University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Erfanian Saeedi, N;Blamey, PJ;Burkitt, AN;Grayden, DB

Title:
Application of a pitch perception model to investigate the effect of stimulation field spread on the pitch ranking abilities of cochlear implant recipients

Date:
2014-10-01

Citation:
Erfanian Saeedi, N., Blamey, P. J., Burkitt, A. N. & Grayden, D. B. (2014). Application of a pitch perception model to investigate the effect of stimulation field spread on the pitch ranking abilities of cochlear implant recipients. Hearing Research, 316, pp.129-137. https://doi.org/10.1016/j.heares.2014.08.006.

Persistent Link:
https://hdl.handle.net/11343/43925

Accepted Manuscript



Application of a pitch perception model to investigate the effect of stimulation field spread on the pitch ranking abilities of cochlear implant recipients

Nafise Erfanian Saeedi, Peter J. Blamey, Anthony N. Burkitt, David B. Grayden

PII: S0378-5955(14)00144-0

DOI: 10.1016/j.heares.2014.08.006

Reference: HEARES 6840

To appear in: *Hearing Research*

Received Date: 5 March 2014

Revised Date: 14 August 2014

Accepted Date: 19 August 2014

Please cite this article as: Saeedi, N.E., Blamey, P.J., Burkitt, A.N., Grayden, D.B., Application of a pitch perception model to investigate the effect of stimulation field spread on the pitch ranking abilities of cochlear implant recipients, Hearing Research (2014), doi: 10.1016/j.heares.2014.08.006.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

***Highlights (for review)**

ACCEPTED MANUSCRIPT

- A computational model of auditory perception is developed
- The effect of stimulation field spread on simulated CI pitch perception is examined
- Sharp stimulation fields lead to better pitch ranking scores

ACCEPTED MANUSCRIPT

Title page
Click here to view linked References

ACCEPTED MANUSCRIPT

Application of a pitch perception model to investigate the effect of stimulation field spread on the pitch ranking abilities of cochlear implant recipients

Nafise Erfanian Saeedi[a,b*], Peter J. Blamey[c,d], Anthony N. Burkitt[a,b,c], David B. Grayden[a,b,c]

[a]NeuroEngineering Laboratory, Dept. of Electrical & Electronic Engineering, University of Melbourne, Australia,

[b]Centre for Neural Engineering, University of Melbourne, Australia,

[c]The Bionics Institute, East Melbourne, Australia,

[d]Dept. of Medical Bionics, University of Melbourne, Australia

*Corresponding author: Centre for Neural Engineering, The University of Melbourne, 203 Bouverie St., Carlton, 3053, Melbourne, Victoria, Australia. Tel: +61424650803

E-mail address: ninaes@student.unimelb.edu.au (N. Erfanian Saeedi)

ACCEPTED MANUSCRIPT

*Manuscript
Click here to view linked References

ACCEPTED MANUSCRIPT

Application of a pitch perception model to investigate the effect of stimulation field spread on the pitch ranking abilities of cochlear implant recipients

Abstract

Although many cochlear implant (CI) recipients perceive speech very well in favorable conditions, they still have difficulty with music, speech in noisy environments, and tonal languages. Studies show that CI users' performance in these tasks are correlated with their ability to perceive pitch. The spread of stimulation field from the electrodes to the auditory nerve is one of the factors affecting performance. This study proposes a model of auditory perception to predict the performance of CI users in pitch ranking tasks using an existing sound processing scheme. The model is then used as a platform to investigate the effect of stimulation field spread on performance.

Keywords: Computational auditory perception model; Pitch ranking; Cochlear implant;

1 Introduction

Cochlear implant (CI) technology has improved significantly during the past three decades. Nowadays, many CI recipients perceive speech very well in quiet conditions. However, implantees still have difficulty with speech in noisy environments and perceiving tones in tonal languages. They also report less satisfaction with music perception compared to normal hearing people. Pitch is important for performing these tasks. Humans use pitch cues to separate target speech from background noise (Qin and Oxenham 2003). Lexical tone perception relies on pitch contour detection (Wang et al. 2011). Melodies can be considered a series of pitch changes and consequently recognizing them requires pitch perception (Galvin et al. 2007; Kang et al. 2009). It is therefore

unsurprising that CI users' performance in the above tasks and their ability to perceive pitch are correlated (Gfeller et al. 2007).

There is large variability in performing pitch-related tasks among CI patients, due to factors such as duration of implant use (Gfeller et al. 2007), residual hearing (Looi and Radford 2011), musical training before deafness (McDermott and McKay 1997; Sucher and McDermott 2007; Maarefvand et al. 2013), post-implant training (Gfeller et al. 2000), and implant properties such as stimulation rate and sound processing strategy (McDermott and McKay 1994; Vandali et al. 2005). Investigating the effect of the above factors on CI recipients' performance and identifying other possible factors requires an insight into how pitch is perceived in a normal and in an implanted ear.

There are two principal theories that explain how normal hearing (NH) listeners perceive pitch (Rose et al. 1967; Loeb et al. 1983; Shamma 1985; Plack et al. 2005). The *place* or *rate-place* theory assumes that pitch is coded by the place on the basilar membrane that has the maximum excitation or neuronal firing rate compared to other locations. The second theory proposes that pitch is a *temporal* code based on the exact timing of the action potentials. That is to say, neurons tend to fire in synchrony with the acoustic stimulus and the subsequent action potentials are separated by intervals equal to the period of the stimulus. Although the two theories seem fundamentally different, it is possible that the central auditory system integrates both cues to perceive pitch. For example, in a study by Oxenham et al. (2004), it was demonstrated that perceiving the pitch of a complex tone required consistent place and temporal pitch information, and that modifying the place information inhibited pitch perception, even if the temporal information was still correct. This suggests that eliciting a precise pitch percept requires both pitch cues, at least for some types of stimuli.

CI users perceive pitch by applying similar pitch cues. Nelson et al. (1995) reported that CI users would reliably judge more basal/apical electrodes as having higher/lower pitches. This outcome is

consistent with the place code of pitch. Tong et al. (1983) showed that a CI recipient's sensation of pitch was monotonic with the rate of stimulation applied to a single electrode. That is to say, higher stimulation rates led to sensation of higher pitches. This is in line with the temporal pitch code in acoustic hearing. They also suggested that the interaction between rate and place information was not significant in the context of speech perception, and that these two perceptual components are in fact orthogonal. The implication of this outcome for pitch perception was later supported by a study by McKay et al. (2000), where the place and temporal pitch cues were found to be processed independently by CI users in a (stimulation) rate and/or (electrode) place discrimination task. However, a more recent study by Luo et al. (2012) showed that CI users would recognize pitch direction changes significantly better with combined place and temporal pitch cues, given that the two cues were consistent (i.e., both cues would elicit a sensation of pitch change in the same direction).

Although implant users are able to extract pitch from cues similar to NH listeners, in electric hearing, these cues are affected by factors that can lead to deterioration in pitch perception. For example, place cues are affected by the limited number of electrodes and the depth of electrode insertion (Blamey et al. 1996). These factors distort the frequency-to-place mapping in multichannel CI users (e.g., Schatzer et al. 2014). Limited number of electrodes in an electrode array (e.g., 22 in Nucleus® device from Cochlear™) results in a range of frequencies being assigned to a single electrode, thereby reducing the frequency resolution and limiting the place cues. Frequency-to-place mapping is more distorted in shorter arrays due to a wide frequency range being mapped to a short cochlear length. Shallow insertion of the electrode array also results in loss of low frequency content (including speech pitch), because the apical regions of the cochlea do not receive sufficient stimulation (Başkent and Shannon 2005).

Temporal cues, on the other hand, are influenced by the existence of an upper limit of rate pitch of about 300-500 pps (Tong et al. 1983; Carlyon and Deeks 2002; Zeng 2002; Macherey et al. 2011;

ACCEPTED MANUSCRIPT

Schatzer et al. 2014). The latter refers to a maximum rate of stimulation, above which, no distinctive change in pitch occurs.

Investigating such limiting factors, and endeavoring to provide solutions for them, have been the subject of many studies in CI research. A brief review of a few studies that are more closely related to the topic of this report – effect of stimulation field spread on CI users' pitch ranking abilities – is provided in the following.

