Implant Technology and TFS Processing in Relation to Speech discrimination and Music Perception and Appreciation

PhD thesis

Defended by

Deborah Ballantyne

Curriculum: Pathologies of the Auditory System

School of PhD in Clinical-Experimental Neuroscience and Psychiatry

Sapienza University of Rome

Rome, 26 February 2018

Advisors

Prof. Elio De Seta, MD
Prof. Patrizia Mancini, MD PhD

DOCTORAL COMMITTEE

Chairman: **Prof. Dr. Marco Salvetti**

Advisor: **Prof. Elio De Seta**

Prof. Patrizia Mancini

Members:

Prof. Maurizio Inghilleri

Prof. Giorgio Di Lorenzo

Prof. Stefano Rufini

External Members: **Prof. Stefano Di Girolamo**

Prof. Marco Alessandrini

ACKNOWLEDGEMENTS

First, I would like to thank to all the study participants and their families without whose contributions this research could not have been carried out.

I wish to express my sincere gratitude to my advisor Prof. Dr. Patrizia Mancini for her continuous support during the course of my PhD study

Besides my advisor, I would like to thank Prof. Dr. Alfredo Berardelli and Prof. Dr. Marco Salvetti, Prof. Dr. Elio De Seta and Prof. Dr. Marco De Vincentiis for giving me the opportunity to work in the context of this PhD program.

I would especially like to thank my colleague Dr. Hilal Dincer D'Alessandro, with whom I had the privilege to have worked in the Cochlear Implant Center of Policlinico Umberto I of Sapienza University of Rome. She has been both an example and a complementary advisor on the more challenging aspects of my thesis. Furthermore, a mention is due to one of my students, Dr. Sabrina Ronconi who has untiringly cooperated in the collection of data and work with patients.

I am grateful to Dr. Patrick Boyle, for his great help with the preparation of the Italian STARR test material and his continuous support and advice for the Italian adaptation and set up of instrumentation

Lastly, I would like to dedicate this thesis to my late husband – Dr. Damiano Scoppettuolo – who constantly supported my desire to improve my knowledge throughout my career.

And to my family and my partner, the joys of my life

PRIOR PUBLICATION OF CONTENT

Material mentioned in this thesis relative to Pitch perception in Cochlear Implant recipients has been partially covered in a previous publication by Dincer D'Alessandro H, Filipo R, Ballantyne D, Attanasio G, Bosco E, Nicastri M, Mancini P (2015): Low-frequency pitch perception in children with cochlear implants in comparison to normal hearing peers. *Eur Arch of Otorhinolaryngol* 272:3115-3122.

Material covered in Chapter 4 has previously been published by: Dincer D'Alessandro H, Ballantyne D, De Seta E, Musacchio A, Mancini P (2016): Adaptation of the STARR test for adult Italian population: A speech test for a realistic estimate in real-life listening conditions. *Int J Audiol* 55:262-267.

Material covered in Chapter 5 has been accepted for publication in Ear and Hearing: Dincer D'Alessandro H, Ballantyne D, *et. al.* (2017): Temporal Fine Structure Processing, Pitch and Speech Perception in Cochlear Implant Recipients. Ear and Hearing (Submitted February 2017, accepted for revision May 2017), accepted for Publication October 2017.

ABSTRACT

Direct stimulation of the auditory nerve via a Cochlear Implant (CI) enables profoundly deaf subjects to perceive sounds. Many CI users find language comprehension satisfactory in quiet and accessible in the presence of noise. However, music contains different dimensions which need to be approached in different ways. Whilst both language and music take advantage of the modulation of acoustic parameters to convey information, music is an acoustically more complex stimulus than language, demanding more complex resolution mechanisms.

One of the most important aspects that contributes to speech perception skills, especially when listening in a fluctuating background, is Temporal Fine Structure processing. TFS cues are predominant in conveying Low Frequency (LF) signals. Harmonic (HI) and Disharmonic (DI) Intonation are tests of pitch perception in the LF domain which are thought to depend on availability of TFS cues and which are included in the protocol on this group of adult CI recipients.

One of the primary aims of this thesis was the production of a new assessment tool, the Italian STARR test which was based on the measurement of speech perception using a roving-level adaptive method where the presentation level of both speech and noise signals varied between each sentence presentation. The STARR test attempts to reflect a better representation of real world listening conditions where background noise is usually present and speech intensity varies according to vocal capacity as well as the distance of the speaker. The outcomes for the Italian STARR in NH adults were studied to produce normative data, as well as to evaluate inter-list variability and learning effects. (Chapter 4).

The second aim was to investigate LF pitch perception outcomes linked to availability of TFS cues in a group of adult CI recipients including bimodal users in relation to speech perception, in particular Italian STARR outcomes. Here it was seen that age had a significant effect on performance especially in older adults. Similarly, CI recipients (even better performers) showed abnormal findings in comparison to NH subjects. On the other hand, the significant effect of CI thresholds re-emphasized the sensitivity of the test to low intensity speech which a CI user can often encounter under everyday listening conditions. Statistically significant correlations between HI/DI and STARR performance were found. Moreover, bimodal benefit was seen both for HI/DI and STARR tests. Overall findings confirmed the usefulness of evaluating both LF pitch and speech perception in noise in order to track changes in TFS sensitivity for CI recipients over time and across different listening conditions which might be provided by future technological progress. (Chapter 5)

Finally, the last and main aspect taken into account in this thesis was the study of the difficulties experienced by CI users when listening to music. An attempt was made to correlate findings resulting from the previous phases of this study both to Speech in Noise and to the complex subjective aspects of Music Perception and Appreciation: correlation analysis between HI/DI tests and the main dimensions of Speech in Noise (STARR and OLSA) and Music Appreciation was performed. (Chapter 6). Interestingly, positive findings were found for the two most complex types of Music (Classical, Jazz), whereas Soul did not seem to require particular competence in Pitch perception for the appreciation of the subjective variables taken into consideration by this study.

Key Words: Speech perception in Noise; Cochlear implants; Pitch perception, Temporal Fine Structure; Music Perception.

RIASSUNTO

La stimolazione diretta del nervo acustico tramite Impianto Cocleare (IC) consente ai sordi profondi di percepire di nuovo i suoni. Diversi portatori di IC trovano soddisfacente la comprensione del linguaggio in silenzio e accessibile in presenza di rumore. Tuttavia, la musica contiene varie dimensioni che debbono essere affrontate con modalità diverse. Mentre sia il linguaggio che la musica si servono della modulazione dei parametri acustici per trasmettere delle informazioni, la musica è uno stimolo più complesso rispetto al linguaggio, e richiede meccanismi di risoluzione a sua volta più complessi.

Uno degli aspetti più importanti che contribuisce alle abilità di Percezione del Linguaggio, in modo particolare quando l'ascolto avviene in un ambiente acustico fluttuante, è noto come Elaborazione Temporale della Struttura Fine (TFS). I cues tratti da questo processo sono prevalenti per la trasmissione dei segnali nelle basse frequenze. Harmonic (HI) e Disharmonic (DI) Intonation sono tests di percezione del Pitch per il dominio delle basse frequenze che si pensa possano dipendere dalla disponibilità dei cues della TFS. Queste prove fanno parte del protocollo eseguito su questo gruppo di portatori di IC.

Uno degli obiettivi principali della presente tesi era la produzione e applicazione di un nuovo strumento di valutazione, ossia la versione italiana dello STARR test che misura la percezione del linguaggio con un metodo adattivo dove il livello di presentazione sia delle frasi che del rumore venivano modificati nella presentazione di ogni frasi. Lo scopo dello STARR test è quello di rispecchiare meglio le condizioni di ascolto nella vita reale, dove è quasi sempre presente il rumore di fondo e l'intensità del linguaggio parlato varia secondo le capacità vocali e la distanza dell'interlocutore. Per la versione italiana dello STARR test i risultati hanno prodotto dei dati normativi insieme alla conoscenza di una eventuale variabilità interlista e di effetti di apprendimento. (Capitolo 4).

