The processing of Dutch prosody with cochlear implants and vocoder simulations

Velde, D.J. van de

Chapter **1**

Introduction

An estimated 360 million people (over 5% of the population) suffer from hearing loss worldwide, according to an estimate by the World Health Organization ("Deafness and hearing loss," 2015). The prevalence of deafness in the Netherlands is approximately 0.7% as of 2016 (Lamoré, 2016). Hearing loss (presumed equivalent to the impairment associated with being 'hard of hearing') is defined as a "hearing disorder, whether fluctuating or permanent, which adversely affects an individual's ability to communicate" and deafness as "a hearing disorder that limits an individual's aural/oral communication performance to the extent that the primary sensory input for communication may be other than the auditory channel" (American Speech-Language-Hearing Association, 1993). Possible causes of hearing loss are hereditary and acquired. Among hereditary causes, the most prevalent is connexin-26 deficiency (DFNB1) in the GJB2 gene. Possible syndromal hereditary causes are Waardenburg's syndrome and Usher's syndrome. Among the acquired causes are meningitis and reactions to ototoxic drugs. One to two in every thousand children is born with bilateral sensorineural hearing loss (Gravel & Tocci, 1998).

Hearing loss can have different repercussions for individual listeners. Following the WHO's International Classification of Functioning, Disability and Health (Stephens & Kerr, 2000), they experience problems detecting, recognizing and identifying sounds, appreciating sound quality, tolerating loud sounds, understanding speech in silence and noise, understanding spoken emotions, and localizing sound sources. Moreover, their education and career opportunities are compromised (Lang, 2002). Neurocognitive effects of (untreated) auditory deprivation have also been reported, such as problems with working memory (Marschark, Lang, & Albertini, 2002) and socio-emotional control (such as psychopathology; Theunissen, 2013), cognitive decline in older listeners and degradation of auditory cortex and its takeover by the visual modality (Glick & Sharma, 2016).

The observations above demonstrate the severity of the problem of hearing loss, both at the level of the individual listener and at the level of global socio-economic functioning. A variety of medical interventions are available to treat hearing loss, such as conventional hearing aids (sound amplification), bone-anchored hearing aids (BAHA; sound conduction through bones) and cochlear (CI) and auditory brainstem implants (ABI), the suitability of which depends, among other factors, on the severity and type of an individual's hearing loss. This thesis focuses on the cochlear implant.

The remainder of this chapter consists of sections introducing core aspects of this thesis. Section 1.1 discusses the goal and history of cochlear implantation and the mechanism behind cochlear implant hearing. Section 1.2 describes the phenomenon of prosody, the aspect of speech which forms the linguistic focus of this dissertation. Section 1.3 covers the distinction between perception and production of speech, both of which are investigated in this dissertation. Section 1.4 focuses on the acquisition of language by children with CIs, as a subset of the studies reported in this dissertation involve that population. Section 1.5 briefly discusses the usage of vocoders for research into CI hearing, a method that was adopted in three of the

studies in this thesis. Finally, an overview of the chapters of this thesis and the corresponding perspectives and hypotheses regarding the processing of prosody with CIs is provided.

1.1 Cochlear implants

Cochlear implants are prostheses of the inner ear partially restoring hearing for severely to profoundly deaf children and adults by providing an electrical reconstruction of sound directly to the auditory nerve. The basic functioning of a CI is based on the vocoder technique (see section 1.5). The functioning involves capturing of sound by a microphone attached near the outer ear, signal analysis by a speech processor, transmission of the processed signal to a transmitter attached to the scalp and subsequent electromagnetic transcutaneous transmission to a receiver on the inside of the skull, and finally to a set of between 12 and 22 electrodes inserted into the cochlea. The array of electrodes mimics the tonotopic organization of the basilar membrane by presenting lower frequencies with electrodes situated at the apical end of the cochlea and higher frequencies at the basal end of the cochlea. Of the many design options that exist, some of the more important ones concern the number of channels (electrodes), the shape of the analysis and synthesis filters, the rate and configuration of stimulation, and the position of the array in the cochlea. Detailed descriptions of CI design and functioning have been provided elsewhere (Wilson & Dorman, 2009).