In multichannel stimulation, multiple electrodes are typically activated sequentially, therefore, a population of neurons may receive stimulation from multiple electrodes. Considering the place theory of pitch perception, electrode interactions reduce the spectral contrast and, therefore, should disrupt the CI users' pitch perception. For example, Townshend et al. (1987) showed that the spread of electrical excitation from the stimulating electrode through the cochlear duct influenced the pitch ranking performance of CI users. Fredelake and Hohmann (2012) demonstrated that increasing the spatial spread of the stimulation field had a significant adverse impact on the speech recognition of a simulated model of an implanted auditory system. Also, through vocoder-based simulations, Stafford et al. (2013) showed that noisy speech intelligibility increased if highly focused stimulation was used instead of conventional monopolar stimulation.

In this regard, various current shaping techniques have been introduced to address the issue of current spread, aiming to improve auditory perception in cochlear implantees. Such techniques tend to excite narrow neural regions by providing more localized and targeted stimulation fields. Generating distinct excitation sites may in fact improve the place cues and generate more precise pitch percepts that are unattainable with broad stimulation fields.

Van Compernolle (1987) proposed a deconvolution method to generate stimulation fields with the desired shape. This approach was used to sharpen place cues (Townshend et al. 1987), which

improved pitch discrimination abilities of CI users. Van den Honert et al. (2007) reported the successful application of phased array stimulation to produce focused stimulation fields in CI users. Landsberger and Srinivasan (2009) used quadrupolar stimulation to localize the stimulation current for better spectral resolution. The effectiveness of quadrupolar stimulation in sharpening the excitation patterns in CI recipients has been demonstrated by Srinivasan et al. (2010). Simulation and field measurements from guinea pigs' cochleae, performed by Jolly et al. (1996), confirmed that partial tripolar stimulation provided focused stimulation fields that might enhance place pitch cues. Landsberger et al. (2012) used partial tripolar stimulation as a current focusing technique to reduce the spread of excitation in CI users. Although stimulation current was successfully focused using this method, a minimum level of focusing was required for targeted neural regions to be localized as well. Srinivasan et al. (2013) used a similar technique to enhance CI users' speech perception in noise. The implantees performed better with partial tripolar stimulation than with monopolar stimulation. In more recent studies, Fielden et al. (2013) and Saoji et al. (2013) were able to create distinctive neural excitation areas using tripolar stimulation and phantom arrays, respectively.

Changing the properties of the stimulation pulses also affects the subsequent neural excitation patterns. For instance, Macherey et al. (2011) reported that customized asymmetric stimulation pulses would generate focused excitation toward the apical cochlear regions. This led to lower pitch percepts, not achievable with symmetric biphasic pulses. Neural excitation patterns can also be shaped by activating nearby electrodes sequentially to generate pitch percepts intermediate to those stemming from each of the electrodes in isolation (McDermott and McKay 1994; Kwon and Van den Honert 2006; Landsberger and Galvin 2011). This stimulation method creates extra perceptual channels, also known as "virtual channels". Virtual channels, therefore, increase the number of place pitch steps beyond the number of physical electrodes and may improve CI users' spectral resolution (Donaldson et al. 2005).

Although the aforementioned methods have been able to provide some functional gain, though of different degrees, some other studies report no benefit from shaped stimulation currents and narrow excitation patterns in the context of auditory perception. For example, McKay et al. (1996) found that even with extensive overlapping neural excitation patterns in a dual-electrode stimulation experiment, CI users would reliably report consistent percepts. In other words, the perceptual spaces were robust against changing degrees of the overlaps between the two stimulated neural populations. This implied that in sound processing strategies that used multiple adjacent electrodes to represent vowel formants, creating non-overlapping excitation patterns would not necessarily lead to better vowel discrimination.

The goal of the current study was to develop a computational model of pitch perception that received as input the activity of the auditory nerve and would accurately reproduce the psychophysical results in pitch ranking tasks. It was hypothesized that such a model could be used as a test platform to investigate the possible effect of stimulation parameters, in particular the spread of the stimulation field, on the pitch performance of CI recipients.

The function of the pitch perception model developed in this paper was based on the place pitch cues. Models that apply the place theory generally use a bank of filters to account for the tonotopic behavior of the cochlea. The pitch of the stimuli is then determined by comparing the spectral profile of the sound with predefined pitch templates (Shamma and Klein 2000) or harmonic sieves (Cohen et al. 1995). Temporal pitch models, on the other hand, derive pitch from the timing information of action potentials generated at the site of the auditory neurons. This is usually done by applying self-similarity measures to the responses of multiple neurons (Licklider 1951; Slaney and Lyon 1993) to find the phase-locking frequency, which is then interpreted as pitch. Combined place and temporal pitch information are also used in some pitch perception models (Loeb et al. 1983; Shamma 1985;

ACCEPTED MANUSCRIPT

Voutsas et al. 2005). These models assume that there is a central pitch processing center that is able to integrate both pitch cues and extract a reliable pitch.

In electric hearing, temporal pitch cues alone provide sufficient information to CI users to detect differences as small as one semitone in the stimulation rate (at rates <200 pps) applied to a single electrode (e.g., Blamey et al. 1984). Place cues, may not provide sufficient information to represent low pitch stimuli (<500 Hz), mainly due to more distortion that occures in the frequency-to-place mapping at low frequency registers (Schatzer et al. 2014). However, there is evidence that place cues alone contain adequate information to discriminate pitch *directions* at intervals as small as one semitone (Maarefvand et al. 2013).

In the pitch perception model used in this study, place pitch is considered an independent pitch cue that presumably contains sufficient information to perform the pitch ranking tasks. The validity of the model is verified by comparing it to the performance of normal hearing and CI listeners. The model is then used to investigate how generating more or less localized stimulation fields translates into pitch ranking capabilities in CI recipients.

This paper is organized as follows: Section 2 describes the database of stimuli used to train and evaluate the model, and explains how normal hearing and hearing with CIs are modelled. Descriptions of the pitch ranking model and the ranking tasks are also provided in this section. Model results are presented in Section 3 and discussed in Section 4. Conclusions are provided in Section 5.

ACCEPTED MANUSCRIPT

2 Methods

2.1 Data

Because training a computational model on pitch categories requires several samples of each category with a precise and reliable pitch, synthetic vowel-like stimuli were used in this study. Synthetic vowels with constant pitch and formant frequencies minimize the real-world acoustic effects such as pitch variations and formant movements in higher pitch stimuli. Also, sustained or sung vowels serve as stimuli with acoustic information on both speech and music, and are desirable for the purpose of this study. To generate the required data, the cascade branch of the KLATT speech synthesizer (Klatt 1980) was used. C-code implementation of this synthesizer (version 3.4) is available at: http://www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/speech/systems/klatt (last accessed on 3 June 2014).

Characteristics of 16 types of vowels, such as the first three formant frequencies and their associated bandwidths, were selected using the reported values of Hawks and Miller (1995). Depending on whether the formant bandwidths were measured or predicted, there were three different bandwidths for each of the three formants reported for each vowel. To create more variety in the data, all three bandwidth values were used. The fundamental frequency (F0) of each vowel took a value of 98 Hz, 139 Hz, 196 Hz, or 277 Hz. The pitch notations for these frequencies are G2, C#3, G3, and C#4, respectively. Fig. 1 presents temporal and spectral representation (using LPC analysis) of an example of the synthesized data with a fundamental frequency of 139 Hz and the first three formant locations associated with the vowel /a/ (viz. 710 Hz, 1150 Hz, and 2700 Hz, respectively). The three variations of the bandwidths are indicated with different line styles in the bottom panel.

The selected notes were 0.5 octave (Oct.) apart and provided a representative range in psychophysical studies investigating the performance of CI users in pitch ranking tasks (Vandali et al. 2005; Wang et al. 2011). It has been shown that for such relatively large intervals between the F0 pairs, CI users were able to detect the differences between the pairs by using only the place pitch information (Laneau et al. 2004). This was particularly important because the pitch perception model described in Section 2.4 relied entirely on place pitch cues. The final database consisted of 192 (16 vowels × 3 bandwidths × 4 F0s) sung vowels, each with a duration of 500 ms, intensity of 60 dB SPL, and sampled at 16 kHz rate.

