Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 57 (2014) 114–125

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Enhancement of speech perception in noise by periodicity processing: A neurobiological model and signal processing algorithm

Neil M. McLachlan [a,*], David B. Grayden [b,c]

[a] Melbourne School of Psychological Sciences, University of Melbourne, Victoria 3010, Australia
[b] NeuroEngineering Laboratory, Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia
[c] Centre for Neural Engineering, University of Melbourne, Victoria 3010, Australia

Received 31 July 2012; received in revised form 16 September 2013; accepted 16 September 2013
Available online 25 September 2013

Abstract

The perceived loudness of sound increases with its tonality or periodicity, and the pitch strength of tones are linearly proportional to their sound pressure level. These observations suggest a fundamental relationship between pitch strength and loudness. This relationship may be explained by the superimposition of inputs to inferior colliculus neurons from cochlear nucleus chopper cells and phase locked spike trains from the lateral lemniscus. The regularity of chopper cell outputs increases for stimuli with periodicity at the same frequency as their intrinsic chopping rate. So inputs to inferior colliculus cells become synchronized for periodic stimuli, leading to increased likelihood that they will fire and increased salience of periodic signal components at the characteristic frequency of the inferior colliculus cell. A computer algorithm to enhance speech in noise was based on this model. The periodicity of the outputs of a Gammatone filter bank after each sound onset was determined by first sampling each filter channel at a range of typical chopper cell frequencies and then passing these amplitudes through a step function to simulate the firing of coincidence detecting neurons in the inferior colliculus. Filter channel amplification was based on the maximum accumulated spike count after each onset, resulting in increased amplitudes for filter channels with greater periodicity. The speech intelligibility of stimuli in noise was not changed when the algorithm was used to remove around 14 dB of noise from stimuli with signal–noise ratios of around 0 dB. This mechanism is a likely candidate for enhancing speech recognition in noise, and raises the proposition that pitch itself is an epiphenomenon that evolved from neural mechanisms that boost the hearing sensitivity of animals to vocalizations.

Keywords: Neurocognitive; Model; Periodicity; Segregation; Algorithm; Speech

1. Introduction

Pitch height is a psychological dimension related to the frequency or period of sounds. Pitch strength may be defined as the certainty of a pitch height (Beerends and Houtsma, 1989; Rakowski, 1996), and has been measured by pairwise comparisons between stimuli of various tonal characteristics where loudness was held constant (Fastl and Stoll, 1979). Pitch strength and loudness judgments are not independent. For pure tones, pitch strength increases linearly with sound pressure level over a range of 60 dB (Fastl, 1989), and both loudness and pitch strength increase with stimulus duration up to around 200 ms (Fastl, 1989; Viemeister and Wakefield, 1991). In contrast to the variability of pitch strength with loudness, the pitch height of a 1 kHz pure tone varies by less than 1% for sound pressure levels between 40 and 80 dB SPL (Zwicker and Fastl, 1999).

Further evidence for a relationship between pitch strength and loudness include the findings that pure tones produce much greater loudness than noise bands (Fastl and Stoll, 1979), such that the sound pressure level of octave band noise with an embedded pure tone is 10 dB less than octave band noise alone with matched loudness

---

* Corresponding author. Tel.: +61 3 8344 6290.
E-mail addresses: mcln@unimelb.edu.au (N.M. McLachlan), grayden@unimelb.edu.au (D.B. Grayden).

0167-6393/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2013.09.007

(Kryter and Pearsons, 1965). Soeta et al. (2004) and Kidd et al. (1989) showed that this effect was preserved even if the width of the noise band was less than one critical bandwidth. Furthermore, Soeta et al. (2007) showed that increases in the pitch strength of iterated rippled noise increased perceived loudness, thereby suggesting that waveform periodicity subserves the relationship between loudness and pitch strength. Furthermore, in studies of auditory evoked potentials, an N1 peak and sustained negative potentials at latencies between 400 and 800 ms were associated with periodicity processing by appending iterated rippled noise to bandpass noise with the same spectral profile (Krumbholz et al., 2003; Seither-Preisler et al., 2006). Increases in the amplitudes of the N1 peak and the longer latency sustained negative potentials are likely associated with enhanced spike rates, and were proportional to the iterated rippled noise pitch strength (Krumbholz et al., 2003; Seither-Preisler et al., 2006).