Cochlear implantation has first been performed by Parisian electrophysiologist André Djourno and otolaryngologist Charles Eyriès in 1957 on a deaf patient (Djourno & Eyries, 1957; Eisen, 2009). With their single-channel implant, the recipient was able to discriminate lower from higher frequencies and environmental sounds but had no speech understanding beyond a small number of words. Otologists William House in Los Angeles, Blair Simmons at Stanford University, and Robin Michelson at the University of California-San

Francisco (UCSF) independently pursued this work with single-channel implants in the 1960s, allowing useful hearing sensations to deaf patients but also encountering issues with biocompatibility of the device. Concerns were raised by scientists regarding the feasibility of electrically reconstructing a signal as complex as that of speech (Jongkees, 1978; Lawrence, 1964; Simmons, 1966). However, in 1975, the National Institutes of Health (NIH) acknowledged the benefits of CI by showing improvements in speech production, lip reading and quality of life, spurring further research and its financial support (Bilger, 1977). Scientists at the UCSF, as well as Graham Clark at the University of Melbourne in Australia developed multichannel CIs, which later became the now commonly used Advanced Bionics Clarion and Cochlear Corporation's Nucleus devices, respectively. In the 1980s, Food and Drug Administration (FDA) approvals were granted for adult CI recipients and children as young as two years of age, allowing research to shift from safety to outcome issues. In 1991, the now common continuous interleaved sampling (CIS) strategy, a design whereby electrodes are never activated simultaneously to reduce channel interactions, was shown to further improve speech understanding (Wilson et al., 1991). Since then, a large variety of implant designs and speech coding strategies have been developed and the scientific and social acceptance of CI have grown considerably (Blume, 1999; Christiansen & Leigh, 2004; Wilson & Dorman, 2008).

The primary aim of CIs is to allow speech understanding. The candidacy criteria for cochlear implantation are multifaceted, evaluated on a case by case basis and differ per country, but in general some of the important eligibility criteria for CI are (i) that individuals and their relatives have realistic expectations of its benefits; (ii) that they are motivated to undergo the surgical procedure and persevere the ensuing rehabilitation; (iii) that they benefit less from conventional hearing aids; and (iv) that there is an absence of medical contra-indications, such as inner ear malformations. On the basis of these criteria, as much as 40% of cases presented led to specialists' decision

not to proceed to implantation in the United Kingdom between 1990 and 1994 (Summerfield & Marshall, 1995). However, due to improved implant technologies and benefits, candidacy criteria have become less stringent over the last decades (Niparko, Lingua, & Carpenter, 2009). The number of CI recipients have grown exponentially since these developments, with over 300,000 users worldwide as of 2014 (Wilson, 2014) and over 6,500 in the Netherlands as of 2015 ("Aantal implantaties in Nederland," 2016). Based on research showing that postlingually deafened adults improved their hearing scores after implantation as they showed up to 80% preimplantation phoneme perception scores, the Leiden University Medical Center's ENT department decided to adopt this preimplantation score as an upper limit for CI indication, as even higher scores provided no benefit of cochlear implantation (Snel-Bongers, Netten, Boermans, Briaire, & Frijns, submitted).

CIs have proven successful in allowing recipients to develop or process spoken language more efficiently than deaf children with a conventional hearing aid (Knoors, 2008; Lenden & Flipsen, 2007). This was shown by a number of outcomes: (i) a vocabulary growth at about 60% of normally hearing (NH) children's rate (Blamey et al., 2001; Geers, 2003); (ii) the production of longer sentences (Geers, 2003); (iii) improved sentence understanding (Geers & Moog, 1994); (iv) improved phoneme production (Geers & Moog, 1994); (v) speech perception abilities in quiet conditions within the norms of normally hearing individuals and communication over the telephone (Beadle et al., 2005); (vi) improved reading skills (Johnson & Goswami, 2010); and (vii) improved production of narratives (Boons et al., 2013; Crosson & Geers, 2001). Implantation can also allow participation in mainstream education and favorable career opportunities (Spencer, Gantz, & Knutson, 2004); however, those results are inconclusive and particularly mixed due to individual variation (Marschark, Rhoten, & Fabich, 2007; Punch & Hyde, 2011; Stacey, Fortnum, Barton, & Summerfield, 2006; Thoutenhoofd, 2006). The effects of CI on quality of life have so far also been inconclusive due to theoretical and

methodological inconsistencies and results between studies (Knoors, 2008). Nevertheless, with the above facts and figures about the device's psychophysical merits taken together, the CI could count as the most successful artificial sensory prosthesis.