Fig. 1: Temporal (top) and spectral (bottom) representations of a C#3 /a/ synthetic sung vowel with F0=139 Hz, F1=710 Hz, F2=1150 Hz and F3=2700 Hz. Amplitude is expressed in arbitrary units. Variations of bandwidths are shown with different line styles.

2.2 Modelling normal hearing

The role of a normal or acoustic hearing model is to simulate the activity of the auditory nerve in response to any given acoustic stimulus. The different models suggested for acoustic hearing almost always follow the same structure, including a bank of band-pass filters representing the tonotopic behavior of the basilar membrane (BM) (Meddis et al. 2010). The model proposed by Zilany and Bruce (2006) is one of the most comprehensive models for the auditory periphery that is capable of describing a wide range of phenomena from the middle ear to the auditory nerve. The model begins with a middle ear filter, followed by wide-band and narrow-band filters in parallel that together represent a single position along the basilar membrane with an assigned characteristic frequency (CF). The outputs of the parallel filters pass through a non-linear model of an inner hair cell (IHC) followed by a synapse model. The neuronal responses from this model are consistent with physiological data recorded from cats. In this study, a slightly modified version of the model was used to create

responses similar to the human auditory periphery. Modifications were made to the middle ear filter, as described by Pascal et al. (1998), and the BM tuning to match human data, as reported by Shera et al. (2002).

The center frequencies of the cochlear filters of the acoustic model were calculated using the Greenwood function (Greenwood 1990),

$$f_{\mathbf{c}} = 165.4 \times \left(10^{2.1 \times \frac{d}{34}} - 1\right),$$

where $f_{\mathbf{c}}$ was the CF of the IHC and also the center frequency of a cochlear filter in the acoustic model and d indicated the position of the IHC along the BM (measured from the apex of the cochlea in mm) and was incremented in 0.1 mm steps to generate different CFs. The cochlea was assumed to have a total length of 34 mm; however, only the cochlear positions in the range 3-22.9 mm were taken into account. Subsequently, there were 200 channels in the acoustic model that generated the output components associated with frequencies from 88 Hz up to nearly 4 kHz. This range contained most of the relevant information of speech. The outputs of the IHCs were interpreted as the level of neural activity or firing rates. Such neural activities, when sorted by location (or CF) and presented in time, create the so-called spatio-temporal response patterns (Shamma 1985). The spatio-temporal response patterns were calculated for each of the vowels. Fig. 2 shows the response pattern for the acoustic stimulus shown in Fig. 1.

The neural activities were averaged over a 100 ms interval in the central portion of the vowel for each channel. The average activities were then used as the input to the pitch perception model. Electrophysiological recordings show that the average neural activities contain information about the most prominent characteristics of the vowel-like stimuli, such as the harmonics of the fundamental

frequency and formant locations (Delgutte and Kiang 1984; Miller and Sachs 1984). Section 2.4 describes how the average neural activities lead to pitch percepts.

Fig. 2: NH spatio-temporal neural activity of the auditory nerve in response to the stimulus shown in Fig.1. Plot of action potentials, with fiber CFs shown along the ordinate (left), and a histogram of the average activity over the 100-ms interval (right). Dark areas indicate more neural activity.

2.3 Modelling hearing with the CI

Similar to the normal hearing model, a CI hearing model also computed the spatio-temporal response patterns. Instead of acoustic stimuli, the input to the model was a sequence of electrical stimulation pulses generated by a sound processor unit and delivered to the electrodes in the cochlea.

The Nucleus® MATLAB® Toolbox (NMT) provides a convenient platform to simulate most of the commercial sound processors efficiently (Swanson 2008). In this study, the NMT was used to generate electrical stimulation, applying the ACE (Advanced Combination Encoder) strategy with its standard configuration. ACE was implemented on a simulated Nucleus® (Cochlear™) device, with an electrode array of 17 cm, including 22 electrodes spaced at 0.75 mm. The strategy selected 12 maxima and activated 12 electrodes out of the 22 available electrodes in each time frame. The stimulation rate was 900 pps per electrode.

The structure of the cochlea and the distribution of the 200 auditory neurons were the same as described in the previous section. However, the depth of electrode insertion varies among implantees (see for example Ketten et al. (1998)) and the Greenwood frequency-position cannot be presumed for electrical hearing (Blamey et al. 1996; Schatzer et al. 2014). Thus, the auditory neurons are represented by their order (1-200, with one being the most apical) and not their CFs.

Fig. 3 shows the simulated stimulation patterns – also known as electrodograms – for the ACE strategy in response to the sound stimulus presented in Fig. 1.

Spatio-temporal response patterns for electric hearing were computed using a model developed by Cohen (2009-b). This model describes the peripheral neural responses to electrical stimulation using the elements proposed by Bruce et al. (1999). However, the basic model is re-formulated to incorporate patients' individual parameters such as electrode distance from the modiolus and neural fiber thresholds. The attenuation of the stimulation field initiated at the site of each activated electrode is described by effective stimulation field (ESF) functions. These functions show the extent to which an electrode's field can propagate and produce neural excitation along the cochlea. ESF functions are derived from CI recipients' evoked compound action potential (ECAP) recordings. In order to match the simulated model outputs to the patients' ESF functions, a scaling factor (SF) is introduced (Cohen 2009-a). This parameter describes the width of the spread function and takes values in the range 1-3, with 1 corresponding to the narrowest field spread and 3 the broadest.

Simulated neural activities were then averaged over a 100 ms interval, similar to Section 2.2. For the simulations in this paper, the electric model properties were the same as the reported values of Cohen (2009-b). The benchmark value for SF was set to the patients' average SF (2.015) estimated for the CI users of the Nucleus® device with straight array. For simplicity, this value is shown as 2 in the remainder of this report. However, in order to investigate the effect of having different degrees of neural excitation spread, as a result of delivering more or less localized stimulation currents to the tissue, other SF values in the range [1-3] were also sampled and employed. Moreover, since the stimulation rate incorporated by the sound processor was higher than 250 pps, refractory behavior was taken into account. In the model, this was done by temporarily increasing the firing threshold of the auditory neuron immediately following an action potential (Cohen 2009-c).

Fig. 4 demonstrates how the electrical model of hearing responds to the stimulation patterns shown in Fig. 3. Similar to the acoustical model, these responses are represented with spatio-temporal maps.

Fig. 3: Simulated electrical stimulation pattern (electrodogram) for the ACE strategy, using the sound stimulus shown in Fig. 1. Time is shown along the abscissa and electrode numbers along the ordinate, with electrode 22 being the most apical. Each vertical bar represents a stimulation pulse with a height associated with the pulse amplitude.

Fig. 4: Spatio-temporal maps for the CI model using the ACE strategy in response to the stimulation pattern presented in Fig. 3. Neuron numbers from 1 to 200 are shown along the ordinate. The average activity over the 100-ms interval is shown on the right side of the map. Dark areas indicate more neural activity.

2.4 Pitch perception model

Fig. 5 shows an overview of the pitch perception model. The model simulated a pitch ranking test; i.e., for each pair of stimuli presented, the model decided which sound had a higher pitch. An artificial neural network (NN) constituted the core of the model. Depending on the type of "listener" being simulated, the NN received as input a pair of average neural activities computed as described in Section 2.2 or Section 2.3, representing NH or CI listeners, respectively.

Throughout the course of training, the network learned to associate the higher (lower) pitch with a high (low) output level. For example, if the first input had a higher F0 than the second input, the first output would ideally be 1, and the second output would be 0. Based on this paradigm, the input-output connecting weights were adjusted using an error back-propagation algorithm (Demuth et al. 2008).

Fig. 5: Flow chart of the pitch ranking model (top). Spatio-temporal maps were calculated for each of the two input sound stimuli, using the acoustic or electric model of hearing, simulating normal hearing or hearing with a CI respectively. Neural activities were averaged over a 100-ms time interval and the result was applied as the

input to the artificial NN. The model decided which of the two stimuli had a higher pitch. The first (second) output neuron represented the first (second) sound. The output neuron with the higher value would indicate the higher pitch. An example of the NN outputs for 10 arbitrary pitch pairs applying NH data (bottom). D1 and D2 indicate the desired output (correct answer) while O1 and O2 show the actual output generated by the NN.