Most research on the neurobiological function of periodicity processing has focused on the importance of periodicity processing for pitch (Cariani, 2001; Dau et al., 1997; Dicke et al., 2007; Guérin et al., 2006; McLachlan, 2009; Meddis and O'Mard, 2006; Nelson and Carney, 2004). However, the ability to hear tones in noise has important implications for speech intelligibility under noisy conditions (Plomp, 1994). In support of this, the detection of common periodicity has been successfully used to segment parts of noisy speech signals that contain vowel sounds (Hu and Wang, 2008), and the use of periodicity-based features was found to enhance the automatic speech recognition rates of voiced components of speech in the presence of noise (Ishizuka and Nakatani, 2006). Other researchers have considered the role of periodicity processing in segregating vowels by autocorrelation and similar algorithms such as recurrent neural networks (Assmann and Summerfield, 1990; Cariani, 2001; de Cheveigné and Kawahara, 1999; Meddis and Hewitt, 1992), and in the formation of stable auditory images of periodic sounds through the cross-channel temporal alignment of the maxima of auditory nerve spike rates (Patterson et al., 1995).

While most techniques that attempt to improve speech intelligibility in noise seek to remove the energy associated with noise by estimating its spectrum (Yoo et al., 2007), a small number of studies have sought to enhance spectral components associated with vowels (Cheng and O'Shaughnessy, 1991; Turicchia and Sarpeshkar, 2005). However, many studies have shown that transient voiced information is very important for speech intelligibility (Hazan and Simpson, 1998; Howell and Rosen, 1983; Richardson et al., 2004; Strange et al., 1983), and algorithms that enhance speech transients given prior knowledge of the speech content have been successful in increasing recognition rates for speech in noise (Yoo et al., 2007). Such algorithms cannot recover non-voiced components of speech such as fricatives, and these components must be deduced from the voiced information when masked by noise.

This paper introduces a neurobiologically inspired model of periodicity processing (Periodic Sampling Model, PSM) that may enhance the loudness of transient formants in speech relative to noise without the long signal durations required to extract fundamental frequency information. Overall, the PSM proposes that inferior colliculus spike rates integrated over multiple waveform periods after each sound onset will be greater at specific sub-bands (or best modulation frequencies, BMF) within auditory frequency channels that contain periodic auditory nerve (AN) spike rate modulations. The specific loudness associated with amplitude modulated filter channels will be greater, in keeping with the relationship between pitch strength and loudness observed by Fastl (1989). The second part of the paper reports a pilot study in which an algorithm based on the PSM is used to enhance periodic components of speech signals embedded in pink noise and speech babble, and speech intelligibility of processed speech stimuli were compared with unprocessed stimuli for normal hearing participants.

2. The Periodic Sampling Model (PSM)

2.1. Neurobiological motivation

This model focuses on brainstem networks that involve the cochlear nucleus (CN), inferior colliculus (IC), and the ventral nucleus of the lateral lemniscus (VLL). The cell types that project to the IC from the CN comprise multipolar (otherwise known as stellate or chopper), fusiform, octopus, and giant cells (Cant and Benson, 2003). Sustained chopper cells exhibit highly regular firing patterns when excited near their characteristic frequency (CF), with a frequency of firing known as the cell's best modulation frequency (BMF; Wiegrebe and Winter, 2001) that is independent of the stimulus frequency (Blackburn and Sachs, 1991).

Octopus, bushy and chopper cells in the CN often synapse to the VLL, which then projects to the IC. Octopus cells are innervated by a wide range of AN fibers and so integrate waveform information from many critical bands (around 2–3 octaves in mice, Oertel et al., 2000). They have very short latencies, and a single cell can phase-lock to acoustic waveforms at frequencies up to around 800 Hz (Ferragamo and Oertel, 2002). Bushy cells have similar temporal response properties to Octopus cells, but are innervated by only a few AN fibers (Oertel, 1997). Neurons in the VLL generally have short temporal integration times (around 5 ms), broad tuning, and exhibit a wide range of response types (Covey and Casseday, 1991). The combined excitatory and inhibitory input to the central nucleus of the IC from the VLL is likely to contain precise waveform timing information that is superimposed on direct inputs from sustained chopper cells in the CN (see Fig. 1; Riquelme et al., 2001; Schofield, 2005; Pickles, 2008).

Neurons in the central nucleus of the IC are arranged in a three-dimensional map with fine and coarse frequency