Il secondo obiettivo era quello di studiare gli esiti delle nostre prove di percezione per le basse frequenze (BF) legate alla disponibilità di TFS cues in un gruppo di portatori di IC, compreso i bimodali, in rapporto alla percezione del linguaggio in presenza di rumore, in modo particolare per lo STARR test. I risultati hanno messo in evidenza l'effetto dell'età sulle performance in modo particolare negli adulti più anziani. Inoltre, I portatori di IC (anche gli star patient) hanno dimostrato degli esiti anormali rispetto ai soggetti NH. D'altra parte, l'effetto significativo delle soglie audio raggiunte dall'impianto, ha potuto sottolineare la sensibilità del test al linguaggio a bassa intensità spesso presente in condizioni quotidiani di ascolto. Erano presenti correlazioni significative tra HI/DI e performance per lo STARR. Inoltre, per entrambi I protocolli erano presenti benefici in modalità bimodale. I risultati sembrano confermare l’utilità di applicare sia le prove di Pitch (HI/DI) sia quelle per la percezione del linguaggio in presenza del rumore allo scopo di poter monitorizzare eventuali modifiche nella sensibilità dell’elaborazione temporale della struttura fine (TFS processing) in soggetti portatori di impianto cocleare sia nel tempo sia in condizioni diverse di ascolto, le quali potrebbero essere previste da ulteriori progressi tecnologici nel campo degli Impianti cocleari. (Capitolo 5)

Infine, l’ultimo e principale aspetto preso in considerazione da questa tesi è stato lo studio delle difficoltà spesso riferite dai portatori di impianto cocleare per le varie dimensioni dell’ascolto della musica: capacità di discriminazione, accesso al significato della musica e apprezzamento soggettivo della musica. Si è tentato quindi di trovare un rapporto tra i risultati delle fasi precedenti di questa tesi anche con gli aspetti complessi ma più soggettivi della percezione e dell’apprezzamento della Musica. E’ stata effettuata l’analisi statistica tra HI/DI e le principali dimensioni della percezione del Linguaggio in Rumore (STARR e OLSA) e Apprezzamento della Musica (Capitolo 6). Esiti significativi sono stati riscontrati per i due tipi più complessi della Musica (Classica, Jazz), mentre il Soul non sembrava richiedere particolari competenze di percezione del Pitch ai fini dell’apprezzamento delle variabili soggettivi di ascolto presi in considerazione da questo studio.

Parole Chiave: Percezione del Linguaggio in Rumore; Impianti Cocleari; Percezione del Pitch; Elaborazione Temporale della Struttura Fine; Percezione della Musica.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AB	Advanced Bionics
ACE	Advanced Combination Encoder
A§E	Auditory Speech Sounds Evaluation
BM	Basilar Membrane
CI	Cochlear Implant
CIS	Continuous Interleaved Sampling
DI	Disharmonic Intonation
EAS	Electro-Acoustic Stimulation
F0	Fundamental Frequency
FSP	Fine Structure Processing
FS4	Fine Structure on 4 channels
FS4-p	Parallel stimulation FS4
HA	Hearing Aid
HF	High Frequency
HI	Harmonic Intonation
HINT	Hearing in Noise Test
HiRes	HiResolution
HiRes-S	Sequential stimulation HiResolution
HiRes-P	Parallel stimulation HiResolution
HiRes 120	HiResolution with the Fidelity 120
HS	Head Shadow
IHC	Inner Hair Cells
IHR	Institute of Hearing Research
ITD	Interaural Time Delay
JND	Just Noticeable Difference
LF	Low Frequency
NH	Normal Hearing
OHC	Outer Hair Cells
OLSA	Oldenburg Sentence Test

PTA	Pure Tone Average
SAS	Simultaneous Analogue Stimulation
SD	Standard Deviation
SM	Scala Media
SNR	Signal-to-Noise Ratio
SP	Speech Processor
SPSS	Statistical Package for Social Sciences
SRT	Speech Reception Threshold
SQ	Binaural Squelch
ST	Scala Tympani
STARR	Sentence Test with Adaptive Randomized Roving Level
SU	Binaural SUmmation
SV	Scala Vestibuli
TFS	Temporal Fine Structure
WRS	Word Recognition Score

LIST OF FIGURES

LIST OF TABLES

Chapter 1 INTRODUCTION

Hearing loss, also known as hearing impairment, is considered as one of the most common human disorders and may arise from dysfunction of any part of the auditory pathway. Nowadays, several medical, surgical and technological tools are available for the treatment of hearing loss depending mainly on type and degree. In the case of cochlear dysfunction, options may include traditional amplifying systems with Hearing Aids (HA) or Cochlear Implants (CI) which bypass the severely impaired cochlea [1] and directly stimulate the acoustic nerve.

Over the last years, cochlear implantation has become a common choice for the rehabilitation of bi-lateral, severe to profound, cochlear hearing loss and CI systems have proved to offer useful auditory information for the perception of environmental sounds, speech and music. The causes of deaf-ness that have been associated with CI recipient individuals range from unknown, genetic or inherited pathology to unpredictable, accidental deafness due to trauma or infection [2].

1.1 Physiology of the Ear and Cochlear Pathology

Deafness is often caused by the absence or degeneration of sensory hair cells in the cochlea and in the case of a pathological cochlea, the logical approach would be to bypass the damaged part and to stimulate spiral ganglion cells electrically with a cochlear implant. The fundamental criterion for cochlear implantation is a majority of missing or non-functioning cochlear hair cells where information from the acoustic environment transmitted through a normal or near-normal middle ear can-not be transduced into effective electrical signals that travel along the body’s natural auditory system to the brain so that comprehension can take place. If there is a retrocochlear pathology, a cochlear implant will not be useful to restore deafness [2].

The part of the inner ear that is concerned with hearing is the cochlea which contains the organ of Corti that sits on the BM. The vestibular (balance) system consists of three semicircular canals and two further structures known as the utricle and saccule. The organ of Corti has hair cells that are the sensory receptors for hearing and these cells are in contact with the nerve cells of the VIII cranial nerve, which connects the peripheral ear to the central nervous system. The auditory branch of the eight nerve is generally called the auditory or cochlear nerve whilst the vestibular branches are often referred to as the vestibular nerve. The conductive system, which consists of the outer and middle ear structures, transduces sound and transmits the stimulus to the inner ear. Hence, the sensorineural system - cochlea and eight cranial nerve - induces the physiological response to the stimulus, activation of the nerve cells and the encoding of the sensory response into a neural signal. The aspects of the central nervous system that deal with this neurally encoded signal are generally called the central auditory nervous system [3].

Under conditions of normal hearing, sound waves which travel through the air reach the tympanic membrane via the ear canal, causing vibrations that move the three ossicles [5]. This produces a coordinated movement of the ossicular chain resulting in a piston-like movement of the stapes. It is the “footplate” of the stapes, which is attached to the oval window whose inward and outward movements induce pressure oscillations in the cochlear fluids, which in turn give rise to a traveling wave, displacing fluids along the BM. This membrane has graded mechanical properties: At the base of the cochlea (near the stapes and oval window) it is narrow and stiff whilst at the other end (near the apex) it becomes progressively wider and less stiff. The resulting traveling wave, which propagates from the base to the apex of the cochlea, is characterized by points of maximal response based on the frequency or frequencies of the pressure oscillations within the cochlear fluids. For an oscillation with a single frequency, the magnitude of displacement increases up to a particular point along the membrane and then drops sharply thereafter. Low frequencies produce maxima near the apex whereas high frequencies produce maxima near the base of the cochlea. Motion of the BM is sensed by the sensory hair cells in the cochlea, which are attached to the top of the BM within the organ of Corti. The cells are arranged in four rows along the total length of the cochlea.

