Investigation into the Enhancement of Voice Perception

With Simulations of Cochlear Implants and Bimodal Hearing

Jessy Ceha

July 2016

Master Thesis
Human-Machine Communication
Department of Artificial Intelligence,
University of Groningen, The Netherlands

Internal supervisor:
Prof. Dr. Marieke van Vugt (Department of Artificial Intelligence, University of Groningen)

External supervisors:
Prof. Dr. Deniz Başkent (Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen)

Prof. Dr. Jacob Jolij (Department of Experimental Psychology, University of Groningen)

university of groningen

faculty of mathematics and natural sciences

Abstract

Cochlear implants (CI) restore hearing in individuals with profound to severe hearing loss via electric stimulation of the auditory nerve. Unfortunately, modern CIs suffer from degraded fine spectrotemporal resolution. As a result, CI users have difficulty perceiving important voice characteristics for higher level speech perception, specifically, the fundamental frequency (F0) and the vocal-tract length (VTL). Research has shown that CI users with residual low-frequency hearing greatly benefit from a hearing aid in the non-implanted ear. This is known as bimodal hearing and provides electric stimulation via the CI, in addition to acoustic amplification for the low-frequency sounds, that an individual may still hear naturally.

The focus of this thesis was two-fold. First, neurofeedback training was developed, aimed at improving VTL perception using CI simulations. The results indicated a necessity for further research into event-related potentials, specifically the P300 waveform, elicited by non-word, vocoded speech stimuli. Second, a study was conducted which investigated how the comparison and integration of the different sound signals in bimodal hearing impacts perception of the F0 and VTL. Normal hearing listeners heard vocoded CI speech simulations in one ear and low-pass filtered (LPF) speech in the other. Three listening conditions (vocoded CI-alone, LPF-alone, and bimodal) were tested, across varying degrees of spectral degradation, implemented through the vocoder, and with two different LPFs applied to the acoustic signal: 150 and 300 Hz. The results showed a significant improvement in F0 perception in the bimodal conditions compared to the vocoded CI-alone conditions, with no increase in improvement above the 150 Hz LPF. Additionally, there was no impact of spectral degradation on the improvement. The results provide evidence for the role of the F0, present in the acoustic signal, in supporting enhanced speech recognition performance in bimodal hearing. Furthermore, it suggests that amplification of residual low-frequency hearing as low as 150 Hz can provide bimodal benefit in quiet, and that overlapping frequency maps between the CI and hearing aid does not cause interference of the signals.

Table of Contents

1 Introduction

1.1 Perception of Voice Characteristics

Normal hearing listeners are able to discriminate and separate competing voices in multi-talker situations. Under these conditions, listeners are required to segregate sounds that are mixed into one signal to obtain the target voice from the interfering background sounds. Masking studies, measuring the ability of a listener to pull apart simultaneously presented sounds, have found that speech recognition performance relies heavily on whether the target and masker voices differ in gender or whether they have the same gender (Brungart, 2001; Brungart, Simpson, Ericson, & Scott, 2001; Feston & Plomp, 1990). Listeners extract information about a particular speaker through voice characteristics such as the pitch and sound quality (Darwin & Hukin, 2000). Specifically, the fundamental frequency (F0) and the vocal-tract length (VTL) that directly influence perceived pitch and sound quality, respectively, provide the strongest cues for voice gender perception (Darwin, Brungart, & Simpson, 2003; Skuk & Schweinberger, 2013; Vestergaard, Fyson, & Patterson, 2009).

1.1.1 Fundamental Frequency

Due to anatomical differences in their speech production systems, male and female voices sound different. The perceived pitch of a voice is directly related to the rate at which the vocal folds vibrate, known as the glottal-pulse rate. Pitch helps with perceiving information on voicing and manner, as well as conveying prosody (Başkent, Gaudrain, Tamati, & Wagner, 2016). The vocal folds are positioned within the larynx and modulate the airflow expelled from the lungs by vibrating, producing speech (Vashishta, Joshi, & Dhawlikar, 2015). The frequency at which the vocal folds vibrate is known as the fundamental frequency (F0). The size and weight of the focal folds impact the F0. As female vocal folds are shorter and lighter than male's, the average F0 for female adults is around 200 - 220 Hz and around 100 - 120 Hz for male adults (Simpson, 2009). As a result, female voices have a higher perceived pitch than male voices.

1.1.2 Vocal-Tract Length

Another anatomical difference between male and female adults is the vocal-tract length (VTL). The VTL is the distance from the vocal folds to the lips, and is on average 14 - 14.5 cm for adult females and 17 - 18 cm for adult males (Simpson, 2009). An alteration in the VTL corresponds to a shifting of the formant frequencies (i.e. the prominent spectral peaks), without changing pitch or F0 (Darwin & Hukin, 2000). A longer VTL results in a shift of the formants towards the low frequencies. On the other hand, a shorter VTL shifts the formants towards the high frequencies (Fant, 1971; Gaudrain & Başkent, 2015; Mackersie, Dewey, & Guthrie, 2011). As female adult VTLs are on average shorter than male adult VTLs, female voices have higher average formant frequencies, which results in a different timbre, or sound quality (Fant, 1971; Simpson, 2009). The VTL voice characteristic is important as it contains place of articulation cues. Additionally it aids in the discrimination of different vowels and fricatives from one another (Başkent et al., 2016).

The F0 and VTL allow for perceptual segregation, and normal hearing individuals use both for higher level speech perception such as gender categorisation in background noise (Skuk & Schweinberger, 2014; Smith & Patterson, 2005).

1.2 Cochlear Implants

Cochlear implants (CIs) provide hundreds of thousands of people worldwide with a sense of sound (National Institute on Deafness and Other Communication Disorders [NIDCD], 2014). CIs are surgically implanted devices for people with severe to profound hearing loss, which can be a result of damage to hair cells and/or auditory nerve fibres (Shannon, Galvin III, & Başkent, 2001). The device allows these individuals to perceive sound through electrical hearing instead of acoustical hearing. A CI consists of an external portion and a surgically implanted internal portion. The external portion is made up of a microphone, speech processor, and transmitter. The internal portion contains a receiver and an electrode array. The receiver receives signals through the transmitter from the speech processor and converts these signals into electrical impulses. These impulses are then sent to the electrode array that excite regions of the auditory nerve, bypassing damaged parts of the ear, which then transports the signals to the brain (NIDCD, 2014).

1.2.1 Degradation of Spectrotemporal Resolution

Unfortunately, modern CIs transmit sound signals that contain degraded fine spectrotemporal resolution (Rubinstein, 2004). Spectral resolution is mainly reduced due to electrical spread of electrodes on the auditory nerve (Boëx, de Balthasar, Kós, & Pelizzone, 2003; Henry & Turner, 2003). The electrical spread of activation defines the interaction of neural activity that happens between individual electrode stimulation sites.

However, it is also impacted by the number of spectral bands provided in the CI system. In CIs, the frequency components in the signal are resolved using bandpass filtering into a number of different spectral bands, also called channels. The spectral shape information is then transmitted through the pattern of stimulation to distinct electrodes for tonotopical stimulation of the auditory nerve. Modern CIs are generally limited to between 6 and 22 stimulating bands, which is not enough to preserve the fine spectral detail in speech (Henry & Turner, 2003). Studies on speech perception in normal hearing listeners using acoustic simulations of CI processing, have shown that high levels of speech recognition can be achieved with 4 to 12 bands, in quiet, but requiring at least 16 to 20 bands when listening in background noise (Friesen, Shannon, Başkent, & Wang, 2001; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). The effect on spectral resolution of different numbers of bands, can be seen in Figure 1. The less bands, the less the formant structure is retained in the processed signal.

Fine temporal structure is lost due to characteristics of the electric stimulation of the auditory nerve. The CI device delivers only the slow-varying temporal envelope of the corresponding spectral band and amplitude modulates it to fixed-rate current pulses. Therefore, they do not extract the fine temporal structure of sound signals (Ching et al., 2007; Rubinstein & Hong, 2003). This is in contrast to the temporal coding of low frequencies that occurs in the normal auditory system, whereby neurons fire action potentials in phase with the sound waves.

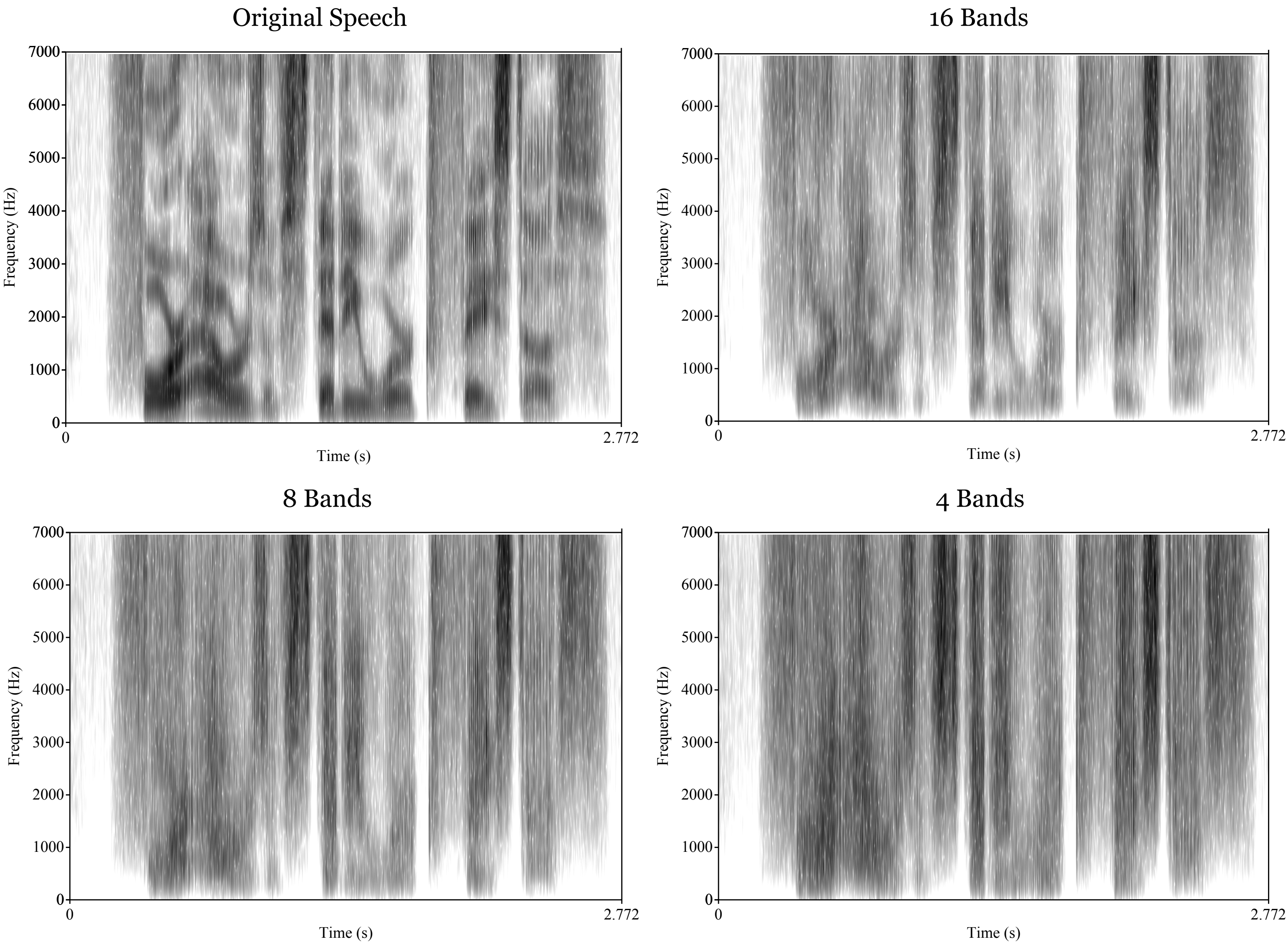


Figure 1: Spectrograms of the original sentence and after processing with 4, 8, and 16 band vocoders simulating a CI. The formant frequencies show up as dark bands running roughly horizontally across the graph. The darker the formant is reproduced in the spectrogram, the more energy there is around its frequency, and the more audible it is. They are very prominent in the original signal, but lose strength with decreasing number of channels. The spectrograms were calculated using PRAAT software (Boersma, 2001) with standard settings, except for the spectral bandwidth of 0-7000 Hz.