2.4.1 Preparing the stimulus pairs

The 192 speech samples that were generated as described in Section 2.1 were divided into four quarters (each 48 samples), with each quarter containing a balanced number of pitch classes (12 of each pitch class) and vowel types (2-4 occurrences of each vowel type). The pitch pairs were drawn from a pool of possible combinations within each quarter (i.e., 48×48 combinations) on the condition that the stimulus pairs had to have different F0s. The vowel types could be the same or different within a pair. Therefore, each quarter would generate (4×12) × (3×12) = 1728 non-equal-F0 pairs, from which, 400 pairs were selected randomly. The network was trained on 75% of the pairs (300 pairs), from which 10% (30 pairs) were applied as validation data to monitor the training process and to avoid over-learning. The remaining 25% (100 pairs) were used as test data to evaluate the model performance. To introduce more variability, for each data quarter, each simulation was repeated five times, using a different set of random initial conditions, totaling 4 (quarters) × 5 (initial conditions) × 100 (test pairs) = 2000 pairs being tested for each experiment.

2.4.2 Training the model

To determine the number of training iterations, human performances in similar tasks derived from available literature were taken into consideration. Looi et al. (2004) reported a performance of about 97% for NH participants and an average performance of about 61% for CI users of ACE and SPEAK strategies in ranking 0.5 Oct. pitch pairs. Sucher et al. (2007) reported an average performance of 89% (SD = 14.7%) for NH listeners in a similar task, with musically-experienced NH listeners

outperforming non-experienced listeners by approximately 16%. CI users of the SPEAK strategy scored about 60.2% (SD = 9.5%) in the same study. In another study, Vandali et al. (2005) showed that ACE users were able to rank 0.5 Oct. pitch pairs correctly about 73% of the time. All the above studies were performed using sung vowels as stimuli. However, the within-pair vowel type was always kept the same. Considering the aforementioned psychophysical results, performance levels of about 90% and 70% were targeted for the NH and CI models, respectively. Preliminary simulations showed that the model would often asymptote at a desired performance, within three iterations of training. In other words, the training process would automatically stop as a result of no further improvements. However, to ensure a balanced training in all experimental conditions (changing SF), training was terminated manually after three iterations

For each test sample, both NN actual outputs (O1 and O2 in Fig. 5) and correct/incorrect answers were saved for further analysis. The correct/incorrect answers were derived from the NN outputs by taking the higher output as 1 and setting the other as 0. If the obtained code matched the desired output (D1-D2), the answer would be considered as correct and vice versa. It should be noted that the NN output and the correct/incorrect answers were two different sets of data. The former can be interpreted as a measure of how easy/difficult it was for the model to make a decision, while the latter can be converted into percent correct scores and be treated as the model accuracy. Both results were considered as primary dependent variables in the statistical analyses performed in Section 3.

The pitch perception model and evaluation of the model were implemented in MATLAB® R2012. Statistical analysis was performed using Minitab® 16 software.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

3 Results

3.1 Model validation

Simulation results for NH and CI models with the benchmark SF (≈2) are summarized in Fig. 6. Pitch ranking scores (presented as % correct) were computed as the ratio between the number of correct responses to the total number of pairs tested in each case. Standard errors of the means are shown with vertical lines. According to the four pitch classes defined in Section 2.1, possible interval sizes were 0.5 Oct. (G2-C#3, C#3-G3 and G3-C#4), 1 Oct. (G2-G3 and C#3-C#4) and 1.5 Oct. (G2-C#4). In order to investigate the effect of interval size on performance, the scores associated with each of the intervals are represented by an individual column. The performance on same-vowel (SV) and different-vowel (DV) stimulus pairs are also separated to observe any possible effect of spectral shape (timbre) variation on the model's judgment of pitch direction. Where available, results from previous psychophysical studies are indicated with overlaying horizontal lines on corresponding columns.

Fig. 6: Simulated pitch ranking performance (% correct) for NH and CI data with SF=2. Within each cluster, mean scores associated with different interval sizes (0.5, 1, and 1.5 Oct.) and same vs. different vowel (SV vs. DV) conditions are represented by individual columns. Vertical bars indicate standard errors of the means within simulation trials. Chance level is 50%. Overlaying horizontal lines show available psychophysical results from previous studies. Study 1, 2, and 3 refer to Looi et al. (2004), Sucher et al. (2007), and Vandali et al. (2005), respectively.

In all the statistical tests, a significance level of 0.05 was assumed. Chi-square test performed on correct/incorrect answers of NH results revealed a strong effect of interval size (Pearson $\chi^2(2, N=2000)=109.50, p<0.0001$) on the performance. No effect of spectral shape variation (SV vs. DV pairs) on pitch ranking scores was observed ($\chi^2(1, N=2000)=0.96, p<0.327$). For a more detailed statistical analysis, analysis of variance (ANOVA) using a general linear model (GLM) was

performed on the actual NN outputs. A logit transformation of the NN outputs (i.e., logit(output) = log(output/1-ouput)) was performed to improve the normality of the residual error prior to conduction the ANOVA. Also, it was necessary to incorporate the direction of pitch changes, rather than the absolute interval size, as a factor when analyzing the NN outputs. Therefore, *pitch-difference* factor with six levels (viz. ±1, ±2, and ± 3, corresponding to 0.5 Oct., 1 Oct., and 1.5 Oct. intervals, respectively, with +/– indicating upward/downward pitch changes) was defined. ANOVA confirmed the strong effect of pitch-difference on the model's outputs ($F(5, 1988)=443.45, p<0.0001$) as well. Also, no significant effect of spectral shape variation ($F(1, 1988)=0.38, p=0.535$) or pitch-difference*spectral shape interaction ($F(5, 1988)=0.30, p=0.916$) was found.

Similar analysis using Chi-square test on CI model's correct/incorrect responses showed a significant effect of interval size ($\chi^2(2, N=2000)=76.24, p < 0.0001$). The effect of spectral shape was only marginally significant for CI data using the benchmark SF ($\chi^2(1, N=2000)=3.80, p = 0.051$).

A one-way ANOVA revealed that the model would perform the pitch ranking task easier with the NH data than with CI data ($F(1, 3998)=3.87, p = 0.049$). This outcome was verified by using a Chi-square test on the average scores as well ($\chi^2(1, N=4000)=367.03, p < 0.0001$).

3.2 The effect of the stimulation field spread

The performance of the CI model vs. different degrees of stimulation field spread (SF) is shown in Fig. 7. Scores associated with different interval sizes are shown with different line styles and markers. The performances over SV pairs are shown with black, while the scores over DV pairs are indicated with grey lines.

Fig. 7: Pitch ranking performance (% correct) vs. degree of stimulation field spread. The width of the stimulation field increases with SF. Different interval sizes are shown with different line styles and markers.

Black (grey) lines indicate SV (DV) pairs. Vertical bars represent standard errors of the means within simulation trials, and where not visible, indicate a very small standard error. Chance level is 50%.

The general trend observed from Fig. 7, and also suggested by the main effect analysis, was that SF and interval size (or pitch-difference) had an effect on the model behavior and performance. For more detailed analysis, ANOVA was performed to investigate the level of significance of these factors and their interaction terms. Although the effect of spectral shape variation was less prominent in the initial analysis, it was decided to include this term as a factor in later analyses to probe possible interactions with other terms.

Using the NN outputs, ANOVA revealed a strong effect of pitch-difference ($F(5, 9940)=202.88$, $p<0.0001$), pitch-difference*SF interaction ($F(20, 9940)=15.4$, $p<0.0001$), and pitch-difference*spectral shape interaction ($F(5, 9940)=2.64$, $p = 0.022$). Fig. 8 (a) and (b) demonstrate the effect of the two interaction terms on NN outputs.

Fig. 8: Interaction plots for SF*pitch-difference (a) and spectral variation*pitch-difference (b) for the CI model. Stimulation field spread increases with SF. Pitch-differences are indicated on the left side of the corresponding lines in (a).

