Chapter I

Introduction

Nawal El Boghdady
Chapter I

1. Preface

Imagine meeting up with a friend at a cocktail party. As you arrive, you are greeted by the din of chatter, background music, and clinking cutlery. You start engaging in conversation yourself and, despite the background interference of other talkers, you are still able to understand what your friend is saying. With normal hearing (NH), you can tell your friend’s voice apart from the conversation taking place next to you. In other words, you can perceive the color of your friend’s voice: the pitch, timbre, accent, manner of articulation, etc., which set out your friend’s voice apart from the other speakers in the background. But what if you have impaired hearing? What would the situation be like?

Consider this analogy: if you normally wear glasses, you would realize, for example, how difficult it would be if you were asked to locate, from a distance, a red ball embedded in a sea of red balloons. Without the glasses, the visual scene would appear too blurry such that you would have extreme difficulty telling the outline of the ball from the balloons in the background because they all share similar visual features. The task would become much simpler once you put on the glasses, which will enhance the optics to render the scene much clearer.

This situation is similar to those who have some form of hearing impairment. For hearing-impaired individuals, female voices, for example, may sound alike, especially in crowded settings (similar to identifying the red ball in the sea of red balloons), and without a tool to enhance the scene, this task would be extremely difficult. However, unlike glasses, hearing aids (HAs) and cochlear implants (CIs), which are neuro-prosthetic devices that attempt to restore hearing by electrically stimulating the auditory nerve, do not restore NH. Rather, they provide some sensation of useable hearing, but the acoustic scene would still largely remain blurred. That said, most CI users can understand words and sentences spoken by a single talker in
Introduction

a quiet environment (like trying without the glasses to locate the red ball if it is propped against a white wall). However, situations such as the cocktail party setting described above (Cherry, 1953), become extremely challenging and effortful for CI users (for an in-depth study on listening effort, see Pals, 2016), which can affect their enjoyment of such social gatherings. CI users have anecdotally reported (e.g., “OPCI: Ervaringen van laatdove mensen,” 2018) that they avoid such events and settings altogether, which significantly affects their quality of life. In a survey of 247 adult CI users (Van Hardeveld, 2010), more than 93% of the respondents indicated their satisfaction with their speech understanding in quiet using the CI. However, this proportion fell to below 30% when asked about their satisfaction with speech understanding in noise. In addition, 63% of the respondents stated that they prefer watching TV with subtitles compared to without because of the added effect of background sound effects.

The prevalence of hearing loss is expected to increase in the next few years, rendering the difficulty of speech intelligibility in adverse conditions a pressing problem. The European Federation of Hard of Hearing People recently published the results of a survey conducted on 391 hard-of-hearing participants from 21 European countries, including Cyprus, Denmark, France, the Netherlands, Germany, Spain, Sweden, and the United Kingdom (European Federation of Hard of Hearing People, 2018). The statistics revealed that about 52.9% of the respondents acquired hearing loss (HL) during the most productive years of their professional lives (between the ages of 26 and 55 years). This is compared to only 17.6% who acquired hearing loss at or after 56 years of age. These statistics suggest that environmental factors in addition to work-related stressors may contribute to late onset of HL. Moreover, the rise in number of the ageing population in Europe is also expected to contribute additionally
to the prevalence of HL, such that the number of hard-of-hearing individuals is expected to increase by around 50% in 2031, reaching about 75 million affected individuals in Europe alone (European Federation of Hard of Hearing People, 2015). These findings suggest that it seems likely for the prevalence of acquired HL to increase in the near future, with implications for an increase in the number of implant users.