1.2.2 **Problems with Speech Perception in Cochlear Implant Users**

The loss in resolution impacts perception of the voice characteristics as perception of the fundamental frequency (F0) relies on temporal fine structure, and perception of the vocal-tract length (VTL) relies on spectral fine structure. One result of these constraints is that CI users seem not to use gender differences to segregate competing speakers (Luo, Fu, Wu, & Hsu, 2009; Stickney, Zeng, Litovsky, & Assmann, 2004). And in contrast to the way in which normal hearing individuals use both F0 and VTL cues to separate concurrent voices, studies have found that CI users base their gender judgements solely on the F0 and are unable to use the VTL cue. This leads them to abnormally categorise the gender of voices (Friesen et al., 2001; Fuller et al., 2014).

Gaudrain and Başkent (2015) investigated whether the VTL voice characteristic can simply not be perceived through the implant due to the poor spectral resolution of the CI. The researchers presented normal hearing individuals with acoustic simulations of CIs. The acoustic simulations were manipulated along two dimensions: the number of electrode stimulation sites, or channels, and the amount of electrical spread on the auditory nerve. As described above, these are two aspects of electrical stimulation which greatly impact the spectral resolution of the implant. The researchers manipulated the number of frequency bands, to simulate the number of channels, and the sharpness of the bandpass filters, to simulate the electrical spread of electrodes.

Gaudrain and Başkent (2015) found that both manipulations influenced VTL perception. They stated that a likely reason for CI users not being able to use the VTL voice characteristic is that VTL 'just-noticeable-differences' (JNDs) for CI users are larger than the typical difference between adult male and female speakers. Their results suggested that reducing the spread of activation by the electrode array might improve VTL perception in CI users.

It may however be possible, without changing the current technology, to improve VTL perception in a different way. Furthermore, although the F0 is more accessible than the VTL, its perception remains degraded compared to normal hearing listeners due to the loss of fine structure in CI signal processing. Two experiments were conducted. The first consisted of a neurofeedback system with the main focus of improving VTL perception, and the second investigated the impact of combining acoustical with electrical hearing on F0 and VTL perception.

2 Theoretical Background

2.1 Experiment 1: Neurofeedback

Electroencephalographic (EEG) biofeedback, or neurofeedback, is a process by which people learn to alter certain aspects of their cortical activity by observing how certain states of cortical arousal feel, and training them to activate such states voluntarily. The goal of neurofeedback is to improve physical or cognitive performance (Vernon, 2005). It is based on the premise that certain states of cortical arousal are associated with specific aspects of behaviour that are thought to be 'optimal'. By training individuals to mirror the pattern of cortical activity that is observed during such 'optimal' states, their performance should improve (Vernon, 2005).

The process of neurofeedback commonly consists of continuously recording and extracting relevant components of an individuals' EEG and supplying them with this information as feedback. Neurofeedback relies on operant conditioning to improve performance. As participants are presented with components of their EEG, they are asked to interact with it in a certain way while being rewarded. For example, the feedback may be in the form of the game 'Snake' whereby a snake shaped icon moves around a computer screen eating 'food', represented by dots. The speed of the snake icon is determined by the amplitude of a certain frequency band in the cortical activity. The aim would then be to have the snake eat as many food dots as possible during the training period (Vernon, 2005).

Recent research by Chang, Iizuka, Naruse, Ando, & Maeda (2014) has shown that a certain type of event-related potential (ERP), related to auditory stimuli, can be used during neurofeedback training to improve auditory discrimination. Chang et al. (2014) found that through neurofeedback training in which participants modulated the amplitude of their mismatch negativity (MMN) brain response via visual feedback, there was a significant improvement in auditory discrimination on the tones used as stimuli. The MMN is a change-specific component of the auditory event-related potential (ERP; Näätänen, Paavilainen, Rinne, & Alho, 2007). It becomes stronger as an index of sound discrimination accuracy and can be elicited by similar sounds that are not consciously discriminated (Chang et al., 2014).

In the study, Chang et al. (2014) calculated the average amplitude of the MMN response and visually presented it on a computer screen as the radius of a solid green disc. During the training, participants were asked to concentrate on making the green disc as large as possible while ignoring the auditory stimuli played through their earphones. A control group was given the same stimuli and instructions as the neurofeedback training group. The difference was that the size of the green disc for participants in the control group did not correspond to their MMN response. The performance in auditory discrimination improved significantly for participants in the neurofeedback group compared to the control group, even though they were not aware of what they were learning nor were they paying attention to the auditory stimuli.

2.1.1 P300

The P300 event-related brain potential (ERP) is elicited with a simple discrimination task in which participants attend to and segregate stimuli that differ from each other, also known as the ‘oddball paradigm’. Two stimuli are presented randomly in a series, with one appearing less frequently, i.e. the oddball (Picton, 1992; Polich & Kok, 1995; Ritter & Vaughan, 1969). The stimuli types are known as target and non-target, occurring with probabilities of 0.20 and 0.80, respectively. Participants are required to pay attention to the target stimuli, for example by mentally counting, and disregard the non-target stimuli. Discrimination between the two stimuli types produces a 10-20 mV positive waveform, found approximately 300 ms post stimulus onset. The P300 ERP has been found to be most clearly recorded from the midline recording sites: Fz, Cz, and Pz (Polich & Kok, 1995).

The P300 and Cochlear Implant Users

The P300 ERP has been successfully elicited in cochlear implant (CI) users (Micco et al., 1995). Using an oddball paradigm with speech stimuli, no significant differences in P300 amplitude and latency were found between nine CI users and nine age-matched controls. As the P300 can be evoked in CI users, its use provides a potential method for improving vocal-tract length (VTL) perception in CI users. This in turn could improve a CI users' ability to differentiate male and female speakers.

P300 Neurofeedback with Cochlear Implant Simulations

Although it remains unclear as to how and why the brain produces the P300, a theoretical account postulates that when a target stimulus is detected, during an oddball task, attentional processes control a change or updating of the stimulus representation (Polich, 2007). This leads to the development of an attentional focus to the stimulus change. As the P300 ERP is generally elicited only as a result of attention to task-relevant stimuli (Picton, 1992), it can be manipulated for use in a neurofeedback system.

A game was designed in which participants were provided with visual feedback which corresponded to the detection of their P300 ERP, elicited by listening to a stream of target and non-target stimuli. The stimuli were vocoded to simulate CIs and the target (oddball) stimulus was set at the participants' VTL just-noticeable-difference (JND) threshold. We hypothesised that by developing an attentional focus to the stimulus change, participants' VTL JNDs would decrease, i.e. VTL perception would improve.

2.2 **Experiment 2: Bimodal Hearing**

As cochlear implant (CI) candidacy has become increasingly lax due to changing audiometric criteria for cochlear implantation, more and more individuals with hearing loss, who receive a CI in one ear, possess residual hearing in the contralateral, non-implanted, ear (Dorman & Gifford, 2010). To supplement the CI, a hearing aid can be fitted to the non-implanted ear. The combination of electric hearing via the CI and acoustic hearing in the contralateral ear, through a hearing aid (HA), is known as 'bimodal hearing'. A similar situation whereby the CI electrode array is implanted less deep, and an HA is worn on the same side as the implant, is known as 'electric-acoustic stimulation' (EAS).

The transmission of low-frequency spectral information is severely limited in CIs (Kong, Stickney, & Zeng, 2005). Bimodal hearing and EAS combine low-frequency information delivered by the HA, with mid- to high-frequency information transmitted through the CI (Ching, van Wanrooy, & Dillon, 2007). Studies have shown that, especially in noise, speech recognition is improved through bimodal hearing and EAS, in real CI patients (Ching, Incerti, & Hill, 2004; Cullington & Zeng, 2010; Dorman, Gifford, Spahr, & McKarns, 2008; Gifford, Dorman, McKarns, & Spahr, 2007; Kong et al., 2005; Sheffield & Gifford, 2014; Shpak, Most, & Luntz, 2013; Zhang, Dorman, & Sphar, 2010) as well as with acoustic simulations (e.g. Chang, Bai, & Zeng, 2006; Kong & Carlyon, 2007; Luo & Fu, 2006; Qin & Oxenham, 2006).

A number of studies suggest that the benefits of EAS and bimodal hearing are the result of F0 information present in the low-frequency acoustic signal (e.g. Chang et al., 2006; Cullington & Zeng, 2010; Qin & Oxenham, 2006; Zhang et al., 2010), while others propose that the F0 is not required for the benefit, and instead preservation of the first formant and other low-frequency cues are important (Kong & Carlyon, 2007).

Cullington and Zeng (2010) assessed a CI patient who retained normal hearing in the contralateral ear. They were interested in the benefit for speech recognition when listening to target speech with a competing talker, in the ear with the CI, as they added low-pass filtered (LPF) information to the contralateral ear. The cutoff frequencies they tested included 150, 250, 500, 1000, 2000, 4000, and 6000 Hz. The patient was tested in three listening conditions: CI-alone, acoustic-alone, and CI + acoustic (bimodal). The results showed that the bimodal benefit was limited to addition of low-frequency LPFs. The high frequency LPFs provided no improvement in performance when combined with the CI.

Zhang et al. (2010) studied adult CI patients who had low-frequency acoustic hearing in the non-implanted ear. They also looked at the three listening conditions: CI-alone, acoustic-alone, and bimodal. The acoustic stimuli conditions were unfiltered and low-pass filtered at 125, 250, 500, and 750 Hz. They found that the addition of low-frequency acoustic information to the CI signal resulted in a significant improvement in word recognition in quiet and sentence recognition in noise. Additionally, the 125 Hz LPF condition provided almost as much bimodal benefit as the unfiltered condition. Both Cullington and Zeng (2010) and Zhang et al. (2010) concluded that the bimodal benefit was most likely due to the presence of F0 information in the low-frequency acoustic signal.

Kong and Carlyon (2007) investigated the contribution of F0 and low-frequency phonetic cues, such as the first formant, to speech recognition. They had normal hearing participants listen to vocoded speech in one ear and low-pass filtered speech in the other. Similar to the other studies, they tested the listening conditions: vocoded CI-alone, acoustic-alone, and bimodal. The researchers found a bimodal benefit but suggested that low-frequency phonetic cues were responsible, and that the F0 was not required for the benefit.

Another study looked at a number of different low-pass (125, 250, 500, and 750 Hz) and pass-band (125-250, 250-500, and 250-750 Hz) filters. Sheffield and Gifford (2014) tested adult bimodal listeners in CI-alone, acoustic-alone, and bimodal conditions. The researchers found a significant bimodal benefit with the 250 Hz LPF for male-talker word recognition in quiet, and a benefit with the 125 Hz LPF for male-talker sentence recognition in multi-talker babble. Additionally, they found that bimodal benefit increased significantly as the LPF increased up to 750 Hz for male-talker word recognition in quiet and female-talker sentence recognition in multi-talker babble, but only up to 500 Hz for male-talker sentence recognition in multi-talker babble. These findings highlight the fact that listening in noise and quiet require different low-pass bandwidths for bimodal benefit, and that the addition of context in sentences compared to monosyllabic words requires different LPFs to obtain the maximum benefit. Lastly, Sheffield and Gifford (2014) found that a significant bimodal benefit could be achieved without the information provided by frequencies below 250 Hz. The researchers suggested that formant structure provided in the higher frequencies may therefore, be just as, if not more important than F0, in providing bimodal benefit.

Although benefit from EAS and bimodal hearing is evident, unclear is which speech cues present in the low-frequency acoustic signal provide the benefit. The goal of the present study was to investigate how additional acoustic information in the contralateral ear to the CI, affects F0 and VTL perception. We presented vocoded speech, to simulate a CI, in one ear and low-pass filtered speech, to simulate a hearing aid, in the other ear. As with previous studies we had three listening conditions: vocoded CI-alone, acoustic-alone, and bimodal. The acoustic low-pass filters (LPFs) tested were 150 and 300 Hz. 150 Hz complemented the frequency range of the vocoder (150 - 7000 Hz), and 300 Hz overlapped with it. Additionally, we examined what influence spectral degradation had on the bimodal benefit in perceiving F0 and VTL. This was done by using either a 4, 8, or 16 band vocoder. Lastly, we had an unprocessed condition to measure the baseline JND.

2.2.1 Hypotheses

(1) Due to the fact that low frequencies contain information about the F0, we predicted F0 perception to improve in the bimodal listening conditions.