The cells in the innermost row are called the Inner Hair Cells (IHC), and the cells in the remaining rows are called the Outer Hair Cells (OHC). The IHCs are closest to the modiolus or “center core” of the cochlea. Each hair cell has fine rods of protein, called stereocilia, emerging from one end. When the BM moves at the location of a hair cell, the rods are deflected as if they are hinged at their bases. Such deflections increase the release of a chemical transmitter substance at the base of the IHCs, whereas deflections in the opposite direction inhibits its release. [4]

In contrast, deflections of the stereocilia of the OHCs produce electromotile changes in the length of the cells, which in turn increase the sensitivity and sharpen the “tuning” of the BM to frequencies that correspond closely to the position(s) of the stimulated cells. Thus, the OHCs act as a highly selective biological amplifier. The increases in chemical transmitter substance at the bases of the IHCs increase discharge activity in the immediately adjacent auditory neurons, whereas reduction in the substance inhibits activity. Changes in neural activity thus reflect events at the BM. These changes are transmitted to the brain via the auditory nerve, which is effectively a collection of all neurons that innervate the cochlea [5].

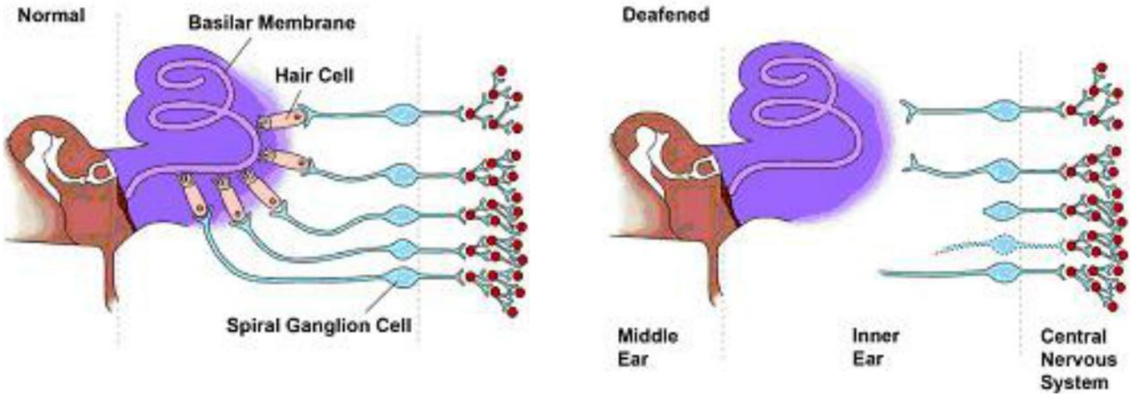


Figure 1. Anatomical structures in normal and deafened ears.
Source: Wilson and Dorman (5)

There are two basic types of hearing loss: conductive and sensorineural hearing loss. Conductive hearing loss results from an abnormality before the cochlea. Any obstruction or malformation which impedes the transfer of sound from the environment through the outer and middle ear, thus attenu-ating it, will result in a conductive hearing loss. This type of loss is within the mild to moderate range, characteristically ranging from 20 dB HL to a maximum of 60 dB HL. The primary effect of a conductive hearing loss is a loss of intensity. Sensorineural hearing loss occurs as a result of dam-age to the inner ear (cochlear hearing loss) or to the auditory nerve pathways (retrocochlear hearing loss) [6]. The principal cause of hearing loss is partial or complete destruction of the sensory hair cells that are extremely fragile structures. The hair cells are subject to a wide variety of damage which include but are not limited to genetic defects, infectious diseases, overexposure to loud sounds, drugs and aging. Damage to the OHCs elevates hearing thresholds and degrades frequency resolution whilst damage to the IHCs produces difficulty in speech perception and is characterized by more profound losses including total deafness. The IHCs are largely or completely absent in the deaf or deafened cochlea, thus inhibiting the connection between the peripheral and central auditory systems.

The CI function is to bypass the pathological hair cells by directly stimulating the neurons in the auditory nerve. Figure 3 is the anatomical illustration of the deafened cochlea (a complete absence of hair cells- an anatomical situation which can be usually faced by CI specialists) in comparison to a normal auditory system including the tympanic membrane, the three ossicles, the oval window, the BM, the IHCs, and the adjacent neurons of the auditory nerve. A small number of cells may remain for some patients, usually in the apical (Low Frequency -LF) part of the cochlea. If not stimulated, the peripheral parts of the neurons undergo “degeneration” and cease to function. Fortunately, even after prolonged deafness or aetiologies such as meningitis, some usually survive [5,7].

1.2 Psychoacoustics of Hearing

1.2.1 Perception of Pitch

Pitch may be defined as “That attribute of auditory sensation in terms of which sounds may be ordered on a musical scale”. (American Standards Association, 1960). In other words, variations in Pitch give rise to a sense of melody. Pitch is related to the repetition rate of the waveform of a sound; for a pure tone this corresponds to the frequency and for a periodic complex tone to the fundamental frequency. However, since Pitch is a subjective attribute it cannot be measured directly. Assigning a pitch value to a sound is generally understood to mean specifying the frequency of a pure tone having the same subjective pitch as the sound. There are two ways in which information about the frequency of a sound stimulus can be transmitted up the auditory nerve: these are Place Coding and Temporal Coding.

Place Coding: As with the basilar membrane and the inner hair cells, so the fibres of the auditory nerve are very sharply tuned. Each fibre has a ‘characteristic frequency’ for which it is most sensitive to stimulation, the fibres innervation the base of the cochlea having the highest characteristic frequencies, those innervating the apex having the lowest. Coding based on this form of frequency selectivity is known as Place Coding.

***Temporal Coding*:** For frequencies up to about 5 kHz, the frequency of the stimulating tone is reflected in the rate of firing in the nerve fibres; the neural ‘spikes’ are said to be phase-locked to the stimulus and the Pitch is determined by the time pattern of these spikes. This is known as Temporal Coding.

1.2.2 Perception of Pitch of Pure tones

Hence, sensation of Pitch elicited by a pure tone is closely related to the frequency of the stimulating tone; the higher the frequency, the higher the pitch.

One of the most remarkable attributes of the ear is the ability to detect very small changes in frequency. For example, if two pure tones, of very short duration (about 500ms), centred at a frequency of 1 kHz are perceived in succession, the trained ear can detect as small a difference as 3kHz. This is known as frequency discrimination.

Frequency discrimination deteriorates abruptly above 4-5 kHz, the upper limits at which temporal coding is possible. Hence, it is probable that pitch perception and frequency discrimination are determined primarily by temporal coding for frequencies up to this level, but by place coding for higher frequencies.

1.2.3 Perception of Pitch of Complex Music Tones

Although pure tones are used extensively in experimental psychology and in all psychoacoustic tests of hearing, the human being practically never listens to pure tones, either in nature or in art, for all the sounds of everyday life are complex tones.

It is the function of the auditory system to break down these complex tones into series of pure tones, or sinusoids. This consists of spectral representation and is known as Fourier analysis. [8 Moore]

In this context, musical instruments emit complex tones which consist, essentially, of a fundamental component and a series of 'overtones' and harmonics. If the frequency of each component is plotted in a graph against its amplitude, an acoustic spectrum of the complex tone is obtained (Figure p.48 it).