Despite these merits, CI hearing faces a number of challenges. The input is degraded relative to normal hearing as a result of, among other factors, a limited number of effective electrodes, channel interactions, the single-sided character of the hearing (in case of unilateral implantation), possible cochlear malformations and dead regions of the auditory nerve, malfunctioning electrodes, and frequency shifts due to shallow electrode insertion depths (Wilson & Dorman, 2009). Of the three main dimensions that the auditory signal is composed of – the temporal, the dynamic and the pitch dimension – variations in the pitch dimension and, to a lesser degree, in the dynamic dimension are difficult to discriminate for CI recipients (Meister, 2011; Shannon, 2002). In the perception of speech, NH listeners rely on some dimensions more than others, depending on the listening task. Reliance means that when a dimension is unavailable for whatever reason, this compromises the recognition of the linguistic information in the speech signal. When a dimension provides information about speech, it is referred to as a 'cue' and the relative reliance by listeners on the dimensions as 'cue weighting'.

Due to CI users' perception difficulties, the voice's pitch (fundamental frequency or F0) and, to a lesser extent, the intensity dimensions pose notorious problems for them, prompting them to weight cues differently than normally hearing people do by balancing their reliance from F0 cues (partly) towards temporal and dynamic cues. These input and sound processing issues compromise their music perception, speech perception, spectral resolution, sound source localization, hearing in noise, the perception of acoustically less prominent morphosyntactic endings in languages such as Dutch and English, such as the suffix *-t* in *werkt* (third person singular of 'to work') which is non-syllabic and short (Hammer, 2010; Nikolopoulos, Dyar, Archbold, & O'Donoghue, 2004; Svirsky, Stallings, Lento,

Ying, & Leonard, 2002), and more general capacities such as verbal working memory and serial data recall (Nittrouer, Caldwell-Tarr, & Lowenstein, 2013; Pisoni, Kronenberger, Roman, & Geers, 2011). In view of these possible consequences, cue weighting is further studied in this thesis.

Linguistic performance by CI users notoriously shows much individual variation (Kane, Schopmeyer, Mellon, Wang, & Niparko, 2004; Peterson, Pisoni, & Miyamoto, 2010), begging the question what factors underlie those differences. For instance, performance on recognition of monosyllables ranges between almost zero percent correct to ceiling level after two years of implant experience, with standard deviations up to 30% (Wilson, 2006). The factors underlying this variation can be divided into demographic factors, psychosocial factors, device factors and neurocognitive factors. Demographic factors are factors such as the duration of hearing loss before implantation, the age at implantation (whereby children implanted at two years or younger tend to outperform the later-implanted children), the duration of implant usage and the family's socio-economic status and size (Anderson et al., 2004; Boons et al., 2012; Colletti, Mandalà, Zoccante, Shannon, & Colletti, 2011; Geers, Nicholas, & Sedey, 2003; Harrison, Gordon, & Mount, 2005; Leigh, Dettman, Dowell, & Briggs, 2013; McConkey Robbins, Green, & Waltzman, 2004; Niparko et al., 2010; Sharma, Dorman, & Kral, 2005; Sharma et al., 2004). Psychosocial factors include the presence of additional disabilities such as mental, emotional and social problems (Edwards, 2007; Shin et al., 2015). Device factors are factors such as the number of electrodes, the analysis and synthesis filter's shape, and the array's insertion position (Geers, Brenner, & Davidson, 2003). Finally, among the neurocognitive factors are (verbal) working memory, and intra- (auditory) and cross-modal (visual) neural reorganization due to auditory deprivation (AuBuchon, Pisoni, & Kronenberger, 2014; de Hoog et al., 2016; Finke, Buchner, Ruigendijk, Meyer, & Sandmann, 2016; Nittrouer et al., 2013; Pisoni, 2000). Of these, the duration of hearing loss, age at implantation, and the duration of implant usage, as

well as socio-economic status tend to surface as some of the main predictors of language performance outcome after implantation (Blamey et al., 2013; Holden et al., 2013; Moon et al., 2014). Although many factors have been identified, the individual variation is still not fully understood.