(2) It is generally assumed that the auditory system can combine the low frequency input with the high frequency information, without any interference between the two signals (Straatman, Rietveld, Beijen, Mylanus, & Mens, 2010). By testing both 150 Hz and 300 Hz LPFs we investigated whether an overlap would cause interference between information provided by the acoustic signal and the vocoded signal. We hypothesised that the frequency overlaps with a 300 Hz LPF would cause an interference, resulting in larger F0 JNDs caused by a loss in spectral resolution.

(3) To test whether lower formant information has any influence on the bimodal benefit, as Kong and Carlyon (2007) and Sheffield and Gifford (2014) suggest, VTL JNDs were also measured. If in fact, the lower formants do play a role in the benefit, we would expect VTL JNDs to improve as a result of added acoustic signal.

3 Methods

3.1 Common Ground

The neurofeedback and bimodal hearing experiments had overlapping methods. Common to both was a just-noticeable-difference (JND) test, the stimuli used and the application of a vocoder to the stimuli. The JND test was conducted pre- and post-training in the neurofeedback experiment and made up the entire bimodal hearing experiment. Below follows a discussion of the stimuli used in both experiments, the technical aspects behind vocoding the stimuli, and the JND test procedure.

3.1.1 Stimuli

61 syllables were used as stimuli. They were the same stimuli as those employed by Gaudrain and Başkent (2015). The syllables were consonant-vowel tokens spliced from Dutch words uttered by a female speaker, taken from the Nederlandse Vereninging voor Audiologie (NVA) corpus (Bosman & Smoorenburg, 1995).

3.1.2 Vocoder

To simulate cochlear implants (CIs), the stimuli were vocoded, i.e. they were created with a voice encoder, or vocoder, which analyses and synthesises speech input. The vocoder extracted the spectral information of the speech and splits the signal, using bandpass filters, into a number of frequency bands, between 150 and 7000 Hz. Greenwood's function (Greenwood, 1990) was used to space the frequency band boundaries according to their estimated frequency-place locations on the basilar membrane of the cochlea.

In the neurofeedback experiment, the speech signal was split into 8 frequency bands as this had been shown to yield voice gender discrimination and speech recognition performance comparable to that of the best-performing CI users (Friesen et al., 2001; Fu, Chinchilla, & Galvin, 2004).

For the bimodal hearing experiment, stimuli were split into either 4, 8, or 16 frequency bands for different conditions. This was done to replicate situations which generate performance comparable to that of the relatively less proficient CI user (4 bands) and performance above what can be reached by the best CI users (16 bands; Friesen et al., 2001).

Vocoders work with a modulator and carrier signal. The carrier signal is a waveform used to transmit information. The modulator signal contains information to be transmitted and modulates the carrier signal by varying one or more of the carrier signals properties. A temporal envelope in each of the frequency bands was extracted using half-wave rectification (i.e. only positive phases of the input signal were maintained in the output) and low-pass filtering, below a cutoff frequency of 300 Hz. 300 Hz was chosen as it mimics the upper boundary of temporal pitch perception in CIs (Zeng, 2002). A white noise carrier signal, common in vocoding acoustic simulations of CIs, was also generated for each frequency band and the temporal envelopes were used to modulate the amplitude of the carrier signals. The modulated carriers were then summed and the overall level of the aggregate sound was adjusted to the same level as the unprocessed sound (Fuller et al., 2014; Gaudrain & Başkent, 2015).

In addition to modulating the speech signal into different frequency bands, to reflect different numbers of electrode stimulation sites, or channels, the vocoder simulated electrical spread of electrodes. As mentioned earlier, both the number of electrodes and the amount of electrical spread of electrodes influence the spectral resolution of the implant (Gaudrain & Başkent, 2015).

The effect of electrical spread was simulated by changing the order, and therefore the sharpness, of the bandpass filter in the vocoder. A Butterworth filter was used which produces a frequency response as flat as mathematically possible in the passband and rolls off to zero in the stopband. Frequency response describes the way in which the output signal relates to the input signal for different frequencies. It can be characterised by the magnitude of the output signal's response, typically measured in decibels (dB), at different frequencies. A Butterworth filter can be of different 'orders'. A first-order filter means it has only one frequency-dependent component, resulting in a frequency response slope of -6 dB/octave. (An octave represents a doubling or halving of the frequency scale). A second-order filter has a frequency response slope of -12 dB/octave, a third-order filter, a frequency response slope of -18 dB/octave, and so forth. The literature on current spread estimates for CIs indicate an average of 2.8 dB/mm corresponding to frequency response slopes of about -40dB/octave (Bingabr, Espinoza-Varas, & Loizou, 2008; Gaudrain & Başkent, 2015). For this reason, an 8th-order filter, corresponding to slopes of -48 dB/octave, was used in the vocoder for both the neurofeedback and the bimodal experiment.

3.1.3 **Procedure**

To determine the discrimination threshold or just-noticeable-difference (JND) for each condition, an adaptive 3-interval 3-alternative forced choice task was used. This task was administered both before and after the neurofeedback training and made up the entire bimodal hearing experiment.

To obtain the JNDs, numerous trials were required. For each trial, three syllables were randomly chosen and concatenated, with 50 ms of silence separating them, to form a triplet. The participants were presented with this triplet three times, separated by 200 ms, with one of the triplets having had the vocal-tract length (VTL) or fundamental frequency (F0) modified, shifting the triplet by the reference female voice towards a male voice. The order of the syllables remained the same in each triplet and the deviant triplet was randomly assigned to one of the three presentation intervals.

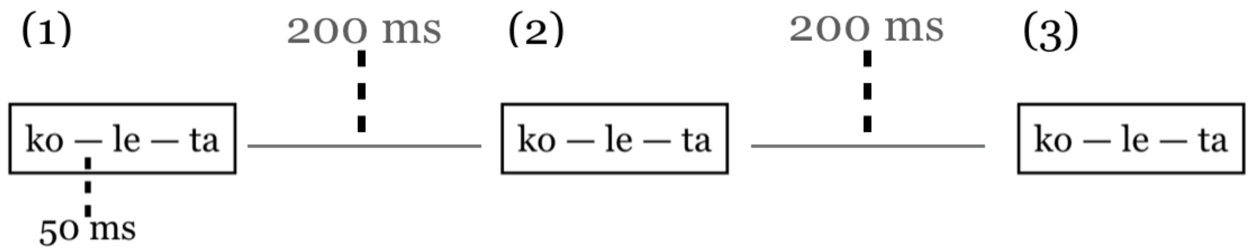


Figure 2: Example of a trial. Three randomly chosen syllables form a triplet. The triplet is presented three times, whereby one of the triplets, randomly assigned to one of the three presentation intervals, has had the VTL or F0 modified.

The reference female voice was set at an F0 of 242 Hz. The average male voice differed from the reference female voice by a VTL difference of 3.8 semitones and an F0 difference of 12 semitones (Gaudrain & Başkent, 2015). A semitone is the smallest distance between two different notes. To achieve each JND measurement, either the VTL was modified with respect to the reference voice, or the F0, but never both at the same time.

The syllables were equalised in root-mean-square for loudness matching whereafter STRAIGHT (Kawahara & Irino, 2005) was applied to decompose them into their F0 contour, spectral envelope, and aperiodicity map. Using STRAIGHT, the F0 and VTL of the three randomly selected syllables were resynthesised using the new F0 and VTL parameter values, even when the original reference voice values were used (Gaudrain & Başkent, 2015).

A staircase procedure was adopted to determine the JNDs. Each threshold measurement started with the deviant triplet differing from the standard triplets by 12 semitones, having either VTL or F0 adjusted, from the reference female voice towards the average male voice. Participants were then asked to determine which of the triplets was different from the other two. They gave their answer by clicking, on a computer screen, one of three boxes numbered '1', '2', and '3', which each corresponded to one of the three triplets they had heard. According to the participant's response, the voice difference of the deviant triplet was modified by a certain step size. Specifically, the difference was reduced by the step size after two correct answers, and the difference was increased by the step size after one incorrect answer. At the beginning of the JND measurement, the step size was 2 semitones. The step size was divided by $\sqrt{2}$, if 15 trials elapsed with the same step size, or when the difference between the stimuli became smaller than 2 times the step size. The JND measurement ended after eight turn-points and the JND was the mean of the last five turn-points. (The stimuli, vocoder parameters, and procedure were replicated from Gaudrain and Başkent, 2015).

3.2 Experiment 1: Neurofeedback

3.2.1 Preliminary Study - Classifier Performance

The classifier during neurofeedback for single-trial analysis used a simplified discriminant analysis method, based on the P3 classifier developed by Bandt, Weymar, Samaga, and Hamm (2009). The goal of the classifier was to distinguish the ERPs elicited by target stimuli from the ERPs elicited by non-target stimuli. To do this it created an average response for each stimuli type, called a template. Before being used for neurofeedback, the analysis method was tested on simulated EEG data.

Simulated Data

EEG signals were simulated for one electrode and 500 time points worth of data were produced for each trial. The signals were created using the Matlab randn function which generates normally distributed pseudorandom numbers. The amount of noise in the signal was manipulated by multiplying randn by a constant. The classifier was trained and tested varying (1) the signal-to-noise ratio and (2) the number of trials used to build the average templates.

Training & Testing

The classifier was trained on a number of target and non-target trials to produce an average ERP template for both stimuli types. Depending on whether it was a target or non-target trial, the signal was manipulated in such a way that there was a clear difference between the two.

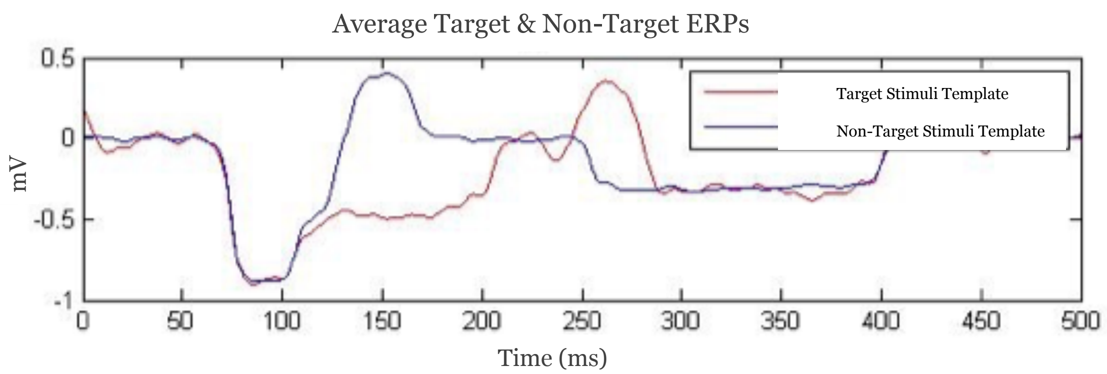


Figure 3: Example simulated data set showing the difference in ERPs between target and non-target stimuli for one electrode.

(1) Signal-to-Noise Ratio

The classifier was first trained and tested with varying levels of noise. In this case, 300 non-target and 300 target trials were used during training, and the same amount for testing.

(2) Number of Trials

Next we trained the classifier by varying the number of trials used to build the templates. The least was 10 trials of each stimulus type, and the most was 300 of each. The classifier was then tested with 300 target and 300 non-target stimuli trials.

3.2.2 Main Study

Participants

Pilot testing was conducted on ourselves in the research group, as well as undergraduate students of the University of Groningen.

Apparatus

The pilot testing was conducted at the Heymans Building of the University of Groningen. Stimuli were presented through speakers placed on either side of the participant. Participants sat approximately 100 cm from a computer screen.

EEG Recording

Event-related potentials were measured with an 8-channel Twente Medical Systems Inc. (TMSI, Oldenzaal, The Netherlands) device. Responses were recorded at a 250 Hz sampling rate from electrodes: T7, T8, P7, P8, Cz, Fz, P3, and P4 (according to the International 10-20 system for EEG electrode placement).

EEG Data Processing

Matlab and OpenViBE (Renard et al., 2010) were used for real-time processing of the electroencephalogram (EEG). Butterworth filters were applied to the incoming signals to distinguish the ERPs of interest from interfering noise. If there were any large amplitudes indicating eye blinks, etc. the trial was discarded. A baseline correction of 200 ms was applied to the signal to eliminate any overall voltage offset, as well as a smoothing function. Signal processing and stimulus presentations were performed in Matlab with a sampling frequency of 48 kHz.