Hearing in noise is thus difficult for many hearing-impaired people, especially if they are CI users. More specifically, situations involving multiple simultaneous talkers are even more challenging for CI users compared to situations in which the background interference is non-speech noise (e.g., Cullington and Zeng, 2008; Stickney et al., 2004), potentially due to the lack of transmission of fine detail by the implant. For a target speech signal masked by a competing masker speech signal (speech-on-speech; SoS), two masking processes are expected to be involved. The first type, termed energetic masking, is a masking phenomenon occurring at the peripheral auditory system due to the energy overlap between the spectrotemporal components of the two speech signals (Brungart, 2001; Pollack, 1975; see Mattys et al., 2012 for a review). The second type, informational masking (Brungart, 2001; Pollack, 1975; Watson et al., 1976; see Kidd et al., 2008 for a review), occurs due to competition between the target and masker signals along more central pathways of auditory processing, such as when linguistic overlap occurs between the two competing signals. These two masking mechanisms, in addition to the degradations imposed on the signal by CI processing, contribute to the added challenge in SoS intelligibility perceived by CI listeners.

Telling various speakers apart relies on the perception of speaker-specific cues (e.g., Abercrombie, 1967), such as, but not limited to, voice differences, manner of articulation, breathiness, the speaker’s accent, etc., which together give the voice of the
speaker its unique character, or as is referred to in this thesis, the *voice color*. The CI limits the transmission of fine spectral and temporal details in the acoustic signal, which can be thought of as a process that smears the voice colors, or rather, expresses certain pigments and inhibits others. In this dissertation, I focus specifically on voice cues which are derived from the anatomy of the human production system and investigate their possible links to deficits experienced by CI listeners in cocktail-party settings.

2. THEORETICAL BACKGROUND

2.1. Fundamentals of auditory scene analysis

In the case of the ball and balloons example given previously, the visual scene (after optic enhancement), could be decomposed into objects according to the Gestalt principles (Wertheimer, 1923), which dictate that objects in a visual scene are grouped based on their visual similarity and proximity, among other attributes. Figure 1 provides an example of such grouping mechanisms. The top portion presents two overlapping sentences; without the aid of any visual cues to separate the sentences into distinct streams, it is difficult to decipher the message conveyed by either sentence. In the bottom portion, introducing spatial and emphasis cues allows much easier parsing of the content of each sentence.

![Figure 1. Example of two overlapping sentences without (top) and with (bottom) the aid of visual cues becoming A cat sits and I sit too (adapted from Bregman, 1990).](image)
Similarly, in the cocktail party scenario described above, the auditory system analyzes the acoustic scene to draw relevant information about the various sound sources present in a process termed *auditory scene analysis* (Bregman, 1990). The acoustic scene consists of multiple streams of sounds that are spectrally, temporally, and spatially related. An object grouped in this manner forms an auditory perceptual stream (Remez, 2005). In the cocktail-party example above, the individual clinking sounds emitted from the plates and cutlery arise from the same location (the nearby table, for example) and share similar spectrotemporal features. The same holds for the speaker’s voice you are trying to attend to. The auditory system then attempts to *group* such similar sounds together into distinct streams by extracting potential cues from the signal, such as onset and offset cues, temporal modulations, frequency components arising from the spectral decomposition in the cochlea, and spatial location from interaural timing and level differences between the two ears (for a review on grouping cues, see Cooke and Ellis, 2001; Darwin and Carlyon, 1995). These cues are then represented in distinct auditory maps (Moore, 1987) that link place of stimulation along the cochlea to its neural representation along the more central auditory pathways, and can encode, for example, harmonic detection, temporal fine structure information, and pitch and timbre cues (Bregman, 1990; Cooke and Ellis, 2001). In addition, central auditory processing also utilizes prior auditory knowledge acquired through experience to draw further information about the nature of the sound source. In that sense, the auditory system utilizes onset and offset cues to group spectrally-overlapping sounds, such as competing speech, and utilizes spectral differences, such as pitch, to group temporally-overlapping sounds (Bregman, 1990; Carlyon, 2004).