1.2 Prosody

Prosody is speech content that cannot be predicted from the information of individual segments or the coarticulation of subsequent segments (Lehiste, 1970; Rietveld & van Heuven, 2009). It is primarily conveyed by means of variations in F0, intensity and durations of any structural level of an utterance. The functions of prosody can be divided into linguistic, on the one hand, and emotional and indexical functions, on the other (Rietveld & van Heuven, 2009; Witteman, van IJzendoorn, van de Velde, van Heuven, & Schiller, 2011). Linguistic prosody pertains to information about the meaning of an utterance, such as phrasing by means of pauses, lengthening and intonation, word stress, information structure by means of pitch accents (the marking of new vs. known information in sentences) and sentence type (statement vs. question). Emotional and indexical prosody convey information about the emotion or attitude (e.g., irony) and demographics, such as identity, gender, age, dialect and health, of the speaker. The importance of emotion understanding in speech has been highlighted by research pointing to a correlation between emotional identification capacities, but not word identification scores, and quality of life (Schorr, Roth, & Fox, 2009).

A third type of prosody, which is not usually acknowledged independently in the literature, could be called basic prosody. Basic prosodic measures have no linguistic, emotional or indexical function. If anything, they could have an emotional or indexical function, but that is only relevant when it has been shown that changes in the parameters correlate with emotion or speaker identification scores in a

listening task. Without such demonstrated function, between-speaker and between-utterance variations could be considered ‘basic’, possibly stochastic prosodic variations. For instance, utterance duration or F0 declination could serve to infer emotion or speaker characteristics, but when such a link is not established, those measures would still count as basic. In Chapter 2 of this thesis, such basic prosodic measures were compared between speech of CI users and NH peers. Measures that would appear to distinguish between the two groups, could then be considered indexical prosodic measures.

Given this central role of prosody in development and usage of language together with CI users’ perceptual problems, it becomes clear that by missing out on important prosodic information such as information structure and indexical (speaker) information (Gilbers et al., 2015; Massida et al., 2011; Meister, Fursen, Streicher, Lang-Roth, & Walger, 2016), this group of language users is at risk of late and/or deviant language acquisition (Chatterjee & Peng, 2008; Giezen, Escudero, & Baker, 2010; Kong, Cruz, Jones, & Zeng, 2004). This warrants further research into the questions of what types of information are available to CI users, what the mechanism behind their capabilities and limitations is, how children acquire prosody, and if a limitations in perception have repercussions for production. This thesis intends to fill in some of these gaps. The last of these issues is discussed in the next section.

1.3 Speech perception and production

The relationship between speech perception and production can be approached from at least two different angles, that of its development influenced by a speaker’s hearing history (this could be called the ‘diachronic’ perspective) and that of its functioning during speech processing (the ‘synchronic’ perspective). First of all, the development of the relationship between speech perception and production seems in part to depend on an individual’s hearing history. For instance, both

congenitally deaf speakers (Osberger & McGarr, 1982) and speakers with acquired deafness (Waldstein, 1990) produce deviant speech, showing that deficient input has ongoing consequences for the output, even after the supposed establishment of an articulation routine. Speakers with acquired deafness, however, continue to produce normal speech for some time following the onset of deafness, which indicates that the acquired articulatory goals are robust enough to support proper production for some time without direct auditory feedback (Guenther, Ghosh, & Tourville, 2006).

Second, the functioning of the relationship between speech perception and production has been modeled by the Directions Into Velocities of Articulators model (DIVA; Guenther, 2006). In this model, which is based on neurolinguistic evidence, articulatory actions are viewed as motor programs for sound, syllables or sequences of syllables. These actions feedforwardly project system-internal abstract predictions of the structure to be produced, against which the auditory feedback provided by the actual output is checked for adequacy. In case of an inadequate output, an error is detected and the feedforward commands are updated. The output can for instance be inadequate as a result of disruption or feedback delay during articulation (Burnett, Freedland, Larson, & Hain, 1998; Perkell et al., 2007; Purcell & Munhall, 2006), because the speaker is still acquiring speech, because the speaker's articulators are still maturing, or because of deafness. The adequacy of the output is based on speech input provided by ambient speech. Deafness, therefore, may result in deviant speech because inappropriate sound structure representations have been established. The process of the evaluation of speech output against internal representations has been labeled 'monitoring' by other researchers (Levelt, 1983).

Together, the above observations suggest that proper speech perception is required for proper speech production and possibly vice versa as well. The 'diachronic' and 'synchronic' aspects of the relationship between perception and production capabilities are both relevant to this thesis, because children with cochlear implants by

definition have an abnormal hearing history and because their auditory input, even if stable since implantation, is degraded in relation to that of normally hearing individuals. Given the possible relationship between language perception and production capabilities, in combination with CI users' deviant perception performance and life history, it is therefore plausible to assume that CI recipients' production performance is also deviant. This hypothesis is tested in Chapter 6 of this thesis.