Procedure

A vocal-tract length (VTL) just-noticeable-difference (JND) measurement was conducted pre-training. The triplets of syllables were created from a batch of 31 randomly selected syllables from the original 61. This threshold measurement was used for the neurofeedback. During neurofeedback, the remaining 30 syllables from the original 61 were used. The original 61 were split into 2 groups to eliminate the possibility of a learning effect. A VTL JND measurement was conducted post-training using all 61 syllables.

Neurofeedback

Classifier Training

In order to provide participants with neurofeedback based on their P300 response, a classifier needed to be trained for single-trial analysis, to distinguish the ERP elicited by the non-target stimuli from the ERP elicited by the target stimuli. To do this, a template for each stimulus type was created. This was done by having participants complete a task, 100 trials at a time (80 non-target and 20 target stimuli, randomly ordered after the first 5 being non-targets), with a short break in between if necessary. A total of 300 trials (240 non-target, 60 target) were required to build reliable templates. The task consisted of counting the number of target sounds in the stream of stimuli. The stimuli, per 100 trials, was a single syllable picked at random from the 61 syllables. The inter-stimulus-interval (ISI) was 1200 ms. The non-target stimulus was the syllable without any VTL modification, and the target stimulus was the syllable with a VTL difference of 12 semitones (in the male voice direction from the reference female voice). The syllables were not vocoded so as to provide a clearly audible difference between the non-target and target stimuli which would help with the clarity of the different ERPs. The stream of stimuli always started with at least 5 non-target stimuli so that participants could clearly recognise the target.

As participants were completing the task, ERPs for the non-target and target stimuli were recorded. EEG was recorded from 8 electrodes (T7, T8, P7, P8, Cz, Fz, P3, and P4) at a sampling rate of 250 Hz. At the end of the task, an average ERP template for non-target stimuli and an average ERP template for target stimuli were calculated for each of the 8 electrodes. The templates were analysed to determine which electrodes recorded the largest difference between the ERPs for non-target and target stimuli, as well as an approximate time window for the P300. This information was then used to achieve the most accurate classification performance during single-trial analysis.

Single-Trial Classification

Game

Once the P300 classifier had been trained, it could be used for single-trial classification of ERPs during the neurofeedback training. The training required participants to be sat in front of a computer screen on which they would play a game. The game consisted of listening to a stream of auditory stimuli presented through speakers, with an ISI of 1200 ms. For each round of the game, a single syllable was chosen from a list of 30 syllables (different from the 31 used during the pre-training JND measurement). The 30 syllables were randomly ordered for each participant, and no syllable was repeated during the game. The syllable was modified with a vocoder into 8 frequency bands, with an 8th-order filter. The target stimuli were produced by adjusting the VTL of the syllable according to each participant's VTL JND threshold recorded during the pre-training measurement.

As with the task used to build the classifier, the stimulus stream during the game, consistently began with at least 5 non-target stimuli. Whilst the stream of 64 stimuli played (48 non-target and 16 target stimuli, randomly ordered after the first 5 non-targets) participants saw, on the computer screen in front of them, a large black square. This square was made up of 16 smaller black squares. Participants were instructed to use their 'Brain Power' to remove these smaller black squares to reveal the image hidden behind (Figure 4), whilst simultaneously listening to the auditory stimuli. After the stream of 64 stimuli was over, the round was complete and participants could take a short break, or continue on to the next round.

The next round consisted of a new image being hidden behind the 16 black squares and on the right-hand side of the screen participants could see how many black squares they removed with their 'Brain Power' the previous round. They were encouraged to remove more black squares each round. This process would continue on until all 30 syllables had been played, i.e. a total of 30 rounds.

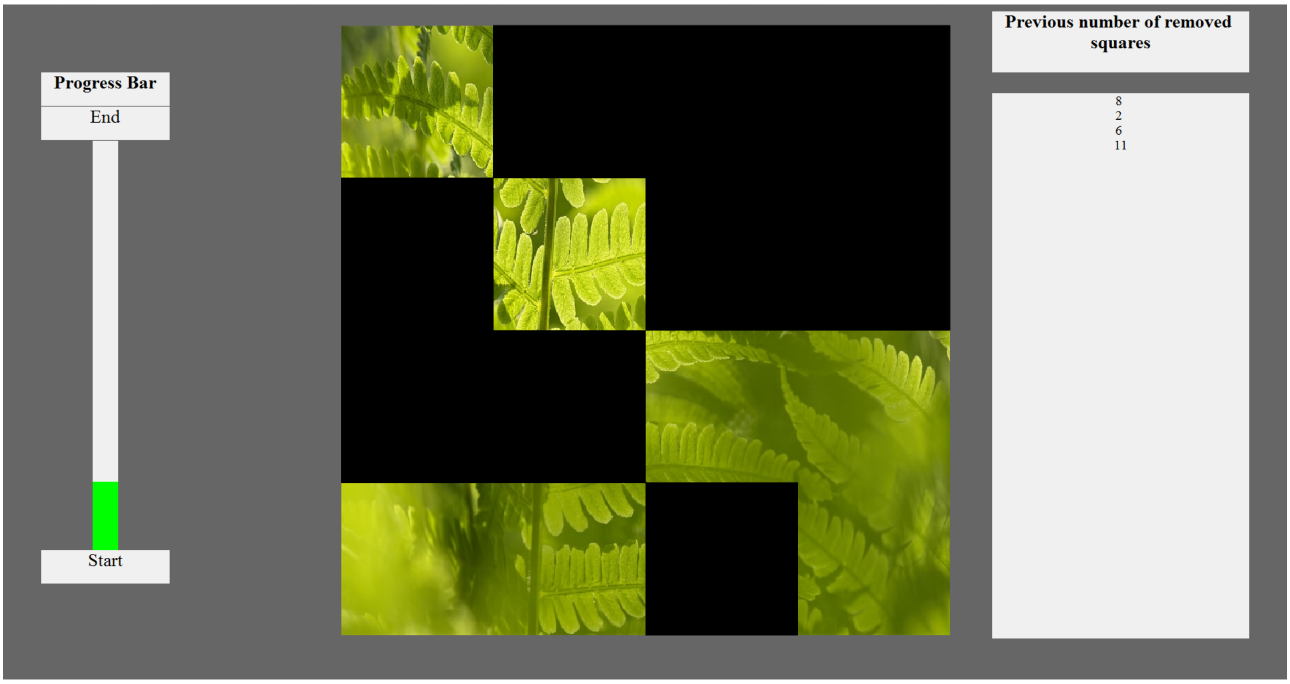


Figure 4: Neurofeedback game. Participants tried to remove as many of the black squares covering the picture behind. On the right hand side, the player could view how many black squares they had removed in the previous rounds. On the left, they could keep track of how far they were with the game.

Classifier

During each round, black squares were removed, in real time, in the following way: for each stimulus presented, i.e. a trial, the classifier analysed the ERP and determined whether a P300 waveform was present or not. It did this through a simplified discriminant analysis method, based on the P3 classifier developed by Bandt et al. (2009). First, for each electrode, the classifier subtracted the ERP template for non-target stimuli from the ERP template for target stimuli, creating a new template. This new template was multiplied by the ERP of the incoming trial to be classified and then the mean was taken over time. This calculation produced a 'score' for the incoming trial ERP which, if it exceeded a certain cut-off value, meant it was classified as a target trial, having elicited a P300, and otherwise as a non-target trial. The value used as a cut-off for classification, was the score half-way between the score for the average target ERP and the score for the average non-target ERP. The classification was sent to the game, and a black square was removed if a target trial had correctly been classified. For the control group, the classification was random.

3.3 **Experiment 2: Bimodal Hearing**

Participants

All 8 participants, aged 19 to 28 (mean 22.5, standard deviation 3.2) had auditory thresholds ≤ 20 dB hearing level at octave frequencies between 125 and 8000 Hz. All participants were native Dutch speakers or had Dutch as one of the languages spoken during their childhood. The participants signed informed consent prior to beginning the experiment and received an hourly wage. The study was approved by the Medical Ethics Committee of the University Medical Center Groningen.

Apparatus

Stimuli were presented through HD600 headphones (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany) via an AudioFire4 sound card (Echo Digital Audio Corp, Santa Barbara, CA) connected to a DA10 D/A converter (Lavry Engineering, Poulsbo, WA) through S/PDIF. Participants were seated in a sound-attenuated booth in the Neuroimaging Center of the University Medical Center Groningen. The sound level in all listening conditions (vocoded CI-alone, acoustic-alone and bimodal) was adjusted to be 62 dB sound pressure level. This was maintained for each participant. Signal processing and stimulus presentations were performed in Matlab with a sampling frequency of 44.1 kHz.

Acoustic Signal

In the bimodal listening conditions, low-pass filtered (LPF) acoustic signal was presented to the left ear, as vocoded CI simulated signal was presented to the right ear. The acoustic signal was processed by a 6th-order Butterworth LPF at cutoff frequencies: 150 and 300 Hz, to retain different amounts of low-frequency acoustic information. The vocoded CI-alone conditions, and the baseline/unprocessed condition, were only presented to the right ear, with no low-pass filtered acoustic signal.

Procedure

In contrast to the single condition (8 bands, 8th-order filter) VTL JND measurement taken for the neurofeedback experiment, the bimodal hearing experiment, consisted of both F0 and VTL JND measurements for a number of different conditions. The vocoder throughout this experiment had an 8th-order filter.

The conditions were as follows:

Acoustic-alone	Bimodal
(1) LPF = 150 Hz	(7) LPF = 150 Hz & Voc = 4 bands
(2) LPF = 300 Hz	(8) LPF = 150 Hz & Voc = 8 bands
Vocoded CI-alone	(9) LPF = 150 Hz & Voc = 16 bands
(3) Voc = 4 bands	(10) LPF = 300 Hz & Voc = 4 bands
(4) Voc = 8 bands	(11) LPF = 300 Hz & Voc = 8 bnads
(5) Voc = 16 bands	(12) LPF = 300 Hz & Voc = 16 bands
Baseline	
(6) Unprocessed sound (No vocoder)	

LPF = Low-pass filter, Voc = vocoder

For each condition, a VTL JND and an F0 JND were measured (in the direction of the male voice with respect to the reference female voice), resulting in 12*2 = 24 threshold measurements. Each threshold measurement was repeated twice, resulting in a total of 48 measurements. The experiment was split into two sessions, lasting two and a half hours each. For each participant, the 48 measurements were randomly ordered, and they completed 24 of the measurements in the first session, and the remaining 24 in the second. Prior to commencing each threshold measurement, participants listened to two sentences, one without modification, i.e. unprocessed, and one modified with the upcoming condition's parameters.

4 Results

4.1 Experiment 1: Neurofeedback

4.1.1 Preliminary Study - Classifier Performance

To asses classifier performance prior to implementation during neurofeedback, we tested the classifier with varying signal-to-noise ratios (SNRs), as well as investigated the minimum number of trials required for a high level of classifier accuracy given a set SNR.

(1) Signal-to-Noise Ratio

Classifier performance was measured, for one electrode, as a function of the signal-to-noise ratio (SNR). The SNR is defined as the ratio between the variance in the signal (maintained at one) and the variance in the noise, as the signal is zero-mean.

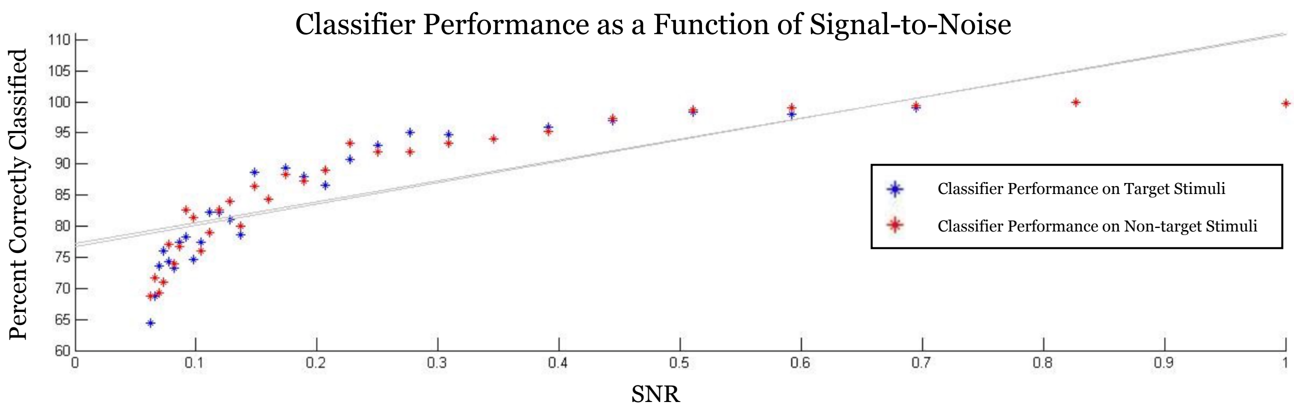


Figure 5: Classifier performance as a function of signal-to-noise ratio (SNR). Each data point represents the average performance of the classifier on either 300 target or 300 non-target stimuli trials.