A special type of auditory scene is that in which multiple talkers are speaking simultaneously and is considered to be more
representative of cocktail-party environments (e.g., Assmann and Summerfield, 2004; Bronkhorst, 2000; Brungart, 2001; Duquesnoy, 1983; Festen, 1993; Festen and Plomp, 1990). Because the target (foreground) and masking (background) signals share similar spectrotemporal structure, in addition to possible linguistic content, a background speech masker is expected to yield both energetic and informational masking. Nevertheless, the NH auditory system has been shown to utilize amplitude dips in the fluctuations of the speech masker signal to glimpse portions of the target speech, thus effectively reducing the local signal-to-noise ratio at the locations of the dips in the masker signal. In fact, it has been shown in the literature that as the number of competing talkers increases, thereby diminishing the overall temporal fluctuations present in the masker, the masking effect increases for NH listeners (Miller, 1947). This conclusion was also strengthened by the observation that a single competing talker or amplitude modulated noise provide a weaker masking effect compared to steady-state (unmodulated) broadband noise (Carhart et al., 1969; Duquesnoy, 1983; Festen and Plomp, 1990; Gustafsson and Arlinger, 1994). In addition to these modulation cues, NH listeners also utilize voice differences, such as gender information (Bregman, 1990; Brungart, 2001; Stickney et al., 2004) or pitch differences (e.g., Brokx and Nooteboom, 1982), between two competing talkers to selectively attend to the target speech. These voice cues can carry crucial information about the speaker, such as their physical characteristics and their emotional states (Kreiman et al., 2005).

In contrast, hearing-impaired (Bronkhorst and Plomp, 1992; Carhart and Tillman, 1970; Duquesnoy, 1983; Festen and Plomp, 1990; Gustafsson and Arlinger, 1994; Hygge et al., 1992; Peters et al., 1998) and CI listeners appear to utilize neither such amplitude modulations in the masker, nor voice differences between competing talkers (Cullington and Zeng, 2008; Stickney
et al., 2004). In this dissertation, I focus on the study of voice cues that arise from the anatomical structures of the speech production system as follows.

2.2. Speech production

2.2.1. Source-filter theory of speech production

Figure 2 (A) shows the anatomy of the speech production system. According to the source-filter theory of speech production (Chiba and Kajiyama, 1941; Fant, 1960), speech can be produced when the glottal pulses arising from the rapid opening and closing of the vocal folds (the source shown in green) are filtered by the vocal tract (shown in blue), which acts as an acoustic filter. Air pushed out by the lungs can be converted into a series of pulses as the speaker controls the rate of opening and closing of the vocal folds, producing a sound pressure wave at audible frequencies (Lieberman and Blumstein, 1988). The rate of these glottal pulses (the glottal pulse rate), which is dictated by the length, mass, and tension of the vocal folds, is responsible for eliciting the percept of voice pitch which is often referred to as the fundamental frequency (F0) of the speaker. Voiced speech produced as a result of these pulses contains frequency components at integer multiples of F0 called harmonics.

The glottal pulses are then filtered through the vocal tract of the speaker, including resonances from the nasal cavity, which together serve to amplify certain frequencies (the formant frequencies) and attenuate others. Changes in the formant frequencies dictate the linguistic content of the signal (e.g., vowels), and such changes can be elicited by the movement of the articulators (the lips and tongue) which influence the shape of the vocal tract, thus changing the filtering function. The vocal tract length (VTL), which is measured from the vocal folds to the opening of the lips, is associated with the physical (Fitch and Giedd, 1999; Patterson et al., 2008) and perceived size of
the speaker (Ives et al., 2005; Smith et al., 2005; see Patterson et al., 2008 for a review).

![Diagram of the human speech production system]

Figure 2. Panel A: Sagittal view of the human speech production system, with the source (F0) and filter (VTL) shown in green and blue, respectively (adapted from Tavin, 2011). Panel B: The effect of changing F0 and VTL on the speech signal. The top half represents the shape of the speech waveform for increasing F0 (going from a to b, and c to d) and for shortening the VTL of the speaker (from a to c and b to d). The bottom half shows the effect of increasing F0 and shortening VTL on the spectral envelope of the signal (adapted from Başkent et al., 2016).