A number of studies have probed the possible correlation between perception and production in CI users. Peng (2005) tested school-aged children on the perception and production of the intonation of sentence type (declaratives vs. statements) and found that children with a good tone production also showed a good tone perception, but not necessarily vice versa, suggesting that for CI children good perception precedes good production. According to Peng (2005), the observations might reflect an indirect relationship between perception and production, in that other factors, such as age at implantation, might differentially underlie perception and production. In a series of experiments, O'Halpin (2010) tested prosody perception and production performance of school-aged children with and without cochlear implants. The participants indicated (a) whether utterances were pronounced as compounds or phrases (e.g., *greenhouse* vs. *green house*), (b) which of two words in a sentence carried focus (*It's a GREEN door* vs. *It's a green DOOR*, where capitals mark focus) or (c) which of three words carried focus (*The DOG is eating a bone* vs. *The dog is EATING a bone* vs. *The dog is eating a BONE*). In another experiment, the participants' production of these phrases was evaluated for appropriateness by a panel of NH listeners. The author reported no correlations between most of the perception and production scores. In a study on 47 primary-school-aged children with cochlear implants and 40 peers with hearing aids, Blamey et al. (2001) found a correlation between word and sentence comprehension performance, on the one hand, and intelligibility measures of spontaneous utterances, on the other hand. Speech

intelligibility scores in prelingually deafened CI users predicted post-implantation speech perception scores, whereas preimplantation speech perception scores with hearing aids constituted a weaker predictor (van Dijkhuizen, Beers, Boermans, Briaire, & Frijns, 2011; van Dijkhuizen, Boermans, Briaire, & Frijns, 2016). Other studies have shown mixed results regarding the correlation between perception and production by CI children, such as a lack of correlation between the Beginner's Intelligibility Test (Osberger, 1994) and the Prosodic Utterance Production test (Bergeson & Chin, 2008) or a correlation between emotion imitation and recognition (Lyxell et al., 2009; also see, Spencer et al., 2004).

These studies together demonstrate that it is at present unclear to what extent perception and production of speech are correlated in children with cochlear implants, as was also concluded in a recent review (Cysneiros, Leal, Lucena, & Muniz, 2016). This thesis joins this debate by studying perception and production of two types of prosody (linguistic and emotional) by CI children controlling for general linguistic and emotional maturation. The next section discusses the general background for this thesis regarding language acquisition by implanted children.

1.4 Language acquisition by children with cochlear implants

Language acquisition is thought to start as early as approximately three months before birth, when the fetus perceives mainly relatively loud and low-frequency (under 1000 Hz) environmental, bodily and some speech sounds from the mother (Graven & Browne, 2008). This is evidenced by newborns' preference for the maternal language over other languages (Mehler et al., 1988; Moon, Lagercrantz, & Kuhl, 2013). Auditory experience further shapes the very early stages of language acquisition by means of infant-directed speech, perceptual tuning in the first 6 months of life (Kuhl et al., 2006; Werker & Tees,

1984), and by guiding the perception of focus, syntactic information and phrase boundaries (Soderstrom, Seidl, Nelson, & Jusczyk, 2003).

Prosody plays a special role in acquisition. As a result of prenatal imprinting, newborns show a preference for native over non-native prosody, showing that the speech information has been processed (Moon, Cooper, & Fifer, 1993). Due to the intrauterine frequency selectivity, the speech sounds that penetrate are mainly prosodic, i.e., rhythmic and intonational. After birth, 'motherese' (prosodically exaggerated child-directed speech by caregivers) draws infants' attention to important components in speech (Liu, Kuhl, & Tsao, 2003; Thiessen, Hill, & Saffran, 2005). Prosody continues to play a pivotal role in language acquisition in the following months and years. At the age of approximately seven months, infants use prosodic patterns to segment the speech stream. Prosody thus paves the way for word learning (Johnson & Jusczyk, 2001). We can therefore conclude that the development of prosody starts early, probably forming the first stage in language acquisition, and proceeds to play an essential role in children's language acquisition until the young-adolescent age.

Given the importance of hearing experience for early language acquisition, it is not surprising that language acquisition develops differently in children with hearing loss. Most deaf children have two hearing parents (Mitchell & Karchmer, 2004), and consequently do not receive native sign language input. Deaf children can have delayed canonical onset and a restricted repertoire of babbling (Kuhl & Meltzoff, 1996; Oller & Eilers, 1988). They possibly do not catch up with NH peers (Vaccari & Marschark, 1997). This inability to catch up after a delay despite intensive efforts is thought to be due to a sensitive period in acquisition, i.e., an age window during which acquisition has to start in order to be able to reach a normal level as the end stage (Lenneberg, 1967; Werker & Hensch, 2015).