Figure 5 shows that, unsurprisingly, with higher SNR, classifier performance increases. With no noise, the classifier can reach close to 100% accuracy. As the amount of noise increases, performance drops. Interestingly, even though the data was not filtered or smoothed, with a lower SNR, the classifier could still achieve very high classification accuracy, e.g. 95% correctly classified target stimuli, at approximately 0.28 SNR.

(2) Number of Trials

Based on Figure 5, assuming we accept a performance accuracy of 80% with a noise level of approximately 0.16 SNR, we tested how many trials we needed to maintain an 80% level of accuracy. We found that a high level of performance accuracy can already be achieved with at least 100 trials of both target and non-target stimuli; whereafter performance levels-off (Figure 6).

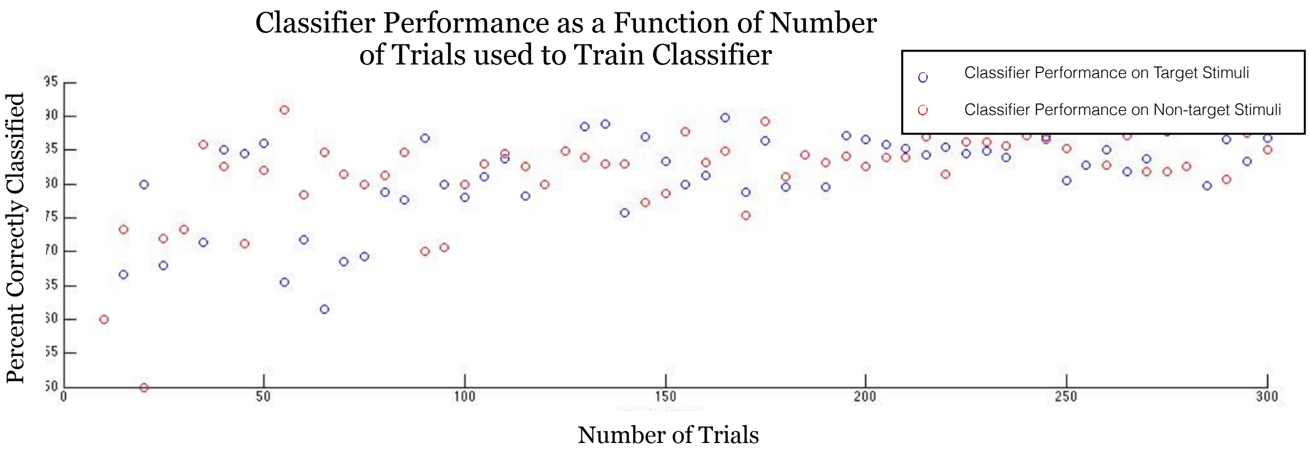


Figure 6: Classifier performance as a function of the number of trials used to train the classifier. Each data point represents the average performance of the classifier on the number of trials tested.

4.1.2 **Main Study**

Neurofeedback

A number of pilot tests were run on the neurofeedback set up. The results from one pilot test are reported here. To recall, the aim of the neurofeedback was to improve vocal-tract length (VTL) just-noticeable-difference (JND). To track this, a pre- and post-neurofeedback training JND measurement was taken.

Pre VTL JND measurement (semitones)	Post VTL JND measurement (semitones)
6.0465	3.4322

Table 1: Results from the pre- and post-VTL JND measurements for one participant. The JND is the difference in semitones from the reference female voice.

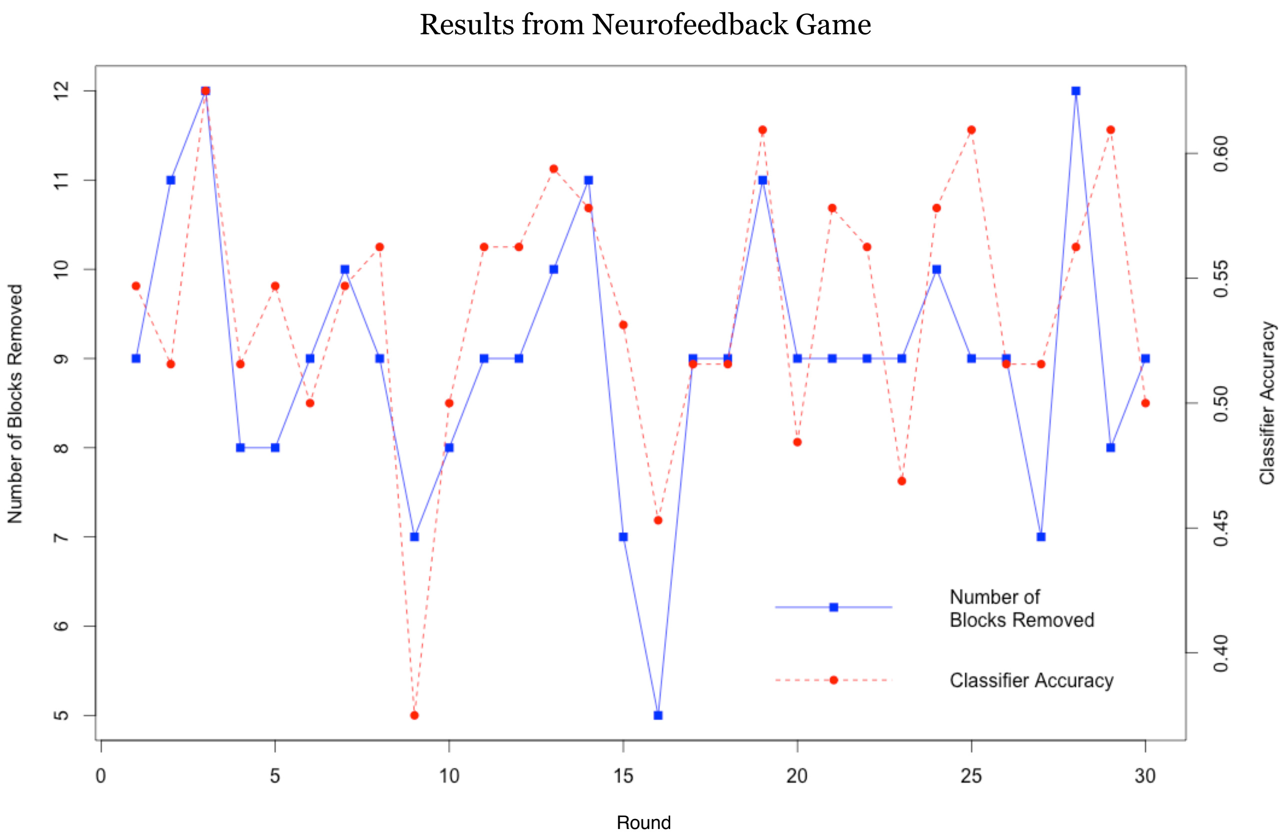


Figure 7: Graph depicting results from one pilot participant. The blue data points, and the axis on the left, represent the number of black squares removed per round during the game. The aim of the game was to remove more black squares each round. The red data points, and the axis on the right, show classifier accuracy per round.

As seen from Figure 7, classifier accuracy remained between 50-60% throughout the duration of the game. Consequently, the participant reported that they thought they had no influence over the game and that it felt random as to when a black square was removed. As seen from Table 1, although the participant's VTL JND measurement improved post-neurofeedback training, we can not make any inferences on these one-participant results as to the reason behind the improvement, or whether it is significant.

Classifier accuracy averaged 50-60% for all pilot tests which triggered concern. After further inspection, we noticed that the target stimuli were not eliciting the P300 as expected, let alone any event-related potentials (ERPs). As there were no clear ERPs to be measured, the templates for the non-target and target stimuli were too similar for the classifier to reliably tell incoming ERPs apart (Figure 8).

Target and Non-Target ERP Templates

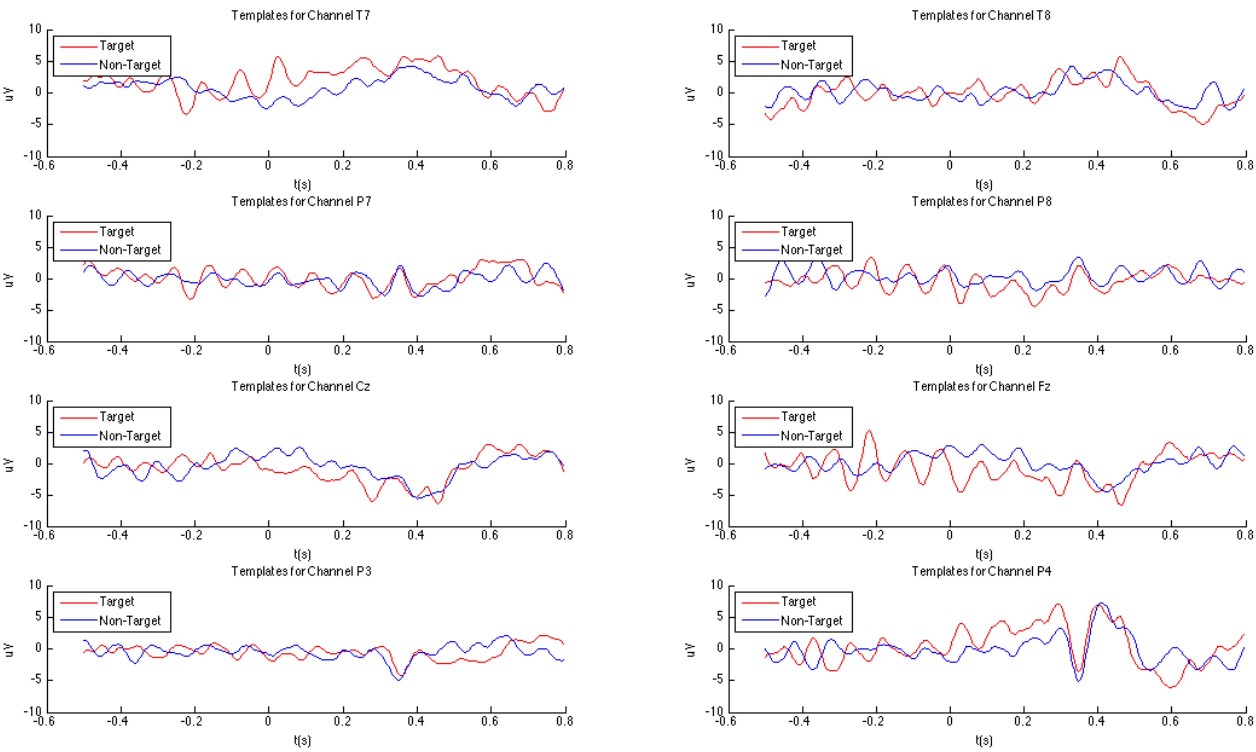


Figure 8: Mean templates for target and non-target stimuli after 300 trials (80% non-target, 20% target). No evidence of ERPs, and no clear difference between the two templates.

Exploratory Tests

We conducted several tests to figure out what the cause of the lack of ERPs may be. The tests are listed in Table 2. None of the tests produced a significant difference to the recorded EEG, i.e. no ERPs were being elicited by target stimuli.

Original Setup	Test
12 semitone difference between stimuli	24 semitone difference
Electrode configuration: T7, T8, P7, P8, Cz, Fz, P3, P4	Electrode configuration: T7, T8, P7, P8, O1, O2, Pz, Cz
Stimuli: - 1 syllable	Stimuli (more word-like): - 1 syllable repeated 3 times to form a triplet - 3 different syllables joined to form a triplet
Analyse EEG data of 0.800 seconds around stimulus onset	Analyse EEG data of 1.000 second around stimulus onset
240 non-target stimuli trials & 60 target stimuli trials	400 non-target stimuli trials & 100 target stimuli trials
Speech stimuli	Pure tones

Table 2: The left side shows the parts of the original neurofeedback setup that were changed, and on the right, the changes made to see whether they would elicit ERPs.