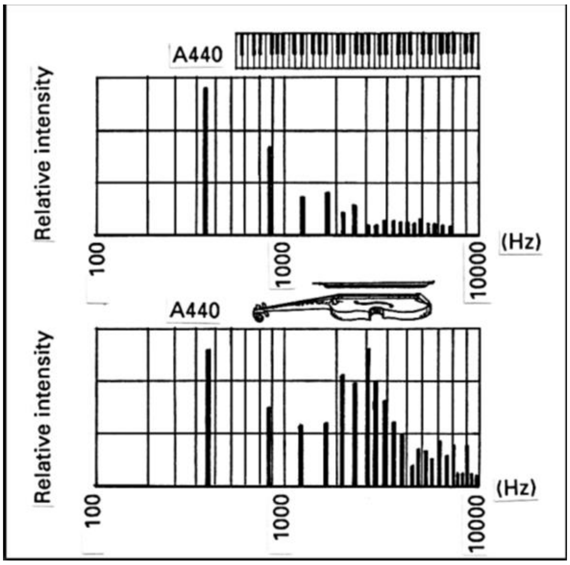


Figure 2 Acoustic Spectra of Piano and Violin
Source: Ballantyne (9)

The relative acoustic spectra of a piano and of a violin, each playing the note ‘A’ (Fa) are shown. In the spectrum of the piano, it will be seen that the first harmonic (the fundamental at 440Hz) is strong, almost twice the intensity of the second harmonic, which is an octave higher (i.e. twice the frequency). The next four harmonics (1320, 1760, 2200 and 2640 Hz – all simple multiples of the fundamental) are considerably weaker, but more or less the same as one another and the remaining twelve are insignificant. In the acoustic spectrum of the violin the 7th harmonic is as strong as the fundamental, the 5th, 6th, and 8th are not much weaker, and at least five other harmonics contribute significantly to the sound structure. Both instruments produced the same sensation of pitch in the listening ear, in this instance to the fundamental note ‘A’(Fa), with a frequency of 440 Hz. [9 Ballantyne D., Handbook of Audiological Techniques, Butterworth-Heinemann 1990 pp 46-50]

It is now assumed that the perception of pitch of complex musical tones is effected in two stages: the first involves the analysis of the lower harmonics and depends on both Place and Temporal in-formation from the cochlea and nerve fibres; in the second stage, some form of ‘pattern recognizer’ is thought to ‘calculate’ the fundamental frequency of a complex tone by matching its lower harmonics with those analysed in the first stage. The pitch perceived will then correspond to the frequency of the ‘internally’ computed fundamental.

However that may be, it is the difference in timbre (sound quality) between different instruments which enables the ear to distinguish one for all others, and the perception of Timbre depends largely on the frequency selectivity of the auditory system.

1.2.4 Perception of Speech

Timbre is also the quality of sound that distinguishes one human voice from another. Like any other wind instrument, the vocal ‘instrument’ emits a fundamental tone (generated in the larynx) and a number of harmonics (generated in the tract above the larynx). The laryngeal tones are generated by vibrations of the vocal cords, and the frequency of these vibrations determines the frequency of the laryngeal tone. In adult males, this averages 120 Hz and one octave higher (ie about 240 Hz) in females. However, fluctuations occur in these laryngeal frequencies during phonation; in other words the voice has intonation. Intonation patterns play an important part role in the expressive elements of speech. For example, they enable us to distinguish a question from a statement. Other characteristics are resonances or ‘formants’. Generally speaking, vowel sounds are lower in frequency but higher in intensity than consonant sounds; and in most Indo-European speech whilst the vowels pro-vide power or energy, intelligibility is dependent largely on the consonants.

In final analysis, perception is a process of identifying and interpreting the information which reaches our consciousness from the world about us by the way of our senses. As Aristotle said ‘Nothing is in the mind that did not pass through the senses’; but it is beyond the capacity of the human brain to deal with all of the many highly complex signal which enter the ears.

Fortunately, speech is highly redundant; in other words, it is possible to dispense with many of the acoustic cues of speech without detriment to our understanding. So that what we actually hear when we listen to speech, is not just a conglomeration of cochlear components, but a series of sound symbols; and we fuse all the many complex sounds of speech into sound ‘patterns’. So how does the brain deal with these patterns?

To date, the only certainty is that the recognition of visual patterns, specialized ‘neurons’ exist in the brain which respond to complex patterns, like the shape of a head; and it has been suggested that similar ‘complex auditory neurons’ may exist, which respond exclusively to complex patterns of sound – possible, for example, to words.

Ultimately, however, speech recognition is accomplished not merely by acoustic cues, but largely by linguistic cues. Speaker and listener must share a common language.

Chapter 2 COCHLEAR IMPLANT TECHNOLOGY

2.1 Development of Cochlear Implants

Subsequent to preliminary experimental studies carried out by Djourno and Eyries in 1957 [10) relating to the direct electric stimulation of the auditory nerve, it was suggested that the activation of the auditory periphery in humans through an electrical device could efficiently provide useful in-formation to the central auditory pathway [11], even though there were some doubts about safety and reliability. As a result, initiatives at stimulating the auditory nerve for clinical benefit began in the United States. In 1972, this led to the production of the first commercially marketed system [12,13,14], known as the 3M/House device which consisted of a single electrode and an external Speech Processor This implant enabled individuals to detect changes in duration and offer sensations of tonality, thus aiding speech-reading.

Multiple-channel devices were introduced in 1984, and the development of single- and multi-channel systems continued to move hand in hand until the 1990s. Single channel implants deliver auditory information through a single electrode whereas multi-channel implants convey different parts of the signal via several distinct channels that stimulate different regions of the cochlea. The terms single and multi-channel are used to describe the number of active electrodes through which different in-formation is sent; whereas single or multi-electrode refers to the number of electrodes in the implant [1]. The single-channel implants experienced successful use in terms of providing basic access to acoustic information where some patients were able to make successful use of the simple stimuli [15].

The development of multichannel systems required advanced technologies of digital signal processing (DSP) chip design, miniaturization, battery consumption and other engineered capabilities. Over time, they outweighed single-channel devices based on enhanced spectral perception and enhanced speech recognition capabilities, as shown in large adult clinical trials [2,11,16,17]. Most patients who have had single channel implants replaced with multi-channel implants have shown varying degrees of improvement in speech and environmental sound recognition [18,19]. CI centers nowadays choose to use multi-channel, intra-cochlear implants as they give better performance [1].

2.2 Basics of Cochlear Implants

Cochlear implantation is mainly based on the following principles: foreign, biocompatible materials can be placed within the human body without being rejected [2,20] and auditory nerve fibres respond to electrical stimulation [2,21].

Figure 3 The essential components of a conventional cochlear implant.

(A) External parts: (1) Behind the Ear Speech Processor (2) transmitter coil.
(B & C) Internal parts: (3) receiver-stimulator (4) electrode array.
Source: Vaerenberg [15].

To date, four major CI manufacturers exist in the market: Cochlear Ltd. (Australia), Advanced Bionics (Switzerland), MED-EL (Austria), Oticon Medical/Neurelec (France). Although the market offers a wide variety in technical and cosmetic features, CI systems basically consist of two parts: an internal part that is surgically implanted and an external part that is called speech processor. Figure 5 shows the essential components of a cochlear implant system.

CI systems share a common working principle. The microphone which is located on the SP picks up environmental sounds and converts these analogue sounds into digital information. This information is sent via radiofrequency transmission from an external coil to an internal receiver-stimulator implanted under the skin. The receiver-stimulator transforms the signals into patterns of electrical stimulation and delivers to the electrode array. The auditory nerve fibres in the cochlea pick up the signals and convey them through the natural auditory pathway to the brain, giving the sensation of the sound.

CI systems use a transcutaneous link. The link is bidirectional to allow transmission of data from the implanted components out to the external components (SP and coil) as well as transmission of data from the SP to the implanted receiver-stimulator. The data sent from the implanted parts to the external parts may include:

- Information about the status of the receiver/stimulator;
- impedance of each single electrode;
- voltages at unstimulated electrodes;
- neural evoked potentials.