2.2.2. Voice cues derived from the source-filter theory

Figure 2 (B) shows the representation of both F0 and VTL in the speech signal for the vowel /a/. The top half shows the waveform representation in the time domain, while the bottom half shows the representation of the signal in the spectral domain. The waveform representation shows the glottal pulses of the vowel: as F0 increases [going from left to right in Figure 2 (B)], the glottal pulses occur more frequently, and thus the temporal spacing between each successive pulse decreases in the time-domain waveform. In the spectral domain, an increase in F0 yields wider spacing between the successive harmonics.
Typically, children have higher F0s compared to adult females, and adult females have higher F0s compared to adult males (Peterson and Barney, 1952). Figure 3 shows a representation of the relative [F0, VTL] space for typical voices of male, female, and children speakers as derived from the data provided by Peterson and Barney (1952) relative to a reference female speaker with an average F0 of about 176 Hz and a VTL of roughly 14.4 cm. Differences in F0 and VTL are represented in semitones (a $12^{th}$ of an octave, and abbreviated as $st$).

![Figure 3](image)

**Figure 3.** Voice space for F0 and VTL differences between typical male, female, and children speakers relative to a reference female speaker provided at the origin of the plane (black cross). The ellipses denote differences across 99% of the population (data replotted from Peterson and Barney, 1952).

Shortening VTL results in a shrinking effect of the glottal pulse resonances in the time-domain waveform of the signal, however, the effect of VTL is more evident in the spectral representation. In the spectral representation, shortening VTL, which corresponds to moving from a taller individual (e.g., an adult) to a shorter individual (e.g., a child), leads to a
stretching effect of the spectral envelope of the signal towards higher frequencies along a linear frequency scale [going from panel e to panel g, and from f to h in Figure 2 (B)]. Elongating VTL would result in the opposite effect, in which the spectral envelope would be compressed towards lower frequencies again along a linear frequency scale. In general, children have shorter VTLs compared to adult females, who in turn, have shorter VTLs compared to adult males (Fitch and Giedd, 1999; Smith and Patterson, 2005). This effect has direct consequences on the formant frequency space defining vowel boundaries (Peterson and Barney, 1952; Turner et al., 2009). When VTL is shortened, the formant peaks in the spectrum are translated towards higher frequencies along a logarithmic frequency scale, thereby changing the individual value of each formant. Likewise, if VTL is elongated, the formant peaks in the spectral envelope would be translated towards lower frequencies along a logarithmic frequency scale. Nevertheless, the auditory system seems to identify a vowel (e.g., /a/) correctly whether it is spoken by a male, female, or child speaker, even though the individual formant values would be quite different. It appears that the auditory system utilizes prior linguistic knowledge regarding the overall vowel pattern rather than the individual formant locations themselves (for a review, see Johnson, 2005), much like it is able to identify a musical chord irrespective of its pitch position (Potter and Steinberg, 1950). In addition, language largely influences the differences between the vowel formants across genders (Johnson, 2005). For example, differences in formant frequencies between male and female talkers are minimal in languages such as Danish, but are quite large in Russian. These data indicate that speaker differences cannot be predicted from anatomical differences of the vocal tract alone.

Nonetheless, while F0 and VTL cues are not the only characteristics for defining the voice of a speaker (Abercrombie,
The focus is given primarily to these two cues because of their direct link with the human anatomy and because manipulations of these two cues were reported in the literature to influence the perceived gender of the speaker (Hillenbrand and Clark, 2009; Skuk and Schweinberger, 2014; Smith and Patterson, 2005). For example, the literature has shown that manipulating both F0 and VTL of a speaker’s voice using the speech processing software, STRAIGHT (Kawahara and Irino, 2005) influenced NH listeners’ perception of the age (child or adult) and size of the speaker (Smith et al., 2007; Smith and Patterson, 2005), in addition to the gender (Fuller et al., 2014; Meister et al., 2016; Smith and Patterson, 2005), and identity (Gaudrain et al., 2009) of the speaker. Moreover, the literature also provided evidence that manipulation of F0 (e.g., Başkent and Gaudrain, 2016; Brokx and Nooteboom, 1982; Darwin et al., 2003; Stickney et al., 2007; Vestergaard et al., 2009) and VTL cues (e.g., Başkent and Gaudrain, 2016; Darwin et al., 2003; Vestergaard et al., 2009) aided NH listeners’ release from speech masking and improved segregation of the two competing talkers into separate streams. In the following section, speech transmission with CIs is described along with evidence from the literature regarding how CI users perceive F0 and VTL cues, and how these patterns of perception may be related to the CI signal processing.