Derived scores were also strongly affected by SF and interval size ($\chi^2(4, N=10,000)=300.70$, $p < 0.0001$ and $\chi^2(2, N=10,000)=300.75$, $p < 0.0001$). For the former, scores monotonically increased with focused excitation fields, including an abrupt jump in scores when moving from SF=2 ($M=[60.2, 74.9, 82.8]$, $SD=[15.2, 12.8, 9.5]$ for 0.5, 1 and 1.5 Oct. intervals, respectively) to SF=1.5 ($M=[71.5, 88.9, 95.2]$, $SD=[13.9, 9.2, 5.0]$ for 0.5, 1 and 1.5 Oct. intervals, respectively).

ACCEPTED MANUSCRIPT

4 Discussion

Given the appropriate type of simulated auditory neural activities, with acoustic model output for NH and electric model output for CI, the pitch ranking model was able to generate performance scores similar to NH or CI listeners. That is to say, the configuration shown in Fig. 5 could replicate NH listeners' scores if the neural activities were computed by an acoustic model of hearing. In the acoustic model, the auditory neurons innervated normal-functioning IHCs distributed over a large portion of the cochlear length. The same model was able to simulate the performance of CI listeners when the peripheral neural excitations occurred in a smaller portion of the cochlear length, generated by electrical stimulations. The excitation patterns were calculated by a model of electrical hearing with clinically-estimated parameters. SF was one such parameter. It described the ability of the stimulation field to propagate and produce neural excitation along the cochlea. The outcome of varying this parameter on the pitch ranking performance of a CI model was the subject of this study.

Before performing simulations with varying degrees of SF, a control condition was designed to ensure that the model to be used as a test platform was valid. In other words, the performance of the model was verified by comparing it to available human performances from the literature. An SF value of about 2 was used as a benchmark in the control simulations. Therefore, it was expected that the performance of the model with CI data and using SF≈2 would be similar to that of the CI users on average.

As shown in Fig. 6, the model was able to reproduce pitch ranking scores that covered most of the psychophysically-obtained results from both NH and CI listeners. Apart from the average scores, the model was able to replicate other phenomena as well. For example, the model could rank the pitch of widely-spaced pairs (1-1.5 Oct.) more easily than the closely-spaced ones (0.5 Oct.). Near-perfect performance was observed in the model using NH data. A ceiling effect for NH listeners in ranking 1

Oct. intervals has also been reported (Looi et al 2004). An increase in performance with increasing the interval size has been reported by Gfeller et al. (2007). Interval size was shown to have a significant effect on the model performance when CI data was used as well. This outcome is also supported by psychophysical results (Looi et al. 2004; Gfeller et al. 2007).

The model's scores in ranking 0.5 Oct. intervals were significantly higher with NH data than with the CI data. This outcome is consistent with data from previous studies (Looi et al. 2004; Sucher and McDermott 2007).

Timbre variation within stimulus pairs has been shown to affect NH listeners' judgments of pitch (Zarate et al. 2013; Caruso and Balaban 2014). Although spectral variation did not seem to have a prominent effect on the model's pitch ranking scores, it should be noted that the stimulus pairs used in this study were spaced more widely than is often used in human psychophysical studies investigating the effect of timbre on pitch perception. For example, Zarate et al. (2013) used intervals smaller than one semitone. Caruso and Balaban (2014) used intervals spaced between one semitone and four semitones. Within this range, they showed that the effect of timbre would gradually diminish as the pitch intervals grew. Since the stimulus pairs used in this study were at least six semitones apart, the latter study could explain why having the same/different types of vowels in a stimulus pair did not affect the model's performance.

The CI model did not seem to be sensitive to within-pair spectral variations either. However, it was observed that the effect of the interval size was stronger for DV pairs compared to SV pairs, especially in ranking 1.5 Oct. intervals. In other words, the model could rank the pitch of DV pairs more easily at substantial pitch differences. The scores associated with SV and DV pairs were still only slightly different.

Focused excitation fields proved to be beneficial for simulated CI hearing, especially when SF was reduced from 2 to 1.5. This indicated that the performance of ACE users could greatly improve if neural excitation regions were slightly narrower. While further reductions in the spread of excitation provided increased benefits, the degree of benefits from further reductions was considerably reduced.

Furthermore, if we define certainty in the model as how distant the two NN outputs are from each other, then it can be inferred from Fig. 8 that wide stimulation fields would cause less certainty in the model. This effect was more pronounced in larger pitch intervals compared to smaller ones.

SF is a patient-specific parameter. CI users of curved arrays have smaller SFs compared to the straight array users (Cohen 2009-a). SF also varies in different locations along the cochlea. Apical regions tend to have larger SFs compared to basal regions. This may partially explain the performance variability among CI users. Moreover, it seems likely that obtaining functional gain by applying techniques that lead to narrow excitation in the cochlea depends also on the existing extent of spread of excitation in individual CI patients. Patients with already narrow spread of excitation (SOE) measures (Cohen 2009-b) may benefit less, from current focusing techniques.

Finally, it should be noted that the model parameters used in the simulations in this study were based on Nucleus® devices from Cochlear™. However, such parameters can be derived for other types of implants or sound processing strategies to simulate any particular configurations.

5 Conclusions

Computational models of the auditory system can help explain human auditory performance and be applied to investigate the factors affecting the performance of CI users. With realistic assumptions, modelling approaches make it possible to estimate the impact of different variables, including those that are physically inaccessible, on the auditory capabilities of CI patients in performing various tasks.

ACCEPTED MANUSCRIPT

A pitch ranking model was developed in this study. The performance of the model was validated according to the human results from both NH and CI users. The model was then used to study the effect of the stimulation field spread on the pitch ranking performance of CI users. It was observed that narrow stimulation fields boosted the place cues and improved the pitch ranking scores.

Acknowledgements

This research was supported by Australian Research Council (ARC) Discovery Grant DP1094830 and the Victorian Life Sciences Computation Initiative (VLSCI). The authors acknowledge the support that the Bionics Institute receives from the Victorian Government through its Operational Infrastructure Support Program.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

Başkent, D., Shannon , R. V. (2005). Interactions between cochlear implant electrode insertion depth and frequency-place mapping. J. Acoust. Soc. Am. **117**, 1405-1416.

Blamey, P. J., Dowell, R. C., Tong, Y. C., Clark, G. M. (1984). An acoustic model of a multiple-channel cochlear implant. J. Acoust. Soc. Am. 76, 97-103.

Blamey, P. J., Dooley, G. J., Parisi, E. S., Clark, G. M. (1996). Pitch comparisons of acoustically and electrically evoked auditory sensations. Hear. Res. **99**, 139-150.

Bruce, I. C., White, M. W., Irlicht, L. S., O'Leary, S. J., Dynes, S., Javel, E., Clark, G. M. (1999). A stochastic model of the electrically stimulated auditory nerve: Pulse-train response. IEEE Trans. Biomed. Eng. **46**, 630-637.

Carlyon, R. P., Deeks, J. M. (2002). Limitations on rate discrimination. J. Acoust. Soc. Am. **112**, 1009-1025.

Caruso, V. C., Balaban, E. (2014). Pitch and timbre interfere when both are parametrically varied. PLoS ONE **9**, e87065.

Cohen, L. T. (2009-a). Practical model description of peripheral neural excitation in cochlear implant recipients: 2. Spread of the effective stimulation field (ESF), from ECAP and FEA. Hear. Res. **247**, 100-111.

Cohen, L. T. (2009-b). Practical model description of peripheral neural excitation in cochlear implant recipients: 4. Model development at low pulse rates: general model and application to individuals. Hear. Res. **248**, 15-30.

ACCEPTED MANUSCRIPT

Cohen, L. T. (2009-c). Practical model description of peripheral neural excitation in cochlear implant recipients: 5. Refractory recovery and facilitation. Hear. Res. **248**, 1-14.

Cohen, M. A., Grossberg, S., Wyse, L. L. (1995). A spectral network model of pitch perception. J. Acoust. Soc. Am. **98**, 862-879.

Delgutte, B., Kiang, N. Y. (1984). Speech coding in the auditory nerve: I. Vowel like sounds. J. Acoust. Soc. Am. **75**, 866-878.

Demuth, H., Beale, M., Hagan, M. (2008). Neural Network Toolbox™, MATLAB Version 6. The Mathworks Inc.