The electrodes and electrode carrier (together called electrode array) for CI systems are placed in the ST, which offers an accessible site that is close to the spiral ganglion. Figure 6 shows a cross-section of the implanted cochlea. It shows the three chambers and the partial insertion of an electrode array into the ST [5,23]. The electrodes should be biocompatible, mechanically stable and should facilitate atraumatic insertion. In general, flexible arrays facilitate insertion. The array is inserted through a drilled opening made by the surgeon in the bony shell of the cochlea overlying the ST and close to the base of the cochlea (called a “cochleostomy”).

Alternatively, the array may be inserted through the round window membrane, which also is close to the basal end of the cochlea and ST. The cochleostomy provides a “straighter shot” into the ST than the round window approach [5]. The number of intracochlear electrode contacts available ranges from 12 to 22. In addition to the intracochlear electrodes, 1 or 2 electrodes are positioned outside of the cochlea. These electrodes serve as reference or ground electrodes [24].

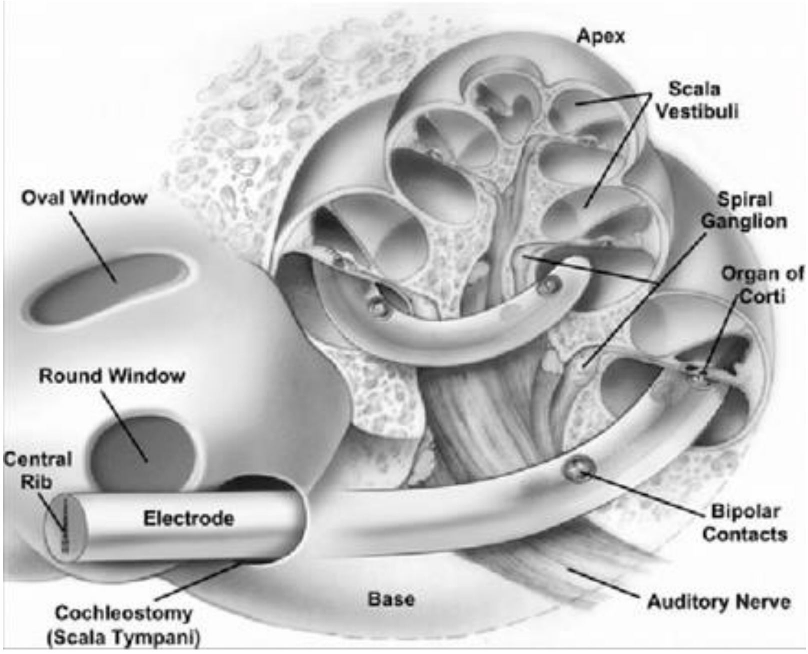


Figure 4. Cross-section of implanted cochlea.
Source: Wilson and Dorman [5]

The depth of insertion can be limited by the decreasing lumen of the ST from base to apex, the curvature of the cochlear spiral, and an uneven and unsmooth lumen, particularly in the apical region. Insertions are typically 18 to 26 mm and no array has been inserted farther than about 30 mm (The length of the human cochlea is typically about 35 mm). In some cases, only shallow insertions are possible, such as when bony obstructions in the lumen impede further insertion. Different electrodes in the implanted array may stimulate different populations of neurons. Neurons at different positions along the length of the cochlea respond to different frequencies of acoustic stimulation in normal hearing subjects. CI systems attempt to mimic or reproduce this "tonotopic" encoding by stimulating basally situated electrodes (first turn of the cochlea and lower part of Figure 4) to indicate the presence of High Frequency (HF) sounds and by stimulating electrodes at more apical positions (deeper into the ST and ascending along the first and second turns in Figure 4) to indicate the presence of sounds with lower frequencies [5,23].

Intracochlear electrodes can be stimulated in a monopolar or bipolar configuration. In the monopolar configuration, each intracochlear electrode is stimulated with reference to a remote electrode, usually in the temporalis muscle or outside of the case of the implanted receiver-stimulator. In the bipolar configuration, one intracochlear electrode is stimulated with reference to another (adjacent) intracochlear electrode. Different pairs of electrodes are used to stimulate different sites along the electrode array [24]. In Figure 4, closely spaced pairs of bipolar electrodes are shown, but all present-day implant systems use the monopolar coupling configuration since it supports performance that is at least as good as bipolar coupling and requires less current and battery power to produce auditory percepts. The spatial specificity of stimulation with an ST electrode most likely depends on multiple factors, including the orientation and geometric arrangement of the electrodes, the proximity of the electrodes to the target neural structures, and the condition of the implanted cochlea in terms of nerve survival and ossification. An important goal of electrode design is to maximize the number of largely non-overlapping populations of neurons that can be addressed within the electrode array.

However, evidence suggests that no more than 4 to 8 independent sites are available with current designs, even for arrays with as many as 22 electrodes [25,26,27]. Most likely, the number of independent sites is limited by substantial overlaps in the electric fields from adjacent (and more distant) electrodes. The overlaps are unavoidable for electrode placements in the ST because the electrodes are "sitting" in the highly conductive fluid of the perilymph and, additionally, are relatively far away from the target neural tissue in the spiral ganglion. A closer apposition of the electrodes to the inner wall of the ST would move them a bit closer to the target cells, and such placements have been shown in some cases to produce an improvement in the spatial specificity of stimulation [28]. However, a large gain in the number of independent sites may well require a fundamentally new type of electrode or a fundamentally different placement of electrodes. Figure 6 shows a complete presence of hair cells (in the labelled organ of Corti) and a pristine survival of cochlear neurons. However, the number of hair cells is zero or close to in cases of total deafness. In addition, survival of neural processes peripheral to the ganglion cells (the “dendrites”) is rare in the deafened cochlea. Survival of the ganglion cells and central processes (the axons) ranges from scarce to substantial. The pattern of survival is usually not uniform, with reduced or sharply reduced cell counts in certain regions of the cochlea and the neural substrate or target for a cochlear implant can be quite different between patients [5].

Despite above mentioned limitations, CI recipients over the last years have benefited from advances in battery, integrated circuit and DSP chip technologies, in that the developments have allowed smaller and more capable SPs and implanted receiver-stimulators. The SPs are available in different models and are usually named according to their wearing styles: body worn processors (worn on the belt or in a pocket usually to offer a more robust solution for young children) and behind the ear processors (usually preferred by adults). The trend in cochlear implantation has been towards achieving better patient performance as well as attempting to improve cosmetic features.

In the 1990s, clinical and basic science investigations produced changes in implant technology and in clinical approaches to cochlear implantation. Electrode and SP designs have evolved to produce encoding strategies that are associated with higher performance levels. Simultaneously, along with device development and observations of safety and reliability there has been emphasis on earlier implantation in children. There is now recognition of the required services for children to optimize implant performance and the structure of the interaction needed among the implanted child, family members, school staff, and implant team professionals. There is now substantially greater potential for open-set speech understanding in children and adults. Technologic advances of the past decade have refined speech encoding strategies and have expanded implant candidacy [11,29,30].

2.3 Cochlear Implant Processing Strategies

A normally hearing ear can discriminate speech by detecting changes in the frequency and intensity (or pitch and loudness referred as their subjective percepts) with time. Thus, the CI aims to represent the sound input so that the recipient can detect pitch and loudness variations and have the ability to understand speech [1]. However, the amount of information that can be presented and perceived with a cochlear implant is much less than that for a normal hearing (NH) person who is listening to an unprocessed acoustic signal, e.g. CI's capacity is restricted by the limited number of electrodes (max. 22 intracochlear electrodes currently) and by the stimulus rate. Additionally, the dynamic range of stimulus amplitudes from auditory threshold to loud precepts is in the order of 10 to 20 dB for electrical pulses in comparison to the order of 100 dB for acoustical stimuli [31].

The SP is usually activated in CI recipients at 3 to 4 weeks after surgery. Initially CIs should be programmed to activate the system and to make the recipient hear sounds. Subsequently, fitting sessions should be done regularly (but less over time), the goal being to achieve the most appropriate configuration for each recipient.