### 2.3. Speech transmission with CIs

In NH [Figure 4 (A)], sound is collected by the pinna (outer ear), travels through the auditory canal, and stimulates the three ossicles in the middle ear, which together perform an impedance matching between the external medium conducting the sound waves (air) and the internal medium within the cochlea (fluid). The movement of the ossicles is translated into stimulation along the basilar membrane within the cochlea (inner ear) in a tonotopic fashion. Figure 4 (A) displays this tonotopic property
of the basilar membrane, such that each location along the basilar membrane is selectively responsive to a certain frequency region: the cochlear apex responds to low frequencies, while the cochlear base responds to higher frequencies. The movement along the basilar membrane at the specific tonotopic frequency then elicits neural stimulation of the auditory nerves. Thus, F0 and VTL cues are expected to be encoded in the location of excitation along the basilar membrane.

![Diagram of the cochlea](image)

Figure 4. **Panel A**: Anatomy of the healthy ear, with cochlea demonstrated in an unrolled fashion (adapted from Munkong and Juang, 2008). **Panel B**: Individual components of a typical CI: 1) Microphone; 2) Battery compartment and signal processor; 3) Radio frequency (RF) transmitter; 4) RF receiver; 5) Pulse generator; 6) Connecting wires; 7) Electrode array; 8) Auditory nerve (adapted from Zeng et al., 2008).

In CI users, this transduction mechanism from a mechanical signal along the basilar membrane into a neural signal is impaired, and is thus accomplished artificially with the aid of the CI. Figure 4 (B) shows the components of a typical CI device. Sound is collected via the microphone and processed by the speech processor worn behind the ear. The speech processor digitizes the acoustic signal and transmits it as a radio frequency (RF) signal to the implanted components, which serve to decode the RF signal into a series of electrical pulses that are transmitted to the electrode array implanted within the cochlea (Zeng et al.,
2008). Figure 5 shows the block diagram of a typical maxima-selection sound coding strategy. The acoustic input is captured by the microphone, preprocessed, and then transmitted to a series of bandpass filters or fast-Fourier analysis block to quantize the audio signal into a series of frequency bands. The temporal envelope of each band is then extracted, and the $n$ bands with the highest energy are selected. The envelopes from the selected bands are then used to modulate a series of pulse trains, which then stimulate the electrodes corresponding to the selected frequency bands.

![Block diagram of CI processing pathway for a typical maxima-selection strategy (adapted from Zeng et al., 2008).]

**Figure 5.** Block diagram of CI processing pathway for a typical maxima-selection strategy (adapted from Zeng et al., 2008).