4.2 **Experiment 2: Bimodal Hearing**

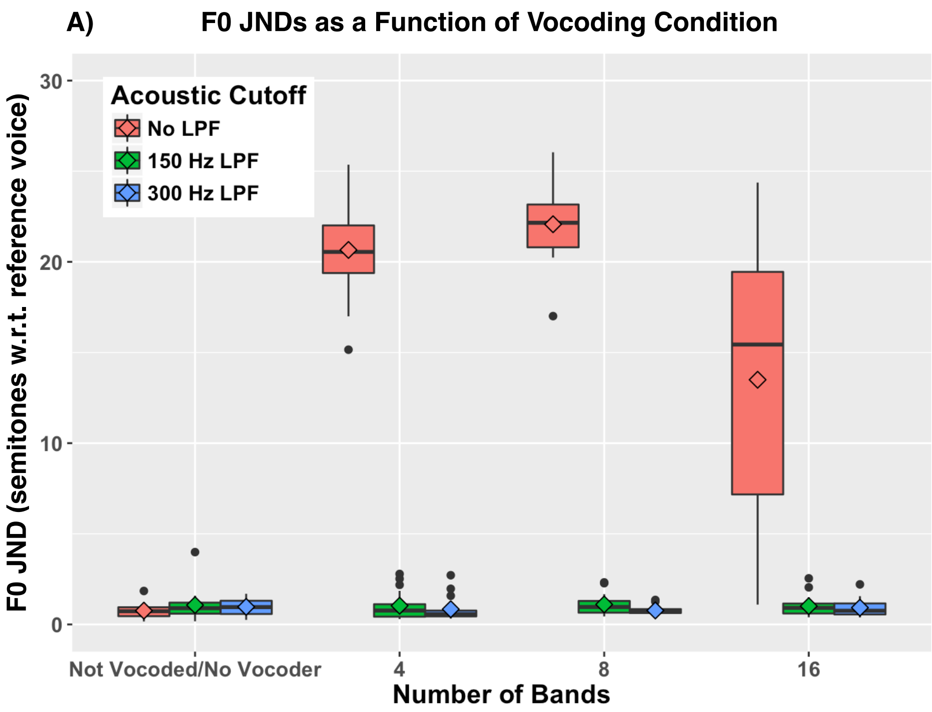
We predicted that fundamental frequency (F0) perception, measured as the F0 just-noticeable-difference (JND), would improve as a result of bimodal hearing. Bimodal hearing was created by combining vocoded signal, simulating a cochlear implant (CI), in the right ear with low-pass filtered (LPF) signal, representing a hearing aid, in the left ear. Three listening conditions were investigated: vocoded CI-alone, acoustic-alone, and CI+ acoustic (bimodal). Two LPFs were tested: 150 Hz and 300 Hz, as well as three levels of spectral degradation, simulated by 4, 8, and 16 bands in the vocoder. A condition with unprocessed sound for baseline measurement was also tested.

In addition to an improvement in F0 JND in the bimodal listening condition, we predicted that a 300 Hz LPF in the acoustic signal would overlap and interfere with the spectral information provided in the frequency range covered by the CI (150 - 7000 Hz), resulting in the measured F0 JNDs being larger than with a 150 Hz LPF.

Condition	**Mean *F0* JND**
Unprocessed	0.7607912
8 bands & 300HZ	0.7699194
4 bands & 300Hz	0.8407638
16 bands & 300Hz	0.9235081
300Hz	0.9681094
16 bands & 150Hz	1.0118763
4 bands & 150Hz	1.0324119
150 Hz	1.0669725
8 bands & 150Hz	1.1025944
16 bands	13.4962250
4 bands	20.6568000
8 bands	22.0748375

Condition	**Mean *VTL* JND**
Unprocessed	0.9941831
16 bands	2.3810500
16 bands & 150Hz	2.9012056
16 bands & 300Hz	3.3301687
8 bands	3.7530750
8 bands & 300HZ	4.8157188
8 bands & 150Hz	5.1540063
4 bands	7.3903187
4 bands & 150Hz	8.5160625
4 bands & 300Hz	8.7351187
300Hz	12.0572000
150 Hz	13.9769375

Table 3: Average F0 and VTL JNDs, in semitones, for each listening condition.



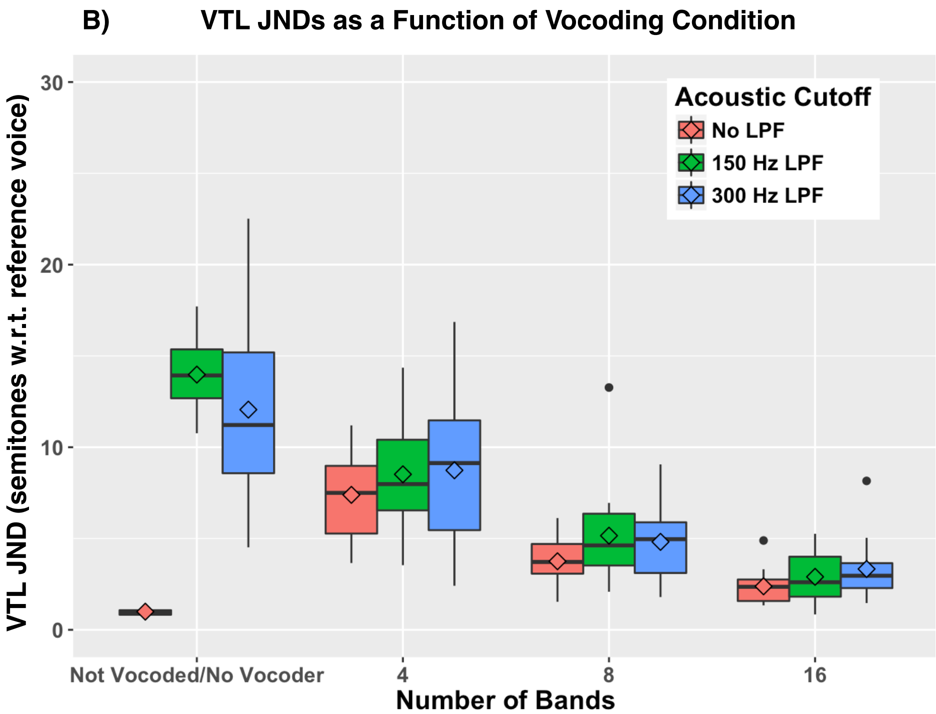


Figure 9: The F0 and VTL JNDs (in semitones with respect to (w.r.t.) the reference voice), in panel A and B respectively, for all participants are represented by box plots. The boxes extend from the lower to the upper quartile, where the midline is the median. The whiskers show the greatest and least values, excluding outliers. The outliers, defined as data points larger than 1.5 times the inner quartile, are the black dots. The diamonds represent the mean JND for F0 and VTL. [LPF = Low-pass filter]

The box plots in Figure 9, A and B, show the aggregate F0 JND and VTL JND measurements, respectively, for all participants as a function of Number of bands in the vocoder (4, 8, 16, and no vocoder) and low-pass filter (no filter, 150 Hz, and 300 Hz). The average JNDs, split by F0 and VTL, are listed in Table 3. As expected, the baseline/ unprocessed condition, no vocoder and no low-pass filter, resulted in the smallest F0 and VTL JNDs. Additionally, as VTL perception is greatly influenced by spectral degradation, VTL JNDs got larger the more spectral degradation was implemented in the vocoded signal (from no degradation, to 16, 8, and then 4 bands). Another difference along the dimension of the two voice characteristics, as seen in Table 3, is that in the acoustic-only listening conditions, perception of F0 was much better than perception of the VTL.

A repeated-measures analysis of variance (ANOVA) on the averaged JNDs was performed, with Number of Bands in Vocoder (16, 8, 4, and No vocoder), Acoustic Low-Pass Filter (No filter, 150Hz, and 300Hz), and Voice Characteristic (F0 and VTL) as repeated factors. The reported p-values were corrected with the Greenhouse-Geisser correction when the sphericity assumption was violated. The generalised eta-squared (η_G^2) measure of effect size (Bakeman, 2005) is also reported. All results from the ANOVA can be seen in Table 4.

Effect	**DF-effect**	**DF-error**	**F**	**p**	η_G^2
NumberOfBands	3	21	72.350369	0.0000000	0.3282825
AcousticCutoff	2	14	121.651926	0.0000000	0.5385604
VoiceCharacteristic	1	7	7.472165	0.0291897	0.0338445
NumberOfBands:AcousticCutoff	6	42	82.915822	0.0000000	0.6849800
NumberOfBands:VoiceCharacteristic	3	21	77.229534	0.0000005	0.5387797
AcousticCutoff:VoiceCharacteristic	2	14	178.440557	0.0000001	0.7910199
NumberOfBands:AcousticCutoff:Voice Characteristic	6	42	14.858395	0.0000504	0.2227928

Table 4: Results of the repeated measures ANOVA with NumberOfBands, AcousticCutoff, and Voice Characteristic as repeated factors. [DF-effect = degrees of freedom in the numerator & DF-error = degrees of freedom in the denominator for the F-statistic]

The analysis showed that the Number of bands in the vocoder significantly altered the JND measurements [NumberOfBands: $F(3,21) = 72.35$, $p<0.0001$, $\eta_G^2 = 0.33$], as did the low-pass filter [AcousticCutoff: $F(2,14) = 121.65$, $p<0.0001$, $\eta_G^2 = 0.54$]. The JNDs for F0 and VTL were also significantly different [VoiceCharacteristic: $F(1,7) = 7.47$, $p<0.05$, $\eta_G^2 = 0.03$].

Of interest is that JND measurements depended on the interaction between the number of bands in the vocoder and the low-pass filter [NumberOfBands X AcousticCutoff: $F(6,42)=82.92$, $p<0.0001$, $\eta_G^2 = 0.68$]. Additionally, the ANOVA showed that the difference between the F0 and VTL JNDs depended on the interaction between the number of bands in the vocoder and the acoustic low-pass filter [NumberOfBands X AcousticCutoff X VoiceCharacteristic: $F(6,42)=14.86$, $p<0.001$, $\eta_G^2 = 0.22$].

Fisher's post-hoc Least Significant Difference (LSD) test at $p \leq 0.05$, confirmed that F0 JNDs significantly improved for each vocoder condition (4, 8, and 16 bands) as a result of added low-pass filtered (LPF) acoustic signal. Specifically, for the 4 and 8 band conditions, the F0 JNDs were reduced by approximately 20 and 21 semitones, respectively, and in the 16 band condition, by approximately 13 semitones. Importantly, there was no significant difference in F0 JNDs in any vocoder condition between adding a 150 Hz and a 300 Hz LPF acoustic signal. Furthermore, there was no significant difference in F0 JNDs between the average for the unprocessed condition and the averages for the acoustic-alone conditions, but also no significant improvement of bimodal hearing over the acoustic-alone conditions.

The results of the post hoc LSD test, also indicated that the VTL JNDs for each vocoder condition (4, 8, and 16 bands) did not significantly change as a result of adding 150 Hz or 300 Hz LPF acoustic signal. Lastly, the average VTL JND was significantly larger in the acoustic-alone condition compared to the unprocessed condition, and as expected, was slightly smaller with the 300 Hz LPF than the 150 Hz LPF.

5 Discussion & Conclusion

The speech signal transmitted via cochlear implants (CIs) is degraded in fine spectrotemporal resolution. An important speech cue that aids higher-level speech perception is voice characteristics, specifically fundamental frequency (F0), determined by the glottal-pulse rate and perceived as pitch, and vocal-tract length (VTL), related to the size of the speaker, perceived as the quality of the sound or timbre. Two experiments were conducted, one in which a neurofeedback training was implemented to improve VTL perception in normal hearing listeners using CI simulations, and the other investigated how the addition of acoustic information in the contralateral ear to the CI, affects F0 and VTL perception in bimodal hearing simulations.

5.1 Experiment 1: Neurofeedback

5.1.1 Preliminary Study - Classifier Performance

Bandt et al. (2009) achieved 88% classification accuracy for target and non-target trials with their P300 classifier. We implemented the same algorithm and tested its performance prior to use during the neurofeedback. The results from the simulated ERP data show that the simple discrimination method can achieve high performance accuracy in spite of a lot of noise.

5.1.2 Main Study

In order for the neurofeedback training to be effective, we required a classification accuracy of at least 70-80%. In this way, the neurofeedback would not be random, and participants would have some influence on the game. Unfortunately, classifier accuracy remained around 50-60%, even after running numerous tests. The most interesting finding was that even pure tones, set up in an oddball task, were not able to elicit the well documented P300 ERP (e.g. Katayama & Polich, 1996).