CI systems offer a choice of different speech coding strategies and variables that can be adjusted during fitting sessions. The parameters that can usually be changed are as follows: thresholds, comfortable levels, active channels or electrodes, stimulation rate, stimulation mode, frequency boundaries, stimulation cycle, sampling rate, pulse width, dynamic ranges, smoothing cut-off filter, automatic channel selection, noise suppression, input dynamic range, compression, rectification mode, pulse rate per channel and more [2].

Initially sound information was transmitted via the use of an analogue waveform with continuous and simultaneous stimulation of the electrodes. Analogue waveform was first used in the Compressed Analogue (CA) strategy of the Ineraid cochlear implant system and later in the Simultaneous Analogue Stimulation (SAS) strategy of the Clarion (Advanced Bionics- AB). The SAS strategy consisted of 16 electrodes used for bipolar stimulation (eight pairs of one active and one reference electrode). The aim of bipolar stimulation was to reduce current spread and to minimize channel interaction.

Current CI models make use of pulsatile waveforms that consist of series of pulses extracted from the incoming signal and delivered to different channels based on their frequency. Each pulse is presented to each channel sequentially in order to minimize channel interactions and to maximize spectral information. Pulsatile strategies do not convey the whole waveform but rapidly updated samples of the sound signal. To represent adequately changes in the signal with time, rapid updating of the incoming signal is required. Most pulsatile strategies use sequential stimulation and fall into two broad categories: Continuous Interleaved Sampling (CIS) strategy and the n of m or spectral maxima type of strategy [1].

2.3.1 CIS Strategy

One of the most effective approaches for representing speech with current CI technology is the CIS strategy [5]. This strategy was the first, fast rate strategy developed by Wilson et al. [32]. CIS gets its name from the continuous sampling of the (compressed) envelope signals by rapidly presented pulses that are interleaved across electrodes. It filters sound information into bands of frequencies with a bank of bandpass filters. Envelope variations in the different bands are represented at corresponding electrodes in the cochlea with modulated trains of biphasic electrical pulses. The envelope signals extracted from the bandpass filters are compressed with a nonlinear mapping function prior to the modulation in order to map the wide dynamic range of sound in the environment (up to about 100 dB) into the narrow dynamic range of electrically evoked hearing (about 10 dB). The output of each bandpass channel is directed to a single electrode, with low-to-high frequency channels assigned respectively to apical-to-basal electrodes, to mimic the frequency mapping in the normal cochlea. The pulse trains for the different channels and corresponding electrodes are interleaved in time so that the pulses across channels and electrodes are non-simultaneous. This eliminates a principal component of electrode interaction, which otherwise would be produced by direct vector summation of the electric fields from different (simultaneously stimulated) electrodes. The corner or “cutoff” frequency of the low-pass filter in each envelope detector is usually set at 200 Hz or higher so that the Fundamental Frequencies (F0s) of speech sounds are represented in the modulation waveforms. All the currently available cochlear implants can be programmed with the CIS strategy but the implementation of the strategy may vary in different implants for parameters such as filtering, envelope extraction, the number of channels, the pulse rate and the update rate. CIS implementations use up to 22 channels and corresponding stimulus sites [5].

2.3.2 The n-of-m, SPEAK and Advanced Combination Encoder (ACE) Strategies

The spectral maxima strategies use a channel-selection scheme in which the envelope signals for the different channels are scanned prior to each frame of stimulation across the intracochlear electrodes in order to identify the signals with the n-highest amplitudes from among m processing channels. Stimulus pulses are delivered only to the electrodes that correspond to the channels with the highest amplitudes. Examples of spectral maxima type strategies are the Nucleus SPEAK and ACE strategies and the Medel n of m strategy. The parameter n is fixed in the n-of-m and ACE strategies and it can vary from frame to frame in the SPEAK strategy, depending on the level and spectral composition of the input signal from the microphone. Stimulus rates typically approximate or exceed 1,000 pulses/sec/selected electrode in the n-of-m and ACE strategies and approximate 250 pulses/sec/selected electrode in the SPEAK strategy. The designs of the n-of-m and ACE strategies are essentially identical and are similar to CIS except for the channel-selection feature. The SPEAK strategy uses much lower rates of stimulation and an adaptive n. The channel selection or “spectral peak picking” scheme used in the n-of-m, ACE, and SPEAK strategies is designed in part to reduce the density of stimulation whilst representing the most important aspects of the acoustic environment. The deletion of low-amplitude channels for each frame of stimulation can reduce the overall level of masking or interference across electrodes and stimulus regions within the cochlea. To the extent that the omitted channels do not contain significant information, such “unmasking” may improve the perception of the input signal by the patient. Furthermore, for positive speech-to-noise ratios, selection of the channels with the greatest amplitudes in each frame may emphasize the primary speech signal with respect to noise [5].

SPEAK is a slow-rate (180-300 pps) spectral maxima strategy. The SP extracts up to nine maxima from the incoming signal and presents these maxima to different electrodes of 20 active electrodes inside the cochlea according to their frequency. The average number of maxima is six but may vary according to the incoming signal. Frequency bands are typically allocated within the range 187-7937 Hz but alternative frequency allocations can be set.

The ACE is a fast rate, flexible spectral maxima strategy. Up to 22 channels can be used and up to 20 maxima, although 8 to 12 maxima are most widely used. The frequency band allocation is variable but usually 187-7937 Hz. Stimulation rates up to 2400 pps per channel are available with a maximum overall rate of 14400 pps. Medel n of m strategy extracts up to 11 spectral peaks from the signal to be delivered to a maximum of 12 electrodes. Fast rates up to a maximum of overall rate of 18000 pps can be used. The spectral peaks are extracted up to 7.5 kHz (1).

2.3.3 HiResolution (HiRes) Strategy and HiRes with Fidelity 120

In 1912, David Hilbert showed that signals can be decomposed into temporal envelope (the relatively slow variations in amplitude over time) and Temporal Fine Structure (TFS, a frequency modulated carrier, rapid oscillations with rate close to the centre frequency of the signal). Figure 5 illustrates an example of such a decomposition.

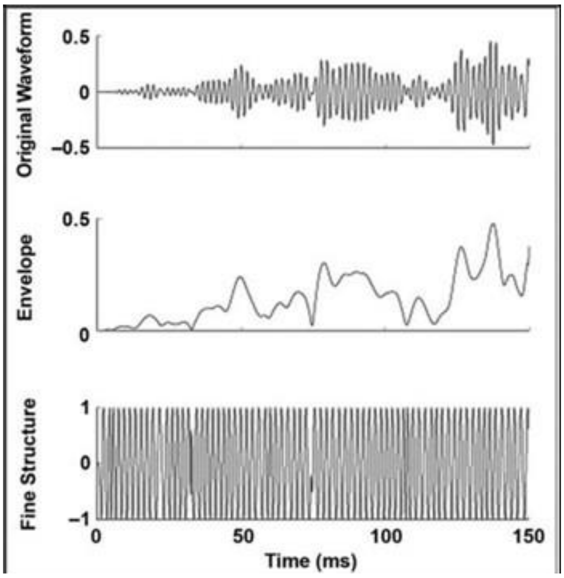


Figure 5. Decomposition of signal using Hilbert transformation.
Source: Wilson and Dorman [5].

In 2002, Smith et al. [33] investigated the importance of envelope and TFS information for auditory perception in NH subjects. They synthesized novel stimuli called "auditory chimeras" which had the envelope of one sound and the TFS of another sound. Hence, the chimeras resulted in conflicting cues—the envelope variations in a given number of bands for one sound versus the TFS variations in the same bands for another sound. Pairings consisted of sentences versus different sentences, sentences versus noise, melodies versus different melodies, and sentences with an Interaural Time Delay (ITD) which corresponded to a sound image to the left side of the subject versus the same or different signal but with an ITD to the right.