### 2.3.1. **Spectrotemporal resolution in CIs**

As mentioned in the preface, CI devices do not restore NH. In fact, because CI processing quantizes of the audible frequency range to stimulate a limited number of electrodes, in addition to disregarding the transmission of temporal fine structure and only focusing on the slowly-varying temporal envelope of speech, the spectrotemporal resolution in CI devices is expected to be
impaired (for a review on spectral resolution in CIs, see Başkent et al., 2016; Friesen et al., 2001; Fu et al., 1998; Henry and Turner, 2003; Winn et al., 2016). The literature has shown that CI users, on average, do not have more than 8 effective spectral channels (e.g., Friesen et al., 2001; Qin and Oxenham, 2003), even though the implanted electrode array is usually comprised of a larger number of physical electrodes [e.g., 22 electrodes in a Nucleus system (Cochlear Ltd., Sydney, Australia), and 16 electrodes in a HiRes system (Advanced Bionics Corp., Stäfa, Switzerland)]. This is attributed to the side-effect of electrical stimulation inside the cochlea, which can induce cross-talk, more commonly referred to as channel-interaction, between neighboring electrodes (Boëx et al., 2003; De Balthasar et al., 2003; Hanekom and Shannon, 1998; Shannon, 1983; Townshend and White, 1987). Moreover, the frequency partitioning boundaries in the bandpass filterbank are seldom optimally matched with their corresponding tonotopic locations along the basilar membrane (e.g., Başkent and Shannon, 2004), however, they are not customized in the clinic for each CI user individually (Fitzgerald et al., 2013; Landsberger et al., 2015; Tan et al., 2017; Venail et al., 2015).

Figure 6. Spectrograms obtained for the Dutch sentence “We kunnen weer even vooruit” [We can move forward again] shown for the unprocessed condition (left panel), vocoded condition with 8 channels (middle panel), and vocoded condition with 4 channels (right panel).
Figure 6 shows the spectrograms for a sample Dutch sentence spoken by a female speaker. The left panel indicates the unprocessed version and demonstrates the fine spectrotemporal features present in the acoustic signal. The effect of CI processing on the signal can be investigated by processing the signal using a vocoder (Dudley, 1939). The use of such vocoder simulations of CI processing with NH listeners (Shannon et al., 1995, 1998) is a widely-used method in the literature to allow better control and manipulation of the spectral and temporal resolution in the output signal (e.g., El Boghdady et al., 2016; Fuller et al., 2014; Gaudrain and Başkent, 2015; Qin and Oxenham, 2003, 2005; Stickney et al., 2007). For example, in the middle and right panels of Figure 6, the effective number of spectral channels was manipulated to observe its effect on the signal. From these manipulations, CI processing appears to degrade the fine spectrotemporal features of the speech signal, which, in turn, is expected to affect voice cue transmission in CI devices. In fact, using such simulations with NH listeners has demonstrated that distortions introduced by CI processing could impair the perception of F0 and VTL cues. As an example of such studies, Gaudrain and Başkent (2015) demonstrated that decreasing the number of vocoder channels, simulating a smaller number of effective spectral channels, yielded a significant deterioration in the sensitivity to VTL cues but not F0. Fuller et al., (2014) also showed that vocoder simulations impaired NH listeners’ ability to utilize VLT cues to categorize the gender of the speaker. Furthermore, Stickney el al., (2007) also showed that degrading the spectral resolution of the signal using vocoder simulations hindered the benefit in target speech recognition from F0 differences between masker and target speakers compared to the unvocoded condition.

2.3.2. Voice cue perception with CIs

When compared with NH listeners, who can discriminate
differences in F0 and VTL of at least 1.95 st and 1.73 st, respectively, CI users were shown to be less sensitive to such small differences in F0 and VTL, with average thresholds of about 9.19 st for F0 and 7.19 st for VTL (Gaudrain and Başkent, 2018). These thresholds indicate the smallest difference that can be detected between two stimuli differing in F0 or VTL, and thus, the smaller the threshold, the more sensitive the listener. Linked to this data is the finding that CI users demonstrate impaired gender judgements based on F0 and VTL cues compared to NH listeners (Fuller et al., 2014; Meister et al., 2016). For example, Fuller et al., (2014) provided evidence that CI users only rely on F0 differences to judge the gender of the speaker, while NH listeners make use of both F0 and VTL cues together to perform the same task.