Initial analysis of the timing of the stimuli and analysing of the EEG data shed light on the fact there was an issue with Psychtoolbox's timing functions and the audioplayer, throwing off EEG data collection and communication between the different computers. This may however only explain part of our results. A study by Furdea et al. (2009) compared a visual and an auditory ERP spelling system. The spelling system was based on previous results showing that the visual P300 ERP can be used to select letters on a computer screen (Farwell & Donchin, 1988). Contrastingly, although both the auditory and the visual system showed the potential for practical written communication, there were some clear differences between the two, especially with regard to the P300 ERP. In the visual system, all 13 participants showed typical P300 ERPs, with an average latency of 349 ms. However, in the auditory speller, only 8 of the 13 participants showed a typical P300 waveform. One major difference was that the mean peak latency of the P300 was almost 200 ms later for the auditory stimuli than the visual. Also, only nine participants using the auditory speller were able to focus their attention in such a way that the ERPs elicited could be reliably detected and classified. As a consequence, users of the the auditory speller had an average accuracy of 65% whereas for the visual speller it was approximately 95%.

These results are important as they highlight some factors that need to be taken into account when using auditory stimuli to elicit the P300 waveform. For example, in the Furdea et al. (2009) study, although the P300 ERPs elicited by the visual stimuli were consistent across participants and at the expected latency, the auditory P300 ERPs were not as consistent across participants. Some users even exhibited negative ERPs to target stimuli, and at varying latencies. Applied to the classifier implemented in our neurofeedback experiment, if this were the case for some participants, the classifier would classify targets as non-targets and non-targets as targets.

Additionally, the P300 classifier we implemented, is designed in such a way that it analyses a pre-defined time window to look for the P300 waveform. If there exist such variations between participants in latencies for auditory stimuli, a more user-centered approach will need to be taken in building the classifier per participant, to take such differences into account. Moreover, studies have found that the more difficult it is to identify the stimuli, the greater the latency of the response (Squires, Donchin, Squires, & Grossberg, 1977). This is an important point when using neurofeedback to train on a just-noticeable-difference (JND). Specifically, our hypothesis was that by setting the target stimuli at the participant's JND they would develop an attentional focus to the stimulus change and eventually improve their JND. In addition to this, the stimuli were vocoded to simulate cochlear implants (CIs). Altogether, this meant that the target stimuli were relatively difficult to identify, which in turn would impact the latency of the response.

Future Directions

P300 neurofeedback training on vocal-tract length (VTL) perception needs to be further investigated. First of all, there has been little research on P300 responses to non-word, speech stimuli. It is possible that the consonant-vowel syllables used as stimuli in our experiment were not eliciting the typical P300 waveform. Second, Chang et al. (2014) used a task which elicited the mismatch negativity (MMN) response rather than the P300. As the MMN is a component of the auditory ERP, it may be a more reliable waveform to track when training on auditory stimuli. Third, although Micco et al. (1995) demonstrated that the P300 can successfully be elicited in CI users with speech stimuli, it is unclear how manipulation of the VTL influences the P300 ERP. This will need to be studied before P300 neurofeedback training on VTL perception can continue. Lastly, if using the P300 with auditory stimuli, due to the variability in the waveform across individuals, future studies aiming to employ the same classifier as we suggested in this study, should explore how adjusting the classifier per participant influences classifier accuracy.

Conclusion

Real-time single-trial analysis of event-related potential (ERP) components for use during neurofeedback is theoretically possible with the simple discriminant analysis method proposed by Bandt et al. (2009). However, to improve vocal-tract length (VTL) perception in cochlear implant (CI) users with the neurofeedback training introduced in this study, the auditory evoked P300, and the ERPs evoked by CI simulated signals will need to be investigated, so that the classifier can be tailored to the specific EEG components varying between individuals.

5.2 Experiment 2: Bimodal Hearing

Bimodal hearing with a cochlear implant (CI) in one ear and a hearing aid in the contralateral ear, provides electric stimulation via the CI to compensate for high frequency hearing loss, while applying acoustic amplification to the low frequency sounds, providing access to the temporal fine structure in the low frequencies. The present study evaluated the benefits of bimodal hearing.

Previous research has shown that bimodal hearing improves speech perception (e.g. Chang et al., 2006; Shpak, Most, & Luntz, 2013). The goal of the present study was to investigate how additional acoustic information in the contralateral ear to the CI, affects fundamental frequency (F0) and vocal-tract length (VTL) perception. Two low-pass filters (LPFs) applied to the acoustic signal were tested, one which supplemented the frequencies covered by the CI (150 - 7000 Hz) perfectly at 150 Hz, and one with overlapping frequencies at 300 Hz. We looked at the effect across different levels of spectral resolution, by applying 4, 8, and 16 band vocoders to the CI simulated signal.

Fundamental Frequency

Our first research question was whether bimodal hearing improved F0 JNDs. As low frequencies contain information about the fundamental frequency, we predicted F0 perception to improve in the bimodal listening conditions. Consistent with previous studies such as Qin and Oxenham (2006) and Sheffield and Gifford (2014), this study demonstrated that low-frequency acoustic sound provides significant benefit when combined with vocoded CI simulations. Specifically, perception of the F0 voice characteristic improved significantly in the bimodal listening conditions compared to the vocoded CI-alone conditions, but not compared to the acoustic-alone conditions. These results show that added low-frequency acoustic sound to the non-implanted ear, produces better F0 perception of speech stimuli in quiet, than is available via a CI alone.

The benefit found in the bimodal conditions was consistent across all levels of spectral degradation, i.e. number of bands in the vocoder. This result is especially interesting as it suggests that even performance comparable to that of the the relatively less proficient CI user (4 bands), can improve as much as it does for the best CI user (16 bands) with the same amount of acoustic information.

Similar to Cullington and Zeng (2010), even unintelligible information below 150 Hz provided significant bimodal benefit. Unexpectedly, average F0 JNDs in all bimodal listening conditions were not significantly different from the baseline/unprocessed condition. This suggests that simulated bimodal hearing provided listeners with as much F0 information as when the signal was unprocessed.

There was no difference in F0 JNDs between the addition of 150 Hz or 300 Hz filtered acoustic signal in all bimodal listening conditions. Both LPFs provided the same bimodal benefit to participants. This was inconsistent with our hypothesis that the 300 Hz LPF would cause interference and suggests that there is no significant loss in spectral or temporal resolution with overlapping frequency maps when a 300 Hz LPF is used in conjunction with a CI simulation transmitting information for 150 - 7000 Hz. This result is in agreement with another study which found that the frequency ranges of the CI and the residual hearing do not need to be spectrally separated to provide maximum benefit (Luo & Fu, 2006). The researchers suggested that this may be due to the reduced spectrotemporal resolution in the CI simulation simply not being susceptible to interference from the high-resolution information provided by the acoustic signal. Another reason could be that listeners ignore the more poorly represented signal when frequencies overlap in bimodal hearing.

Based on previous research by Gaudrain and Başkent (2015), the F0 JNDs recorded here were much larger than would be expected. As Gaudrain and Başkent (2015) reported only on a 6 band vocoder, we are limited to making assumptions about what would be expected with 4, 8, and 16 bands. For a 6 band vocoder with a noise carrier signal, the average F0 JND they found was around 8 semitones. We also used a noise carrier signal in our vocoder, however for the vocoded CI-alone conditions of 4, 8, and 16 bands, the average F0 JNDs were 20.7, 22.1, and 13.5 semitones, respectively. We would have predicted that the F0 JNDs were at least less than 12 semitones, as this should be a large enough difference to hear, and that JNDs were closer to 8 semitones than to 20, for 4 and 8 bands, as this is what Gaudrain and Başkent (2015) found with a 6 band vocoder.

Although each JND measurement started with a 12 semitone difference between stimuli, there were a number of cases, as described above, in which participants' F0 JNDs were above, and relatively close to, the starting difference. This resulted in them not having the same amount of 'training' time leading up to their JND measurement. It was assumed, prior to starting the study, that participants would be trained by the adaptive JND procedure as it supposedly always starts easy and becomes difficult over the course of the measurement. One question this provokes is what influence it has on our results. If these measurements are not representative of the actual vocoded CI-alone thresholds, the bimodal benefit to F0 perception may be much smaller or not statistically significant.

The cause of these very large JND thresholds could be the result of participants having to change strategy for each measurement, and not being prepared for the new condition. This is caused by the conditions being in a random order. If from one measurement to the next they go from acoustic-alone in one ear, to vocoded CI-alone in the other, it may take them time to adjust to the sound.

Furthermore, there were large individual differences in the measured F0 JNDs for the 16 band vocoder CI-alone condition, as shown by the box plot in Figure 9A. The JND measurement for each condition was repeated twice. We found that for the 16 band vocoder CI-alone condition, participants' F0 JNDs often differed drastically from one threshold measurement to the next, as shown in Figure 10.

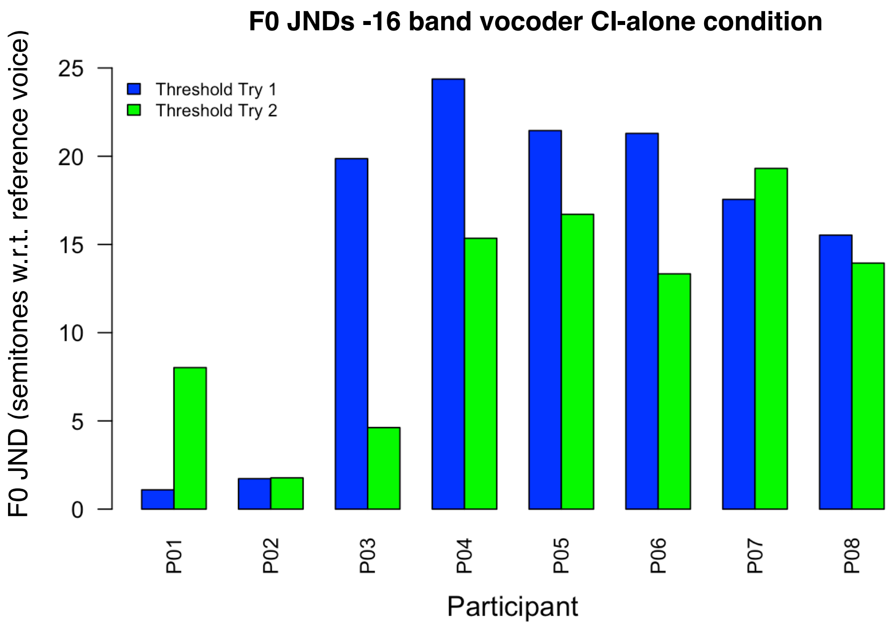


Figure 10: Individual participant F0 JNDs for the 16 band vocoder CI-alone condition.

A small pilot study was conducted, with one participant, to investigate whether a block design of the conditions rather than random would provide more consistent threshold measurements. The initial results of this experiment indicated that there seemed to be no difference between the two designs.

Vocal-Tract Length

As expected, due to the severe loss of spectral resolution with a low-pass filter (LPF), vocal-tract length (VTL) JNDs became significantly larger, compared to baseline, in the acoustic-alone conditions. Similarly, the 300 Hz LPF provided slightly more spectral resolution than the 150 Hz LPF, and consequently produced a slightly smaller average VTL JND, 12.06 compared to 13.98 semitones, respectively.

Interestingly, neither the addition of the 150 Hz, nor the 300 Hz LPF, significantly changed VTL JNDs compared to the CI-alone conditions, across all vocoder bands. By measuring VTL JNDs we were investigating the suggestion made by researchers such as Kong and Carlyon (2007), that low-frequency phonetic cues, such as the first formant, available in the acoustic signal, contribute more to the bimodal benefit than F0. Although first formant cues generally occur above 300 Hz (Hillenbrand et al., 1995), it is possible that due to filter roll-off properties, some weak first formant cues are present below 150 Hz (Cullington & Zeng, 2010). If it were the case that weak first formant information was present in the acoustic signal with a 150 Hz or 300 Hz LPF, and contributed to the bimodal benefit, we would have expected the VTL JNDs to become smaller in the bimodal conditions compared to the vocoded CI-alone conditions. This is because VTL perception relies on formant structure. The results from our study however, do not support the contribution of weak first formants to the bimodal benefit, as no significant differences in VTL JNDs were seen in the bimodal conditions. This suggests that the F0 plays a more important role, at least with added acoustic signal that is low-pass filtered at 150 and 300 Hz, and when listening in quiet to non-word speech stimuli.