Congenitally deaf children with cochlear implants present an interesting case of atypical language development, since they experience a clear-cut delayed onset of spoken language acquisition, while enjoying – in most cases – a normal upbringing. For

congenitally deaf implanted children, the onset of spoken language acquisition coincides with the activation of the implant (Connor, Craig, Raudenbush, Heavner, & Zwolan, 2006; Tye-Murray, Spencher, & Woodworth, 1995). The study of pediatric CI recipients therefore allows the investigation of the effect of a delayed onset on language acquisition and the role of early non-linguistic maturation. CI children's language acquisition is delayed and can also be deviant relative to that of NH peers (Geers, Nicholas, Tobey, & Davidson, 2016; Robinson, 1998). Cochlear implantation improves speech production but after several years of implant usage, in some recipients, it still deviates from that of NH peers (Geers, Tobey, Moog, & Brenner, 2008).

Despite these differences, several studies observed a similar prosodic development in CI and typically developing (TD) children (Snow & Ertmer, 2009, 2012; Vogel & Raimy, 2002; Wells, Peppé, & Goulandris, 2004). Snow and colleagues (Snow & Ertmer, 2009, 2012) modeled children's intonational development until 24 months of age in terms of stages in F0 range on word accents. They found that CI children matched TD children's alternation between stages of increased and decreased pitch range. However, the CI recipients' development shows an interaction between implantation age and duration of implant usage, whereby children implanted after 24 months of age showed a development that was more advanced than would be expected based on their hearing age (i.e., the time since implantation) and whereby children implanted before 24 months of age showed a delay in their development. This suggests that maturation plays a role in prosody development in that some components of it continue without auditory input.

In one of the experiments in this study, long-term effects of cochlear implantation on emotional and linguistic prosody perception and production are investigated by comparing school-age CI with NH children. Apart from probing possible deviations or delays in the acquisition of these four quadrants of prosody processing (linguistic prosody production, linguistic prosody perception, emotional prosody

production, and emotional prosody production) and the correlations between them, we test the hypothesis that emotional prosody is less delayed than linguistic prosody because the former is supposedly less dependent on rule-learning derived from input than the latter.

1.5 Vocoders

Sound processing in cochlear implants is based on the channel vocoding technique. Channel vocoders (short for voice encoder) are signal processing algorithms designed to reconstruct a sound signal in a parametrized way. The signal processing procedure follows two basic steps: analysis and resynthesis. In the analysis step, incoming sound is band-pass filtered into a number of contiguous frequency bands (channels). In the resynthesis step, the signal is resynthesized (with a reduced information load) by multiplying the dynamic envelope of each channel with a chosen source signal, band-pass filtering the resulting channels by the same filters as for the analysis part, and finally adding those channels together. The signal source can either consist of noise (noise vocoder) or of a sinewave (tone vocoder) (Loizou, 2006).

In CI models, variation exists in the settings that the vocoding technique allows to manipulate. Most importantly, the number of channels is typically between 12 and 22 and the source signal consists of a constant train of pulses delivered to the electrodes with a rate of several hundreds to several thousands of pulses per second per electrode. Moreover, the shape of the analysis and synthesis filters influences the amount of spectral smearing between filters. Steeper filter slopes cause less overlap than shallower filter slopes, improving discriminability of frequencies coded in different bands (Friesen, Shannon, Baskent & Wang, 2001).

Researchers use vocoder simulations of CIs to study CI hearing. This allows them to recruit participants with normal hearing, who are more numerous and form an audiologically more uniform

group than CI users. Moreover, it allows researchers to manipulate and study signal processing parameters that cannot be manipulated in CI users, since the settings in their devices are fixed. Results from studies using vocoders could, however, inspire the design of implants with improved settings. In this thesis, for the above reasons, vocoders were used to test the effect of filter slope on the discriminability of intonational and rhythmic variants of spoken sentences and musical fragments.

Limitations of vocoders as CI simulations should, however, be taken into account. The details of the signal processing procedure, the functioning of the ear, and the audiological background of the participants all differ between hearing and implanted individuals. Results from vocoder simulations cannot therefore be generalized to the population of CI users without caution. Ideally, tests with vocoders are followed up by tests with actual CI users in order to elucidate which vocoder settings most closely model the performance by the clinical population. These limitations of vocoder simulations will be dealt with in more detail in the respective chapters.