As pertains to SoS perception, while NH listeners were shown to gain an increase in target speech intelligibility as the voice separation between target and masker speakers increased (Başkent and Gaudrain, 2016; Brokx and Nooteboom, 1982; Darwin et al., 2003; Drullman and Bronkhorst, 2004; Vestergaard et al., 2009), CI users were shown not to draw such benefit (e.g., Cullington and Zeng, 2008; Stickney et al., 2004, 2007). Together, these findings indicate that such a deficit in voice cue perception by CI users may be linked to the CI signal processing, and if so, could be addressed by optimizing the signal processing pathway. Thus, the aim of this dissertation is to investigate the links between such a deficit in voice cue perception and the underlying CI signal processing operations.

3. **Study Aims of this Dissertation**

   The main aim of this dissertation is to assess the relationship between voice cue perception, SoS performance, and CI signal processing by addressing the following research questions, to each of which a separate chapter was dedicated:

   1. Is sensitivity to F0 and VTL cues related to CI users’
Chapter I

SoS performance?
2. If so, could that relationship be influenced by the amount of inherent channel interaction in the implant?
3. If channel interaction is found to influence the perception of such cues, can advanced signal processing techniques that enhance the spectral content of the signal help improve the perception of such cues?
4. In addition to optimizing the signal processing strategy, can a signal processing parameter like the frequency-to-electrode quantization map improve the perception of such vocal cues, specifically VTL?

The research questions were addressed in the individual chapters as follows:

- In Chapter 2, the first research question was addressed by utilizing three measures of F0 and VTL perception. SoS intelligibility of the target speaker was measured both in NH and CI users as a function of the F0 and VTL difference between the two competing talkers. An additional SoS perception test was also used to assess overall sentence comprehension as a function of the F0 and VTL difference between the two speakers. This SoS comprehension test was used in order to tap more closely to the CI users’ psychophysical function in case the first SoS task yielded performance close to floor levels. Additionally, this second SoS task allowed the capture of both comprehension accuracy and speed, which together could yield more information about the difficulty level of the task compared to accuracy measures alone. Finally, F0 and VTL just-noticeable-differences (JNDs) were measured, which aim to quantify the participants’ sensitivity to differences along those two voice cues. The correlations between the performance on both SoS tasks and the JND task were investigated.

- In Chapter 3, the second research question was addressed by investigating the effect of channel interaction on SoS perception and the sensitivity to F0 and VTL differences using the same tests as in the previous chapter. Because simulated channel interaction was found to affect
the sensitivity to VTL cues (Gaudrain and Başkent, 2015), the hypothesis was that it could also affect SoS perception and voice cue sensitivity in actual CI users. The channel interaction was implemented in actual CI systems by deploying three electrode stimulation techniques simulating low, medium, and high levels of channel interaction. This was done by stimulating one, two, and three simultaneous channels, respectively. The larger the number of simultaneously-stimulated channels, the larger the current spread and resulting channel interaction. The hypothesis was that increased channel interaction would be expected to negatively impact both SoS perception, in addition to voice cue sensitivity, especially those related to VTL.

- In Chapter 4, the third research question was addressed by investigating whether enhancing the spectral contrast of the signal using a spectral contrast enhancement strategy (SCE) could improve the sensitivity to voice cue differences and their relationship with SoS perception. This was also obtained using the same tests as in Chapter 2. The SCE strategy deployed in this chapter served to enhance the contrast between the peaks and troughs in the spectral envelope, thereby enhancing the resolution of individual formants.

- In Chapter 5, the fourth research question was addressed in which the effect of varying the frequency-to-electrode allocation map on the sensitivity to VTL differences was investigated in vocoder simulations of CI processing. Because of the large space of possible setups for the frequency-to-electrode allocation map, vocoder simulations were deployed as a first step to better control the parameter settings before testing with actual CI users.

- Finally, in Chapter 6, an overarching discussion is presented of the findings of this dissertation.
Chapter I

REFERENCES


Introduction

273–279.


Duquesnoy, A. (1983). “Effect of a single interfering noise or speech source...
Chapter I


Introduction


Chapter I

118, 3816–3822. doi:10.1121/1.2118427


Chapter I


Chapter I