Donaldson, G. S., Kreft, H. A., Litvak, L. (2005). Place-pitch discrimination of single-versus dual-electrode stimuli by cochlear implant users. J. Acoust. Soc. Am. **118**, 623-626.

Fielden, C. A., Kluk, K., McKay, C. M. (2013). Place specificity of monopolar and tripolar stimuli in cochlear implants: The influence of residual masking. J. Acoust. Soc. Am. **133**, 4109-4123.

Fredelake, S., Hohmann , V. (2012). Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation. Hear. Res. **287**, 76-90.

Galvin, J. J., Fu, Q. J., Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. Ear Hear. **28**, 302-319.

Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., Olszewski, C. (2007). Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. Ear Hear. **28**, 412-423.

ACCEPTED MANUSCRIPT

Gfeller, K., Witt, S., Stordahl, J., Mehr, M., Woodworth, G. (2000). The effects of training on melody recognition and appraisal by adult cochlear implant recipients. J. Acad. Rehab. Audiol. **33**, 115-138.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species – 29 years later. J. Acoust. Soc. Am. **87**, 2592-2605.

Hawks, J. W., Miller , J. D. (1995). A formant bandwidth estimation procedure for vowel synthesis. J. Acoust. Soc. Am. **97**, 1343-1344.

Jolly, C. N., Spelman, F. A., Clopton, B. M. (1996). Quadrupolar stimulation for cochlear prostheses: modeling and experimental data. IEEE Trans. Biomed. Eng. **43**, 857-865.

Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., Won, J. H., Worman, T., Yueh, B., Rubinstein, J. (2009). Development and validation of the University of Washington Clinical Assessment of Music Perception test. Ear Hear. **30**, 411-418.

Ketten, D. R., Skinner, M. W., Wang, G. E., Vannier, M. W., Gates, G. A., Neely, J. G. (1998). In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays. Ann. Otol. Rhinol. Laryngol. **107**, 1-16.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. **67**, 971-995.

Kwon, B. J., Van den Honert, C. (2006). Dual-electrode pitch discrimination with sequential interleaved stimulation by cochlear implant users. J. Acoust. Soc. Am. **120**, EL1-EL6.

ACCEPTED MANUSCRIPT

Landsberger, D. M., Galvinm, J. J. (2011). Discrimination between sequential and simultaneous virtual channels with electrical hearing. J. Acoust. Soc. Am. **130**, 1559-1566.

Landsberger, D. M., Padilla, M., Srinivasan, A. G. (2012). Reducing current spread using current focusing in cochlear implant users. Hear. Res. **284**, 16-24.

Landsberger, D. M., Srinivasan, A. G. (2009). Virtual channel discrimination is improved by current focusing in cochlear implant recipients. Hear. Res. **254**, 34-41.

Laneau, J., Wouters, J., Moonen, M. (2004). Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. J. Acoust. Soc. Am. **116**, 3606-3619.

Licklider, J. C. R. (1951). A Duplex Theory of Pitch Perception. EXPERIENTIA **VII**, 128-134.

Loeb, G. E., White, M. W., Merzenich, M. M. (1983). Spatial cross correlation: a proposed mechanism for acoustic pitch perception. Biol. Cybernet. **47**, 149-163.

Looi, V., McDermott, H. J., McKay, C. M., Hickson, L. (2004). Pitch discrimination and melody recognition by cochlear implant users. Proceedings of the VIII International Cochlear Implant Conference, Indianapolis, Indiana. Elsevier **1273**, 197-200.

Looi, V., Radford, C. J. (2011). A comparison of the speech recognition and pitch ranking abilities of children using a unilateral cochlear implant, bimodal stimulation or bilateral hearing aids. Int. J. Pediatr. Otorhinolaryngol. **75**, 472-482.

Luo, X., Padilla, M., Landsberger, D. M. (2012). Pitch contour identification with combined place and temporal cues using cochlear implants. J. Acoust. Soc. Am. **131**, 1325-1336.

ACCEPTED MANUSCRIPT

Maarefvand, M., Marozeau, J., Blamey, P. J. (2013). A cochlear implant user with exceptional musical hearing ability. Int. J. Audiol. **52**, 424-432.

Macherey, O., Deeks, J. M., Carlyon, R. P. (2011). Extending the limits of place and temporal pitch perception in cochlear implant users. J. Assoc. Res. Otolaryngol. **12**, 233-251.

McDermott, H. J., McKay, C. M. (1994). Pitch ranking with nonsimultaneous dual electrode electrical stimulation of the cochlea. J. Acoust. Soc. Am. **96**, 155-162.

McKay, C. M., McDermott, H. J., Carlyon, R. P. (2000). Place and temporal cues in pitch perception: Are they truly independent? ARLO **1**, 25-30.

McKay, C. M., McDermott, H. J., Clark, G. M. (1996). The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees. J. Acoust. Soc. Am. **99**, 1079-1090.

Meddis, R., Lopez-Poveda, E. A., Fay, R. R., Popper, A. N. (2010). Computational Models of the Auditory System, Springer.

Miller, M. I., Sachs, M. B. (1984). Representation of voice pitch in discharge patterns of auditory-nerve fibers. Hear. Res. **14**, 257-279.

Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., Levine, S. (1995). Electrode ranking of "place pitch" and speech recognition in electrical hearing. J. Acoust. Soc. Am. **98**, 1987-1999.

Oxenham, A. J., Bernstein, J. G., Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. Proceedings of the National Academy of Sciences of the United States of America **101**, 1421-1425.

ACCEPTED MANUSCRIPT

Pascal, J., Bourgeade, A., Lagier, M., Legros, C. (1998). Linear and nonlinear model of the human middle ear. J. Acoust. Soc. Am. **104**, 1509-1516.

Plack, C. J., Oxenham, A. J., Fay, R. R., Popper, A. N., Eds. (2005). Pitch Neural Coding and Perception. New York, Springer.

Qin, M. K., Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. J. Acoust. Soc. Am. **114**, 446-454.

Rose, J. E., Brugge, J. F., Anderson, D. J., Hind, J. E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol. **30**, 769-793.

Saoji, A. A., Landsberger, D. M., Padilla, M., Litvak, L. M. (2013). Masking patterns for monopolar and phantom electrode stimulation in cochlear implants. Hear. Res. **298**, 109-116.

Schatzer, R., Vermeire, K., Visser, D., Krenmayr, A., Kals, M., Voormolen, M., Van de Heyning, P., Zierhofer, C. (2014). Electric-acoustic pitch comparisons in single-sided-deaf cochlear implant users: Frequency-place functions and rate pitch. Hear. Res. **309**, 26-35.

Shamma, S. A., Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. J. Acoust. Soc. Am. **107**, 2631-2644.

Shamma, S. A. (1985). Speech processing in the auditory system I: The representation of speech sounds in the responses of the auditory nerve. J. Acoust. Soc. Am. **78**, 1612-1621.

Shera, C. A., Guinan, J. J., Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. PNAS **99**, 3318-3332.

ACCEPTED MANUSCRIPT

Slaney, M., Lyon, R. F. (1993). On the importance of time-a temporal representation of sound. Visual representations of speech signals, John Wiley.

Srinivasan, A. G., Landsberger, D. M., Shannon, R. V. (2010). Current focusing sharpens local peaks of excitation in cochlear implant stimulation. Hear. Res. **270**, 89-100.

Srinivasan, A. G., Padilla, M., Shannon, R. V, Landsberger, D. M. (2013). Improving speech perception in noise with current focusing in cochlear implant users. Hear. Res. **299**, 29-36.

Stafford, R. C., Stafford, J. W., Wells, J. D., Loizou, P. C., Keller, M. D. (2013). Vocoder simulations of highly focused cochlear stimulation with limited dynamic range and discriminable steps. Ear Hear. **35**, 262-270.

Sucher, C. M., McDermott, H. J. (2007). Pitch ranking of complex tones by normally hearing subjects and cochlear implant users. Hear. Res. **230**, 80-87.

Swanson, B. A. (2008). Pitch perception with cochlear implants. Ph.D. Thesis, The University of Melbourne, Australia.

Tong, Y. C., Blamey, P. J., Dowell, R. C., Clark, G. M. (1983). Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant. J. Acoust. Soc. Am. **74**, 73-80.