1.6 Overview of this thesis

This thesis investigates the processing of prosody by CI users from a number of perspectives, covering the mechanism and development of perception and production of the major types of prosody. These perspectives are covered by a number of broadly stated hypotheses of which more specific formulations are tested throughout different chapters. The motivations for these hypotheses will be stated in the chapters in which they are tested.

First of all, we investigate prosody by making a distinction between three major types, namely linguistic, emotional, and basic prosody, and studying one of them separately (basic prosody, in one study, Chapter 3) or comparing linguistic and emotional prosody (in three studies, Chapters 4, 5, and 6). There are fundamental differences

between linguistic and emotional prosody; e.g., knowledge of emotional prosody is possibly innate and universal, its cerebral processing right-lateralized and its realization of a gradient nature, whereas linguistic prosody is probably learned, less lateralized (Witteman et al., 2011) and its realization more discreet and rule-based. They might therefore be perceived and produced differently. A third type, basic prosody, is postulated as a rest category of prosodic measures that are performed without linking them to a linguistic or emotional function and is separately tested. We hypothesize that emotional prosody is differently recognized (**Hypothesis 1a**) and realized (**Hypothesis 1b**) than linguistic prosody. Second, emotional prosody perception and linguistic prosody perception are compared to music perception (elaborated below). It is predicted that emotional prosody is less correlated to music than linguistic prosody (**Hypothesis 1c**). Finally, we hypothesize that emotional prosody perception and production are less correlated than linguistic prosody perception and production (**Hypothesis 1d**).

The second perspective entails the distinction and relationship between speech perception and production. Perception (in three studies, Chapters 3, 4, and 5) and production (in one study, Chapter 1) are studied separately or in direct comparison (in one study, Chapter 6). We hypothesize that both perception (**Hypothesis 2a**) and production (**Hypothesis 2b**) are deviant in CI users, because they develop as an integrated system, which surfaces as a within-participant correlation between perception and production scores (**Hypothesis 2c**).

The third perspective is that of the relationship between prosody perception and music perception, two disciplines in which the acoustical dimensions of rhythm and melody are fundamental. In one study (Chapter 4), the hypothesis that NH listeners can be cue-specifically trained with musical materials to recognize musical melodies based on either melody or rhythm cues is tested (**Hypothesis 3**). Further, this training effect could transfer to reliance on the non-trained cue in melody perception (cross-cue transfer), on the trained

cues in prosody perception (cross-domain transfer) or to prosody perception for both cues (cross-cue plus cross-domain transfer) (as this does not involve a directional hypothesis, this issue is referred to as the **Transfer Issue**).

The fourth perspective is that of the mechanism of CI prosody hearing. CI users weight the cues they use to process prosody differently than NH listeners do. In this thesis, we compare prosody perception with the availability of temporal and F0 related cues by these two groups. Based on previous literature, **Hypothesis 4a** holds that of these two cues, CI users rely relatively heavily on temporal cues, as compared to their NH peers. **Hypothesis 4b** states that this cue weighting is reflected in speakers' speech output in that F0 related basic prosodic measures of CI users will deviate more than temporal prosodic measures. Within perception, it is hypothesized (**Hypothesis 4c**) that reduced channel interaction, as manipulated by steepening of channel filter slopes in vocoder simulations of CI hearing, will improve F0 perception, but not temporal perception.

The final perspective is that of the development of prosody in children. Two of the studies in this thesis were (retrospectively) performed with children with and without CIs (Chapters 2 and 6). We conjecture that language acquisition of CI children is delayed relative to that of NH peers by as much as the time until implantation (**Hypothesis 5a**), but that this delay is longer for prosody perception than for prosody production (**Hypothesis 5b**) and longer for linguistic prosody than for emotional prosody (**Hypothesis 5c**), and that CI children (partially) catch up with increasing experience with their device (**Hypothesis 5d**).

Chapter 2 reports a retrospective study of basic prosodic measures of prosody in spontaneous speech recordings of control children without and hearing-aged matched children with cochlear implants. The prosodic measures are categorized, from 'easy' to 'difficult' for CI users, as temporal, intensity related and F0 related and measured at 18, 24 and 36 months after implantation (for CI recipients) or birth (for

NH children). This study combines the perspectives of production, mechanism and development and tests **Hypotheses 2b**, **4b**, **5a**, and **5d**. It is predicted that production differs most for F0 related, less for intensity related and least for temporal measures and that any delay that exists with hearing-aged matched controls will be (partially) caught up after 36 months of CI experience, but more so for 'easier' measures.