Performance depended on the type of sounds in each pairing and on the number of processing channels. Speech was identified by its envelope information for a minimum of eight channels whereas the TFS information was more important for one or two channels. For intermediate numbers of channels, both envelope and TFS information contributed to sentence recognition. On the other hand, melodies – which required greater spectral resolution - were recognized almost exclusively by their TFS information for a minimum of 32 channels and envelope cues became dominant at 48 and 64 channels. Lateralization of sentences was difficult for subjects with a small number of channels, but performance improved with increasing numbers up to the test limit of 32. Lateralization was cued by the TFS information in all cases. These findings indicated the importance of TFS information for speech perception using less than about 8 processing channels and for music perception using less than about 40 channels. In addition, the findings indicated that ITD cues may be represented by TFS but not envelope information for any number of channels up to (at least) 32. Present-day electrode arrays and processors do not support more than 4 to 8 channels of perceptually separable information. In this 4 to 8 range, both envelope and TFS information contribute to speech perception, whereas music information is conveyed almost uniquely by TFS cues [5,33].

In recent years, an important aim of CI research has been to develop strategies for enhancing the representation of TFS information so that implanted patients could benefit better pitch perception and sensitivity to ITD. Improved pitch perception should help music appreciation and convey prosody cues in speech. Thus, it may improve speech perception in particular among speakers of tonal languages, where pitch is crucial to distinguish different words. Better ITD sensitivity may help bilateral CI recipients in taking advantage of binaural cues that NH listeners use to distinguish speech among competing sound sources [33].

The HiRes strategy has been the first approach among these strategies. It uses relatively high rates of stimulation and high envelope cut-off frequencies to improve TFS information. Although only envelope information is presented with the processing strategies, frequencies included in the envelopes generally range up to 200 to 400 Hz or even higher in the HiRes strategy. Thus, substantial TFS information is represented and may be at least partially perceived within LF range [5]. An alternative approach has been to represent the TFS information within bands using multiple sites of stimulation for each band instead of a single site for each band. This approach is called the HiRes with the Fidelity 120 (HiRes 120) and is a variation of HiRes strategy. It makes use of “virtual channels” in order to increase the number of discriminable sites beyond the number of physical electrodes. The term “current steering” is also used to refer “virtual channels”. This concept was first introduced by Wilson et al. in the early 1990s [34,35,36]. With virtual channels or current steering, adjacent electrodes can be stimulated simultaneously. In this way, the perceived pitch can be shifted in any direction with respect to the percepts elicited with stimulation of either of the electrodes alone. Studies with CI recipients revealed that pitch could be manipulated through various choices of simultaneous and single-electrode conditions [37]. For example, if the most apical electrode was stimulated alone, subjects reported a low pitch. If the next electrode in the array was stimulated alone, a higher pitch was reported. For the majority of subjects, an intermediate pitch was perceived by stimulating the two electrodes together with identical in-phase pulses.

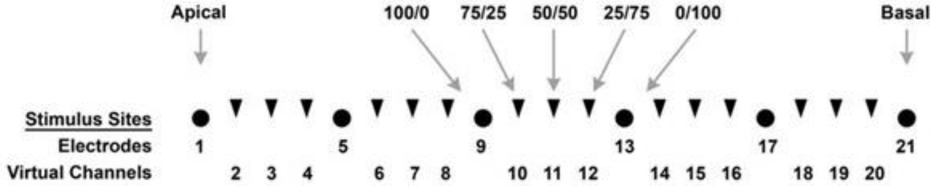


Figure 6. Diagram of stimulus sites used in virtual channel interleaved sampling processors and other similar processors.
Source: Wilson et al. [37].

The concept of virtual channels can be extended to include a quite high number of sites and corresponding pitches by using varying ratios of the currents delivered between electrodes that are simultaneously stimulated. As shown in Figure 8, stimulus site 1 is produced by stimulation of electrode 1 alone, stimulus site 2 by simultaneous stimulation of electrodes 1 and 2 with a pulse amplitude of 75 percent for electrode 1 and 25 percent for electrode 2, and so on. The total number of sites and corresponding pitches that might be produced for a good subject in the illustrated case is 21, with 6 intracochlear electrodes [37].

In the HiRes 120 strategy, 8 sites are allocated to each of 15 bandpass ranges in order to form 120 sites. The different sites for each channel are created with eight different ratios of currents delivered to the two adjacent electrodes assigned to that bandpass range. One of each of the eight ratios is used at a time, and the stimuli for the different channels are presented in a non-overlapping sequence, as in the CIS strategy. However, unlike the CIS strategy, two electrodes are stimulated simultaneously (with the selected amplitude ratio) at each update, rather than stimulation of a single electrode at each update.

The HiRes 120 strategy suggests that a higher number of available pitches may result in greater spectral resolution hence giving patients access to relatively small frequency differences thus enhancing speech perception particularly under adverse conditions, and music perception, which is generally quite poor with the CIS and other related strategies, as might be expected from the findings of Smith et al. [33]. However, a high number of available pitches or discriminable sites does not guarantee a high number of effective channels in CI recipients and furthermore “virtual pitches” may well be inherent in standard CIS and related strategies using sequential stimulation, in that intermediate pitches can also be created with non-fsimultaneous stimulation of adjacent (or more distant) electrodes so long as the pulses are relatively close in time [5,38,39,40].

2.3.4 Fine Structure Processing (FSP) Strategy

More recently, new processing options have been introduced. In one approach, the FSP strategy, the timing of positive zero crossings in the output of the bandpass filter with the lowest center frequency, or in the outputs of up to four bandpass filters with the lowest center frequencies, is marked with the presentation of a short group of pulses for the corresponding channel(s) and site(s) of stimulation instead of the continuous presentation of pulses for CIS channels. The overall amplitude of the pulse bursts for these processing channels is determined by the magnitude of energy in the band for each channel, like in CIS. The remaining higher frequency channels make use of CIS processing. The pulses for the lower frequency channels are also interleaved across electrodes, including the electrodes presenting the CIS stimuli [5,41,42].

There is some evidence that the FSP and related approaches provide an advantage compared to CIS and other envelope-based strategies to the extent that single pulses or short groups of pulses represent temporal events in the LF channel(s) better than the continuous and time varing modulations for the same channels in envelope-based strategies [5,41,42,43,44].

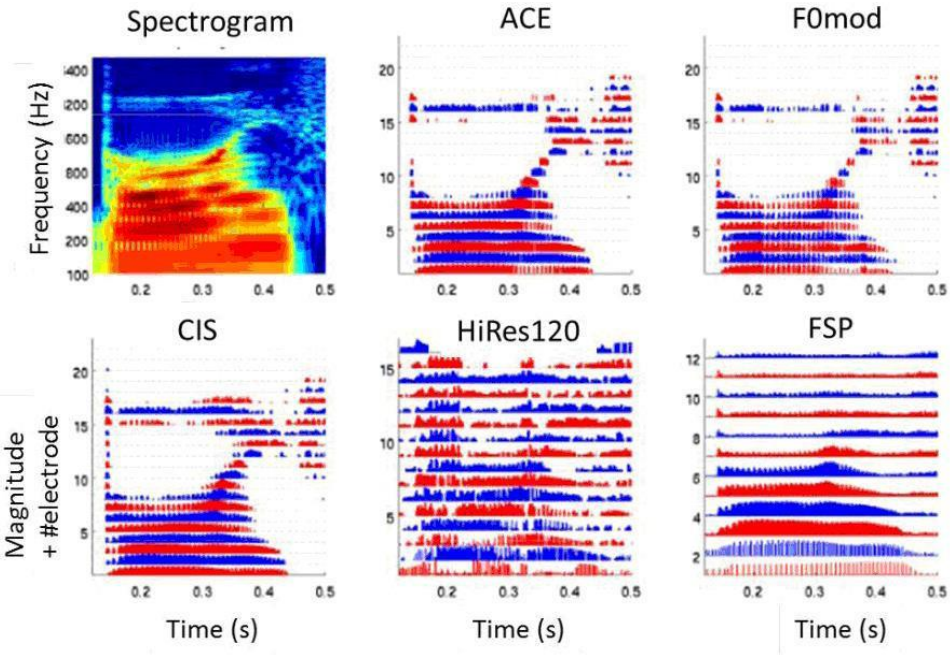


Figure 7 Spectrogram and electrodograms of the word “boy”. The signal was presented at an average RMS level of 60 dB SPL. For the electrodograms, the vertical axis indicates the channel, and the height of each vertical line represents the magnitude of the pulse. The magnitude is expressed in different units for different strategies. The red and blue colors visually distinguish adjacent channels. Source: Wouters et al. [45].