Townshend, B., Cotter, N., Van Compernolle, D. S., White, R. L. (1987). Pitch perception by cochlear implant subjects. J. Acoust. Soc. Am. **82**, 106-115.

Van Compernolle, D. S. (1987). Method and apparatus for providing spread correction in a multi-channel cochlear prosthesis, Google Patents.

Van den Honert, C., Kelsall, D. C. Kelsall (2007). Focused intracochlear electric stimulation with phased array channels. J. Acoust. Soc. Am. **121**, 3703-3716.

Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W. D., McDermott, H. J. (2005). Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies. J. Acoust. Soc. Am. **117**, 3126-3138.

Voutsas, K., Langner, G., Adamy, J., Ochse, M. (2005). A brain-like neural network for periodicity analysis. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics **35**, 12-22.

Wang, W., Zhou, N., Xu, L. (2011). Musical pitch and lexical tone perception with cochlear implants. Int. J. Audiol. **50**, 270-278.

Zarate, J. M., Ritson, C. R., Poeppel, D. (2013). The effect of instrumental timbre on interval discrimination. PLoS ONE **8**, e75410.

Zeng, F. G. (2002). Temporal pitch in electric hearing. Hear. Res. **174**, 101-106.

Zilany, M. S. A., Bruce, I. C. (2006). Modelling auditory-nerve response for high sound pressure levels in the normal and impaired auditory periphery. J. Acoust. Soc. Am. **120**, 1446-1466.

ACCEPTED MANUSCRIPT

Fig. 1
Click here to download Figure(s): R1_Fig 1.docx

ACCEPTED MANUSCRIPT

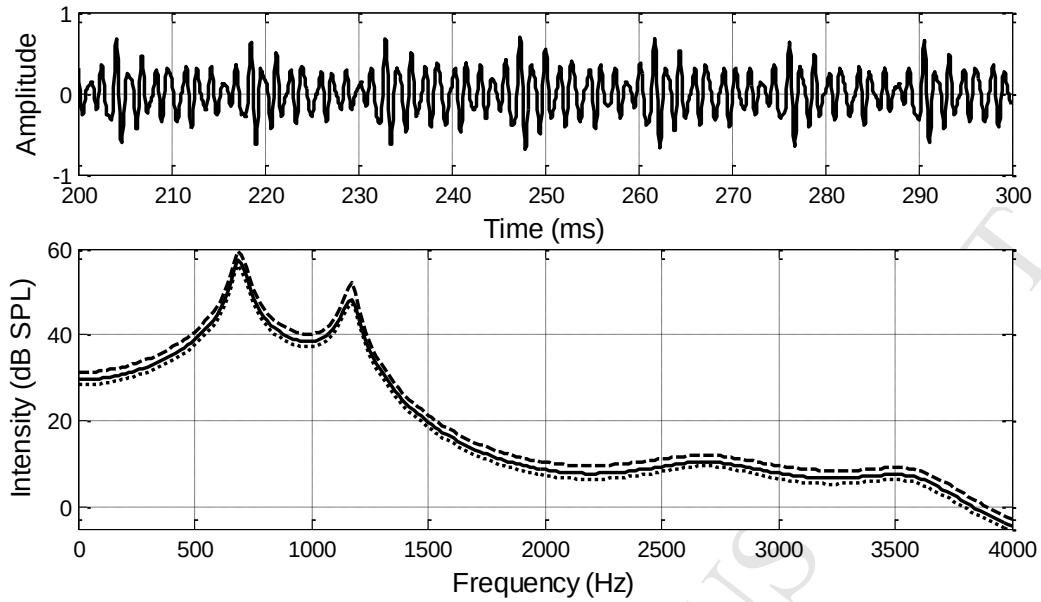


Fig. 1: Temporal (top) and spectral (bottom) representations of a C#3 /a/ synthetic sung vowel with F0=139 Hz, F1=710 Hz, F2=1150 Hz and F3=2700 Hz. Amplitude is expressed in arbitrary units. Variations of bandwidths are shown with different line styles.

ACCEPTED MANUSCRIPT

Fig. 2
Click here to download Figure(s): R1_Fig 2.docx

ACCEPTED MANUSCRIPT

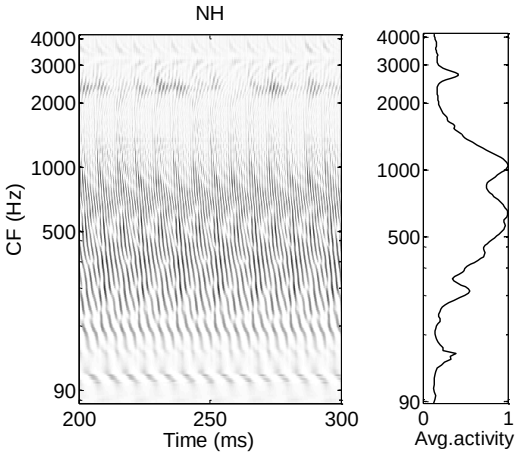


Fig. 2: NH spatio-temporal neural activity of the auditory nerve in response to the stimulus shown in Fig.1. Plot of action potentials, with fiber CFs shown along the ordinate (left), and a histogram of the average activity over the 100-ms interval (right). Dark areas indicate more neural activity.

ACCEPTED MANUSCRIPT

Fig. 3
Click here to download Figure(s): R1_Fig 3.docx

ACCEPTED MANUSCRIPT

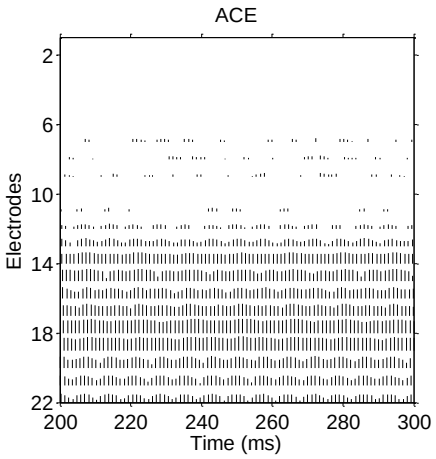


Fig. 3: Simulated electrical stimulation pattern (electrodogram) for the ACE strategy, using the sound stimulus shown in Fig. 1. Time is shown along the abscissa and electrode numbers along the ordinate, with electrode 22 being the most apical. Each vertical bar represents a stimulation pulse with a height associated with the pulse amplitude.

ACCEPTED MANUSCRIPT

Fig. 4
Click here to download Figure(s): R1_Fig 4.docx

ACCEPTED MANUSCRIPT

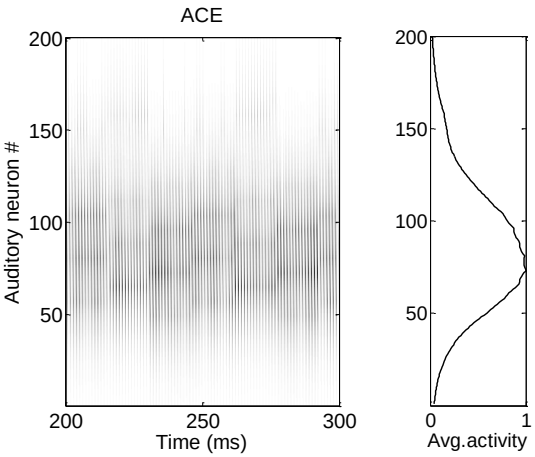


Fig. 4: Spatio-temporal maps for the CI model using the ACE strategy in response to the stimulation pattern presented in Fig. 3. Neuron numbers from 1 to 200 are shown along the ordinate. The average activity over the 100-ms interval is shown on the right side of the map. Dark areas indicate more neural activity.