Limitations

There are a number of limitations of this study preventing generalisation of the data to the larger population of CI users and those employing bimodal hearing. First of all, the current study is based on a relatively small sample size (n=8).

Second, although bimodal hearing provides both electric and acoustic signals to the listener, similar to electric-acoustic stimulation (EAS), the comparison and integration of the sound signals may be different in the case of ipsilateral EAS. For this reason, the results found in this study cannot be extrapolated to the other population with certainty.

Third, the stimuli used in this study consisted of syllables rather than words, and were presented in quiet rather than noise, so as to keep the situation as simple as possible. Although this is not realistic it allows for minimal additional brain processing besides that required for the conditions we were interested in. As Sheffield and Gifford (2014) found, the low-pass bandwidths required for bimodal benefit differ between quiet and noise, and between sentence and word recognition. Therefore although the conditions tested in this study show a significant effect on F0 perception in quiet, it may not hold for listening in noise, or with meaningful words or sentences as stimuli.

Fourth, simply low-pass filtering sound signals and presenting that to normal hearing listeners cannot fully replicate the residual acoustic hearing available to some CI users. Additionally, the conditions tested here are specifically based on a CI which transmits sound information in the 150 - 7000 Hz region, and simulates residual hearing either below 150 or 300 Hz. The frequency range and placement of the implant, as well as the amount of residual hearing, realistically will depend on the brand of the implant as well as individual sensori/neural hearing loss in both ears.

Lastly, although this type of vocoder, with 8th filter order simulating spread of excitation on the cochlea, provides functional similarity to actual implants, it does not precisely replicate the processes taking place in the device. It is simply a useful tool to show how degraded spectral cues can affect normal hearing individuals (Fuller et al., 2014).

Future Directions

Future research can build upon these initial results and look at the bimodal benefit when perceiving words and sentences instead of non-word speech, as used in this study. Moreover running the measurements in noise, such as multi-talker babble, and with different talker genders, may produce different results, as seen by Sheffield and Gifford (2014).

This study only tested modifications to the fundamental frequency (F0) and vocal-tract length (VTL) in the direction of the male voice with respect the the reference female voice. This was due to the large number of conditions we were already testing, without adding the reverse direction to the battery of conditions as well. However, this meant that in every threshold measurement run, the ‘deviant’ triplet was always the one with lower pitch or what may be perceived as sounding similar to lower pitch in the case of the VTL cue. In the future it would be of interest to see whether the threshold measurements may change if also tested in the opposite direction.

Although VTL JNDs did not improve in the bimodal conditions compared to the vocoded CI-alone conditions, wider LPFs applied to the acoustic signal, may show that weak formant information, available at higher frequencies, provides improvement in VTL perception. Furthermore, F0 may only be important in quiet and not in noise, as Kong and Carlyon's (2007) results were achieved from testing in noise. Therefore, we do not rule out the possibility that both F0 and weak formant structure play roles in bimodal benefit, however it will need to be further tested.

Conclusion

Bimodal benefit was found in the improved perception of fundamental frequency (F0) in quiet, when cochlear implant (CI) simulations in one ear were supplemented by 150 and 300 Hz low-pass filtered (LPF) acoustic sound, presented to the contralateral ear. The bimodal benefit was not affected by the level of spectral degradation in the CI, simulated by varying the number of bands in the vocoder, nor did it improve or deteriorate with increasing low-pass bandwidth above the 150 Hz low-pass filter. The results provide evidence for the role of the F0 present in the acoustic signal, in supporting the enhanced speech recognition performance in bimodal hearing. Furthermore, it suggests that amplification of residual low-frequency hearing as low as 150 Hz can provide bimodal benefit in quiet, and that overlapping frequency maps between the CI and hearing aid does not cause interference of the signals.

References

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*(3), 379-384.

Bandt, C., Weymar, M., Samaga, D., & Hamm, A. O. (2009). A simple classification tool for single–trial analysis of ERP components. *Psychophysiology*, *46*(4), 747-757.

Başkent, D., Gaudrain, E., Tamati, T. N., & Wagner, A. (2016). Perception and Psychoacoustics of Speech in Cochlear Implant Users. In A. T. Cacace, E. de Kleine, A. G. Holt & P. van Dijk (Eds.), *Scientific Foundations of Audiology: Perspectives from Physics, Biology, Modeling, and Medicine* (285-319). San Diego, CA: Plural Publishing, Inc.

Bingabr, M., Espinoza-Varas, B., & Loizou, P. C. (2008). Simulating the effect of spread of excitation in cochlear implants. *Hearing Research*, *241*(1), 73-79.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot international, 5*(9/10), 341-345.

Boëx, C., de Balthasar, C., Kós, M. I., & Pelizzone, M. (2003). Electrical field interactions in different cochlear implant systems. *The Journal of the Acoustical Society of America, 114*(4), 2049-2057.

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology, 34*(5), 260-284.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101-1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110*(5), 2527-2538.

Chang, J. E., Bai, J. Y., & Zeng, F. G. (2006). Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise. *IEEE Transactions on Biomedical Engineering, 53*(12), 2598-2601.

Chang, M., Iizuka, H., Naruse, Y., Ando, H., & Maeda, T. (2014). Unconscious learning of auditory discrimination using mismatch negativity (MMN) neurofeedback. *Scientific Reports, 4.*

Ching, T. Y., Incerti, P., & Hill, M. (2004). Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear and Hearing, 25*(1) 9-21.

Ching, T. Y. C., van Wanrooy, E., & Dillon, H. (2007). Binaural-Bimodal Fitting or Bilateral Implantation for Managing Severe to Profound Deafness: A Review. *Trends in Amplification, 11*(3), 161–192.

Cullington, H. E., & Zeng, F. G. (2010). Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear. *Ear and Hearing, 31*(1), 70.

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America, 114*(5), 2913-2922.

Darwin, C. J., & Hukin, R. W. (2000). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *The Journal of the Acoustical Society of America, 107*(2), 970-977.

Dorman, M. F., & Gifford, R. H. (2010). Combining acoustic and electric stimulation in the service of speech recognition. *International Journal of Audiology, 49*(12), 912-919.

Dorman, M. F., Gifford, R. H., Spahr, A. J., & McKarns, S. A. (2008). The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiology and Neurotology, 13*(2), 105-112.

Fant, G. (1971). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations* (Vol. 2). Walter de Gruyter.

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology, 70*(6), 510-523.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech–reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America, 88*(4), 1725-1736.

Friesen, L. M., Shannon, R. V., Başkent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 110*(2), 1150-1163.

Fu, Q. J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology, 5*(3), 253-260.

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q. J., Free, R. H., & Başkent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology, 15*(6), 1037- 1048.

Furdea, A., Halder, S., Krusienski, D. J., Bross, D., Nijboer, F., Birbaumer, N., & Kübler, A. (2009). An auditory oddball (P300) spelling system for brain–computer interfaces. *Psychophysiology, 46*(3), 617-625.

Gaudrain, E., & Başkent, D. (2015). Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America, 137*(3), 1298-1308.

Gifford, R. H., Dorman, M. F., McKarns, S. A., & Spahr, A. J. (2007). Combined electric and contralateral acoustic hearing: word and sentence recognition with bimodal hearing. *Journal of Speech, Language, and Hearing Research, 50*(4), 835-843.

Greenwood, D. D. (1990). A cochlear frequency–position function for several species—29 years later. *The Journal of the Acoustical Society of America, 87*(6), 2592-2605.

Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *The Journal of the Acoustical Society of America, 113*(5), 2861-2873.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5), 3099-3111.

Katayama, J. I., & Polich, J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. *International Journal of Psychophysiology, 23*(1), 33-40.

Kawahara, H., & Irino, T. (2005). Understanding principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In Divenyi, P. (Ed.), *Speech separation by humans and machines* (167-180). Springer US.

Köbler, S., & Rosenhall, U. (2002). Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification. *International Journal of Audiology, 41*(7), 395-400.

Kong, Y. Y., & Carlyon, R. P. (2007). Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. *The Journal of the Acoustical Society of America, 121*(6), 3717-3727.

Kong, Y. Y., Stickney, G. S., & Zeng, F. G. (2005). Speech and melody recognition in binaurally combined acoustic and electric hearing. *The Journal of the Acoustical Society of America, 117*(3), 1351-1361.

Luo, X., & Fu, Q. J. (2006). Contribution of low-frequency acoustic information to Chinese speech recognition in cochlear implant simulations. *The Journal of the Acoustical Society of America, 120*(4), 2260-2266.

Luo, X., Fu, Q. J., Wu, H. P., & Hsu, C. J. (2009). Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hearing Research, 256*(1), 75-84.

Mackersie, C. L., Dewey, J., & Guthrie, L. A. (2011). Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss. *The Journal of the Acoustical Society of America, 130*(2), 1006-1019.

Micco, A. G., Kraus, N., Koch, D. B., McGee, T. J., Carrell, T. D., Sharma, A., ... & Wiet, R. J. (1995). Speech-evoked cognitive P300 potentials in cochlear implant recipients. *Otology & Neurotology, 16*(4), 514-520.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*(12), 2544-2590.

National Institute on Deafness and Other Communication Disorders. (2014). *Cochlear implants*. Retrieved from http://www.nidcd.nih.gov/health/hearing/pages/coch.aspx

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9*(4), 456-479.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128-2148.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology, 41*(2), 103-146.

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., ... & Lécuyer, A. (2010). Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. *Presence, 19*(1), 35-53.

Ritter, W., & Vaughan, H. G. (1969). Averaged evoked responses in vigilance and discrimination: a reassessment. *Science, 164*(3877), 326-328.

Rubinstein, J. T. (2004). How cochlear implants encode speech. *Current Opinion in Otolaryngology & Head and Neck Surgery, 12*(5), 444-448.

Rubinstein, J. T., & Hong, R. (2003). Signal coding in cochlear implants: exploiting stochastic effects of electrical stimulation. *The Annals of Otology, Rhinology & Laryngology. Supplement, 112*(191), 14-19.

Shannon, R. V., Galvin III, J. J., & Başkent, D. (2002). Holes in hearing. *Journal of the Association for Research in Otolaryngology, 3*(2), 185-199.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304.

Sheffield, S. W., & Gifford, R. H. (2014). The benefits of bimodal hearing: effect of frequency region and acoustic bandwidth. *Audiology and Neurotology, 19*(3), 151-163.

Shpak, T., Most, T., & Luntz, M. (2014). Fundamental frequency information for speech recognition via bimodal stimulation: cochlear implant in one ear and hearing aid in the other. *Ear and Hearing, 35*(1), 97-109.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass, 3*(2), 621-640.

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language, and Hearing Research, 57*(1), 285-296.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America, 118*(5), 3177-3186.

Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America, 116*(2), 1081-1091.

Straatman, L. V., Rietveld, A. C. M., Beijen, J., Mylanus, E. A. M., & Mens, L. H. M. (2010). Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid. *The Journal of the Acoustical Society of America, 128*(4), 1884-1895.

Squires, N. K., Donchin, E., Squires, K. C., & Grossberg, S. (1977). Bisensory stimulation: Inferring decision-related processes from the P300 component. *Journal of Experimental Psychology: Human Perception and Performance, 3*(2), 299-315.

Qin, M. K., & Oxenham, A. J. (2006). Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. *The Journal of the Acoustical Society of America, 119*(4), 2417-2426.

Vashishta, R., Joshi, A. S., Dhawlikar, N. (2015). *Vocal cord and voice box anatomy*. Gest, T. R. (Ed.). Retrieved from http://emedicine.medscape.com/article/1948995-overview

Vernon, D. J. (2005). Can neurofeedback training enhance performance? An evaluation of the evidence with implications for future research. *Applied psychophysiology and biofeedback*, *30*(4), 347-364.

Vestergaard, M. D., Fyson, N. R., & Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *The Journal of the Acoustical Society of America, 125*(2), 1114-1124.

Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: current designs and future possibilities. *Journal of Rehabilitation Research and Development, 45*(5), 695.

Zeng, F. G. (2002). Temporal pitch in electric hearing. *Hearing research*, *174*(1), 101-106.

Zhang, T., Dorman, M. F., & Spahr, A. J. (2010). Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. *Ear and Hearing, 31*(1), 63.