Chapter 3 uses vocoder simulations of cochlear implant hearing to test the role of spectral smearing for intonation perception by normally hearing Dutch adults. Spectral smearing is the effect whereby the activation in a channel overlaps the area of a neighboring channels resulting in mixed (frequency) percepts. Sharper channel filters (i.e., with a steeper filter slope, expressed in dB/octave) reduce overlap and guarantee better F0 and intonation perception. Noise vocoder simulations are used instead of actual CI users, because they allow the manipulation of sound processing parameters (such as filter slopes) that could play a role in CI hearing but that the device of a given user does not allow to be manipulated (they could, however, be manipulated by redesigning a device). This study combines the perspectives of perception and mechanism and tests **Hypotheses 2a** and **4c**. Participants decide if naturally recorded but manipulated utterances that differ only in their F0 contour sound as a surprise, as news or as a predictable utterance. This setup, in which participants are asked to pay attention to the interpretation of the utterance, maximizes the likelihood that they listen to the stimuli as linguistic (intonational) and not just as acoustic (frequency varying) stimuli. It is hypothesized that intonation identification will be more accurate with a 40 dB/octave than with a 20 dB/octave condition, but that for both conditions it will be less accurate than in a control condition without vocoding.

Chapter 4 uses the same setup as the experiment described in Chapter 3 but extends its scope by using more different filter slopes (ranging

between 5 and 160 dB/octave), by making a distinction between emotional and linguistic prosody, and by making either temporal, F0 related or both cues available. This study combines the perspectives of the distinction between the two major types of prosody (emotional and linguistic), that of the perception and that of the mechanism and tests **Hypotheses 1a**, **2a**, **4a**, and **4c**. In this pair of experiments, NH Dutch adults decide (focus test) which of two words in a phrase carries sentential focus, or (emotion test) which of two emotions (happy or sad) is expressed in a phrase, whereby the phrases are highly similar to those in the focus test in order to justify a comparison between results of those two tests. These tests are repeated with and (as a control condition) without noise vocoding. It is hypothesized that intonation discrimination will improve with increasing filter slope and that this effect is smaller when temporal cues are available than when only F0 cues are available. The pattern of results might or might not differ between emotional and linguistic prosody. This experiment also functions as a validation for the stimuli, which are also used in several experiments in Chapter 6. Near-ceiling performance with the non-vocoded condition shows which of the stimuli appropriately convey focus position and emotions, thereby validating them for usage in further experiments.

Chapter 5 compares music perception to prosody perception. For the musical task, NH Dutch adults receive a short training to enhance their perception of either temporal (one group) or frequency (second group) perception of tone-vocoded stimuli and subsequently decide which of four possible well-known melodies was heard in conditions with only the rhythm of the melody available, only the tonal changes (but with all notes having the same duration) or both. They are also tested on emotional and linguistic prosody perception with the same cue conditions. The linguistic tasks are similar to those performed in the experiments in Chapter 4. This study combines the perspectives of the distinction between emotional and linguistic prosody, perception, the mechanism and music, and tests **Hypotheses 1a**, **1c**, **3**, **4a** and the

Transfer Issue. It is hypothesized that NH participants' perception in post-training tests is selectively enhanced for the trained cue. Further, this training effect could either transfer to non-trained cues in the same domain (i.e., within music; cross-cue transfer), in another domain but only for the same cue (i.e., to language; cross-domain transfer) or to another domain and another cue (cross-domain and cross-cue transfer).

Chapter 6 reports a set of experiments performed with young school-age children with and without CIs. They performed four core tests gauging their capabilities in the perception and production of both emotional and linguistic prosody. In the perception tests, temporal and F0 cues or both cues were made available. Additionally, participants performed three control tests aimed at probing their baseline level of non-verbal emotional development, of general linguistic development, and of basic picture identification and naming skills. Parents or caregivers completed a questionnaire about their children's language and medical background and the parents' socio-economic status. This set of experiments combines most of the perspectives of this thesis, viz. the distinction between linguistic and emotional perspectives, perception and production, the mechanism, and the development. It tests **Hypotheses 1a**,**b**,**d**; **2a**,**b**,**c**; **4a**,**b**; and **5a**,**b**,**c**,**d**.