ACCEPTED MANUSCRIPT

Fig. 5
Click here to download Figure(s): R1_Fig 5.docx

ACCEPTED MANUSCRIPT

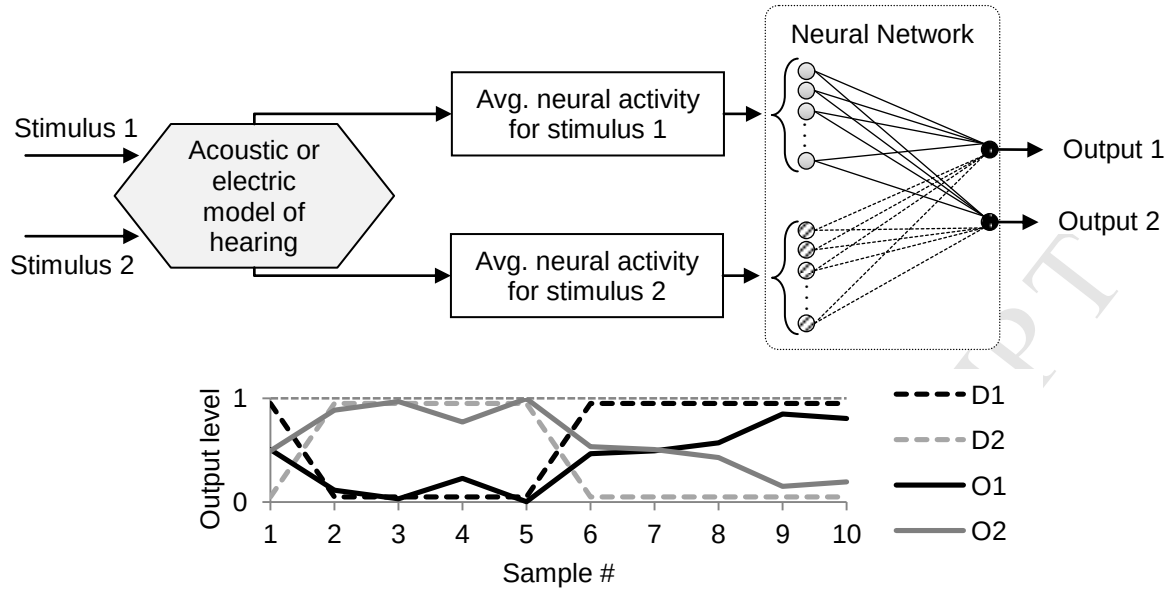


Fig. 5: Flow chart of the pitch ranking model (top). Spatio-temporal maps were calculated for each of the two input sound stimuli, using the acoustic or electric model of hearing, simulating normal hearing or hearing with a CI respectively. Neural activities were averaged over a 100-ms time interval and the result was applied as the input to the artificial NN. The model decided which of the two stimuli had a higher pitch. The first (second) output neuron represented the first (second) sound. The output neuron with the higher value would indicate the higher pitch. An example of the NN outputs for 10 arbitrary pitch pairs applying NH data (bottom). D1 and D2 indicate the desired output (correct answer) while O1 and O2 show the actual output generated by the NN.

Fig. 6
Click here to download Figure(s): R1_Fig 6.docx

ACCEPTED MANUSCRIPT

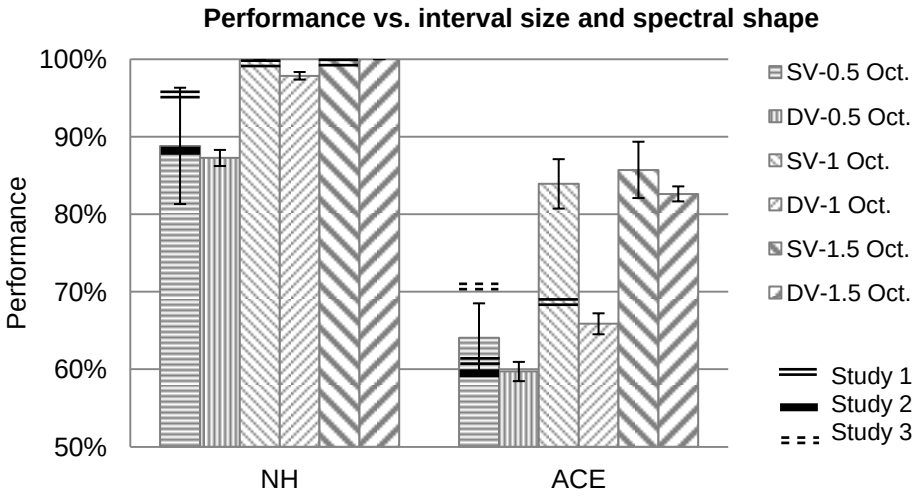


Fig. 6: Simulated pitch ranking performance (% correct) for NH and CI data with SF=2. Within each cluster, mean scores associated with different interval sizes (0.5, 1, and 1.5 Oct.) and same vs. different vowel (SV vs. DV) conditions are represented by individual columns. Vertical bars indicate standard errors of the means within simulation trials. Chance level is 50%. Overlaying horizontal lines show available psychophysical results from previous studies. Study 1, 2, and 3 refer to Looi et al. (2004), Sucher et al. (2007), and Vandali et al. (2005), respectively.

ACCEPTED MANUSCRIPT

Fig. 7
Click here to download Figure(s): R1_Fig 7.docx

ACCEPTED MANUSCRIPT

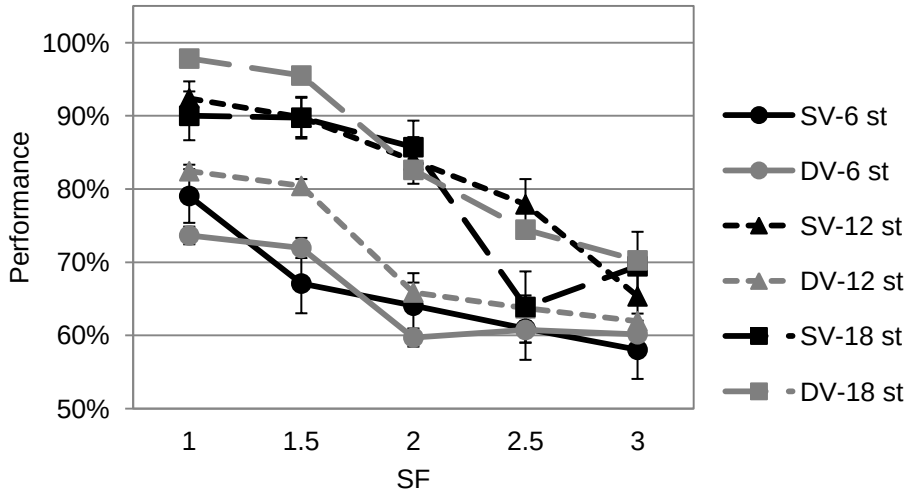


Fig. 7: Pitch ranking performance (% correct) vs. degree of stimulation field spread. The width of the stimulation field increases with SF. Different interval sizes are shown with different line styles and markers. Black (grey) lines indicate SV (DV) pairs. Vertical bars represent standard errors of the means within simulation trials, and where not visible, indicate a very small standard error. Chance level is 50%.

ACCEPTED MANUSCRIPT

Fig. 8
Click here to download Figure(s): R1_Fig 8.docx

ACCEPTED MANUSCRIPT

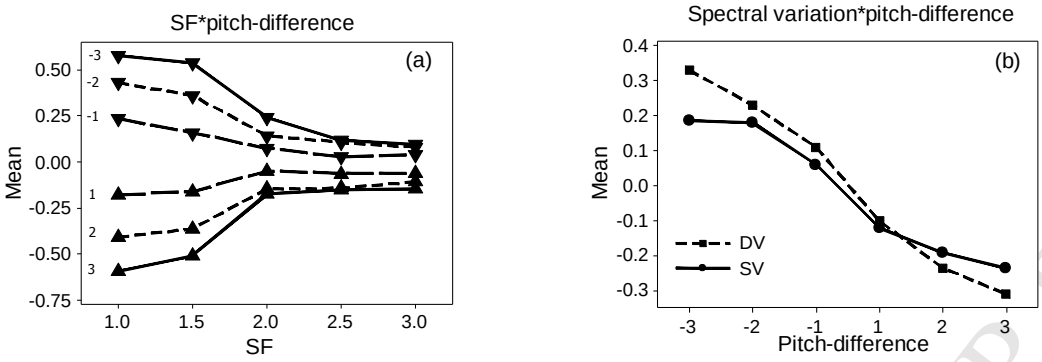


Fig. 8: Interaction plots for SF*pitch-difference (a) and spectral variation*pitch-difference (b) for the CI model. Stimulation field spread increases with SF. Pitch-differences are indicated on the left side of the corresponding lines in (a).

ACCEPTED MANUSCRIPT