Hills Rd., Cambridge, England, United Kingdom
[c] Anglia Ruskin University, East Rd, Cambridge, England, United Kingdom
[d] Dept. Engineering, University of Cambridge, Trumpington Street, Cambridge, England, United Kingdom

A R T I C L E I N F O

Article history:
Received 5 July 2015
Received in revised form
30 November 2015
Accepted 2 December 2015
Available online 17 December 2015

Keywords:
Sound textures
Envelope discrimination
Cochlear implants

A B S T R A C T

We assessed auditory sensitivity to three classes of temporal-envelope statistics (modulation depth, modulation rate, and comodulation) that are important for the perception of ‘sound textures’. The textures were generated by a probabilistic model that prescribes the temporal statistics of a selected number of modulation envelopes, superimposed onto noise carriers. Discrimination thresholds were measured for normal-hearing (NH) listeners and users of a MED-EL pulsar cochlear implant (CI), for separate manipulations of the average rate and modulation depth of the envelope in each frequency band of the stimulus, and of the co-modulation between bands. Normal-hearing (NH) listeners' discrimination of envelope rate was similar for baseline modulation rates of 5 and 34 Hz, and much poorer than previously reported for sinusoidally amplitude-modulated sounds. In contrast, discrimination of model parameters that controlled modulation depth was poorer at the lower baseline rate, consistent with the idea that, at the lower rate, subjects get fewer ‘looks’ at the relevant information when comparing stimuli differing in modulation depth. NH listeners could discriminate differences in co-modulation across bands; a multidimensional scaling study revealed that this was likely due to genuine across-frequency processing, rather than within-channel cues. CI users' discrimination performance was worse overall than for NH listeners, but showed a similar dependence on stimulus parameters.

1. Introduction

In recent years there has been an emerging interest in the perception of a class of stimuli known as “sound textures”. These include familiar environmental sounds such as fire, wind, rain and running water. Research from two groups (McDermott, 2009; Turner and Sahani, 2010; McDermott and Simoncelli, 2011; McDermott et al., 2013) indicates that listeners process and identify sound textures using information derived from the statistics of the envelopes in each frequency band, and from the relative amplitudes of, and correlations between them. Evidence for statistics-based perception comes from the fact that models that use a small number of stochastic parameters can generate sounds that can be readily identified as originating from different categories (McDermott, 2009; Turner and Sahani, 2010; McDermott and Simoncelli, 2011).

A large number of studies have investigated the detection of differences in stimulus envelope parameters by normal-hearing listeners. These include the detection of both first- and second-order amplitude modulation, the discrimination of changes in modulation rate, and the detection of differences in comodulation between pairs of narrowband carriers (Füllgrabe and Lorenzi, 2003; Grant et al., 1998; Lee, 1994; Moore and Emmerich, 1990; Richards, 1987). The vast majority of those studies have used deterministic modulators, and/or have imposed the modulation on either a single carrier or on a small number of carriers. This differs from many

Abbreviations: CI, Cochlear Implant; NH, Normal Hearing; MDS, Multidimensional Scaling; SMD, Statistical Modulation Depth; rms, root-mean-square; DL, Discrimination Limen; ANOVA, Analysis of Variance

* Corresponding author. MRC Cognition & Brain Sciences Unit, 15 Chaucer Rd., Cambridge, England, United Kingdom.

E-mail addresses: Philip.Gomersall@mrc-cbu.cam.ac.uk (P.A. Gomersall), ret26@cam.ac.uk (R.E. Turner), dmb29@cam.ac.uk (D.M. Baguley), John.Deeks@mrc-cbu.cam.ac.uk (J.M. Deeks), Hedwig.Gockel@mrc-cbu.cam.ac.uk (H.E. Gockel), Bob.Carlyon@mrc-cbu.cam.ac.uk (R.P. Carlyon).

http://dx.doi.org/10.1016/j.heares.2015.12.013

environmental sounds, which contain stochastic modulation in multiple contiguous frequency bands. In contrast, research on sound-texture perception has either required listeners to make categorisation judgements (McDermott and Simoncelli, 2011; Turner and Sahani, 2010), or, where discrimination experiments have been performed, have involved the simultaneous manipulation of more than one stimulus parameter (McDermott et al., 2013). In an attempt to bridge this gap by providing basic psychophysical measures of sensitivity to the envelope parameters important for the perception of sound textures, the present study measured listeners' sensitivity to differences in the individual parameters of one generative model, previously described by Turner and Sahani (2010). We then compare the results of both discrimination and multi-dimensional scaling experiments to the predictions of a simple auditory model, in order to gain insight into the auditory cues that listeners use when distinguishing between sound textures. For example, we find that increasing the co-modulation between the envelopes in different frequency bands increases the modulation depth both in individual auditory filter outputs and in the summed envelopes of all auditory filters that respond to the sound, but that listeners are either primarily or exclusively sensitive to the summed (across-channel) cue. We also demonstrate a formal mathematical relationship between the summed-channel cue and one in which each channel's envelope is correlated with every other, and discuss the implications of this relationship for the interpretation of results from other paradigms, such as the discrimination of correlation applied to widely-spaced narrowband carriers, and to the co-modulation detection difference (CDD: Cohen and Schubert, 1987; McFadden, 1987).

An understanding of the perception of the envelopes in each frequency region of broadband sounds, and of the correlation between those envelopes, is arguably of even more importance to the study of hearing by cochlear implant (CI) users. The speech-processing strategies implemented in all contemporary CIs discard the temporal fine structure in either all or the majority of frequency channels (McDermott et al., 1992; Wilson et al., 1991) leaving the listener to rely primarily on the channel envelopes to distinguish between the sounds encountered in everyday life. As is the case for normal acoustic hearing, research on envelope perception by CI users has typically used deterministic envelopes applied either to a single carrier or to a small number of carriers (Cazals et al., 1994; Chatterjee and Oba, 2004; Chatterjee and Robert, 2001; Chatterjee, 2003; Fu, 2002; Lorenzi et al., 1998, 1997; Richardson et al., 1998; Shannon, 1992; Won et al., 2011). We are not aware of any psychophysical investigation into sound-texture identification by CI users. We therefore repeated a subset of the experiments with CI listeners, as a first step towards an understanding of how this population perceive differences between sound textures.

Our investigation of stochastic envelope processing by CI users is also relevant to a long-term clinical goal of the study, related to the alleviation of tinnitus in CI users by the presentation of competing sounds ("sound therapy"). The environmental sounds used in this form of therapy correspond very closely to those used in the study of sound textures and that are effectively reproduced by generative models. Although the evidence for the overall effectiveness of sound therapy is equivocal (Hobson et al., 2010), exposure to low-level auditory textures can alleviate sleep handicap in tinnitus sufferers with acoustic hearing (Handscomb, 2006). Unfortunately, the benefit of such stimuli to CI users with tinnitus has not been established, and, in a preliminary stage of a previous study (Carlyon et al., 2010), CI users reported that the environmental sounds used to alleviate tinnitus in acoustic hearing sounded similar to each other. Indeed, many of the fast fluctuations present in these stimuli were degraded by the processing of the CI. By using a generative model, rather than simply selecting from a range of available sounds, it may be possible to produce stimuli that evoke a wider range of percepts, and, potentially, lead to a method where an individual CI patient can select an effective sound by varying a small number of parameters. Important pre-requisites to this goal include an understanding of listeners' sensitivity to changes in the parameters of a generative model, and of how these acoustic changes are processed by the electrically stimulated auditory system.

2. Generative model

The stimuli were created using a statistical audio texture model (Turner and Sahani, 2010). The model generates signals by summing a set of quickly varying band-limited noise carriers that have slowly-varying modulation imposed upon them by a set of stochastic modulators. The statistics of the signal are controlled by a set of parameters that intuitively correspond to the bandwidths and centre frequencies of the narrow-band noise carriers, the modulation-depth (or sparsity) and the rate of envelope fluctuations, and the dependencies between the modulators.

The full statistical audio texture model contains a large number of parameters that control fine details of the statistical structure of the generated sounds. In the original work, these parameters allowed the model to match the statistics of target textures, such as running water, crackling fire, or howling wind. Here the goal is rather different since we are interested in exploring perceptual sensitivity to the three main classes of envelope statistics, namely the modulation depth, modulation rate, and the dependencies between the modulators (comodulation). For this reason, we use a simpler version of the model with fewer parameters that is nevertheless able to produce a range of textural sounds that were subjectively identified by the authors to sound fairly naturalistic, and that result in similar range of statistics as natural sounds at the output of an auditory model (see Section 6.2, also examples available at http://www.mrc-cbu.cam.ac.uk/wp-content/uploads/2015/11/Wind.wav, http://www.mrc-cbu.cam.ac.uk/wp-content/uploads/2015/11/Water.wav, http://www.mrc-cbu.cam.ac.uk/wp-content/uploads/2015/11/Rain.wav). We focus on sensitivity to temporal fluctuations rather than to differences in spectral shape, which are already reasonably well accounted for by existing auditory models. Therefore all stimuli had the same long-term spectrum.

As the focus of this paper is on the perceptual sensitivity to the statistical properties of the modulators, the statistics of the narrow-band carriers were fixed throughout. Specifically, the carriers were produced by filtering Gaussian noise through band-limited filters (defined by second order autoregressive functions), with centre frequencies equally spread in log space between 500 and 4000 Hz. The bandwidths and centre frequencies used for the 20-carrier stimuli employed in most of our experiments are shown in Table 1a, and the filters imposed on each carrier are plotted in Fig. 1a.

The modulators were generated in three stages. The first stage produces low-pass Gaussian noise, which is both smooth and slowly varying, and the second stage converts this into a positive modulator. The third stage introduces dependencies between the modulators in different frequency channels.

In more detail, in the first stage (Fig. 2a) the low-pass Gaussian noise is produced using a filter that has a Gaussian shape,

$$W(f) = Ae^{-f^2/(2f_o^2)} \quad (1)$$

The filter is centred on 0 Hz and has a parameter f_0 that controls the width of the low-pass filter and therefore the rate of fluctuation