2.4 Performance with cochlear implants

Postoperative outcomes for the perception of meaningful speech information are very satisfactory in CI recipient adults and children. Current CI technology offers good opportunities for formal and informal language acquisition in deaf children [46,47,48,49] as well as very good speech understanding in quiet environments in post-lingually deafened adults [50]. Indeed, a principal conclusion of the 1995 National Institutes of Health (NIH) Consensus Conference on Cochlear Implants in Adults and Children [29] was that "A majority of those individuals with the latest speech processors for their implants will score above 80 percent correct on high context sentences, even without visual cues." [5].

However, CI recipients usually complain about speech understanding under noisy listening conditions which are very common in everyday communication environment. Findings indicate that though many of them show ceiling effects for understanding speech in quiet, their performance decreases considerably when testing takes place in the presence of background noise [51]. Another common complaint of CI recipient people is the music perception and appreciation. Although rhythm perception usually is good, they generally have difficulty hearing pitch or melody [52,53]. On the other hand, adult CI users commonly complain that they do not enjoy listening to music [54,55,56,57]. Furthermore, it may be the case that children who acquire hearing with a cochlear implant may encode sounds differently than adults. Consequently, children may be able to use acoustic cues that are not perceptible or available to implanted adults to hear and appreciate music. In a more recent research by Bruns et al, it would seem that this discrepancy in enjoyment of music between post-lingually deafened adults and prelingual children could be attributed to lack of preconception for acoustic cues and memory for melodies [58].

The trend in research on cochlear implantation over the past years has been towards improvements in speech perception in the presence of noise by implementing developments such as directional microphones and noise reduction algorithms as well as by introducing advanced settings in technical parameters such as Automatic Gain Control and Input Dynamic Range [59,60,61].

An important technological improvement in most of the SPs that are available on the market regards the use of multiple microphones in order to increase the selectivity of the directional pattern. With two microphones, sounds originating between and in front of the microphones produce microphone outputs that are in phase with each other, whereas sounds originating at other locations produce microphone outputs that are not. Summation of such microphone outputs produces larger signals for the in-phase conditions, emphasizing sounds in front of the microphones and suppressing sounds from other locations. It has been shown that the addition of a second microphone to a CI system improves speech reception performance under difficult listening conditions with reduced Signal-to-Noise Ratio SNR [5].

Binaural hearing provides important benefits in comparison to monaural hearing especially under challenging listening conditions [62,63,64]. One main advantage of binaural hearing is defined as the improvement in speech perception in the presence of noise. Three specific binaural effects are believed to benefit NH listeners and those with a hearing loss when listening to speech in noise: Head Shadow (HS), binaural SQuelch (SQ), and binaural SUmmation (SU). Although initially cochlear implantation used to be monolateral, over the past years the trend has been towards implanting patients bilaterally to make use of the binaural hearing advantages provided by the capacity of the central auditory system to process stimuli received from each ear and to reproduce it with a higher SNR by comparing interaural time and intensity differences or by the physical placement of the head which acts as an acoustic barrier and leads to an increase in SNR in the ear far from the noise when signal and noise are spatially separate.

Research in normal hearing subjects indicated a 3 dB improvement in SQ for the binaural Speech Reception Threshold (SRT) and an average increase of 3 dB SNR for HS which is more dominant for attenuation of high frequencies and can cause even 8 to 10 dB of improvement [65]. Several studies indicate that these effects may improve speech recognition in bilateral CI recipients [66]. Schleich et al. [67] evaluated SRTs in adults with bilateral CIs under three listening conditions and found an average improvement of 6.8 dB from HS, of 0.9 dB improvement for SQ and a 2.1 dB improvement from SU.

Similarly, Livotsky et al. [68] and Buss et al. [69] showed that HS was resulting in the greatest effect. An important factor that affected differences in performance was thetiming of implantation: bilateral CI use was found to be more effective when implantation was done simultaneously or sequentially with the shortest possible time interval [70].

On the other hand, the extension of indications for candidature to severe as opposed to uniquely profound hearing losses has led to an increasing number of CI recipients wearing a HA to make use of LF residual hearing on the contralateral side. Hence, contralateral HA use offers an alternative to bilateral cochlear implantation in that unilateral CI recipients benefit from the LF cues provided through acoustic signals from contralateral HA in addition to electrical signals from the CI. This has been named as “bimodal benefit” [71]. Bimodal benefit in CI recipients has recently received much attention and previous studies have shown a significant positive effect on speech recognition in noise and on functional performance in daily life as well as on the improvement of localization, pitch and music perception skills [65,72,73,74].

Previous studies showed discrepancies for correlations between bimodal benefit and audiological outcomes such as unaided pure tone average and aided free field audiometry as well as the duration of CI experience and the duration of HA experience prior to cochlear implantation. There were also studies showing the positive effect of degree of LF residual hearing or longer duration of HA experience prior to implantation on bimodal benefit [65].

Some studies even found an adverse effect of better hearing thresholds at mid-to-high frequencies [71,75]. Bimodal findings were promising for unilateral CI recipients with profound hearing loss and with no LF residual hearing especially in countries where bilateral implantation is still not reimbursed. On the other hand, in countries without any financial restrictions, the decision depends more on evaluation of the amount of benefit that a second CI or a contralateral HA can provide for individual subjects by taking into consideration better time-based cues that HA can convey to an ear with LF residual hearing in comparison to CI [76,77].

A wide range of outcomes has been found for the various multichannel implants currently on the market. Different patients using identical implant devices may show quite different speech perception scores. This indicated the importance of patient variables in the design and performance of implant systems. Such variables include differences among patients in the survivial of neural elements in the implanted cochlea, proximity of the electrodes to the target neurons, depth of insertion for the electrode array, integrity of the central auditory pathways, and pre-existing cognitive and language skills (11,31).

Chapter 3 AIM OF THE STUDY

3.1 AIM

To test the hypothesis whereby CI technology and TFS processing may influence Speech discrimination and appreciation of Music in Cochlear Implant recipients in relation to spectral resolution, pitch perception and listening mode.

3.2 OBJECTIVES

- To introduce the Italian adaptation of the STARR test, a speech assessment tool based on adaptive randomized roving levels across sentences which mimics challenging real world listening conditions. The STARR test could be a supplement to the Italian speech assessment battery for use with hearing-impaired populations with auditory prostheses and could contribute to cross-language studies.
- To study links between CI technology and TFS processing in adult CI recipients

  - The LF pitch perception outcomes linked to TFS processing capacities, in a group of CI recipients;

  - The Italian STARR outcomes in relation to another Speech in Noise test (OLSA), the interaction between TFS processing and speech perception outcomes in particular those of the Italian STARR;

  - The correlations between Pitch Processing and speech perception outcomes in relation to the amount of monoalateral and bimodal benefit;
- To analyse outcomes for Appreciation of Music in Adult CI recipients in relation to Pitch perception;

  - Outcomes for low frequency Pitch tests (Harmonic Intonation-HI / Disharmonic Intonation – DI) thought to depend on availability of TFS cues crucial for Speech perception;

  - Relate Music trials with outcomes for Pitch tests in study group;

  - Attempt to correlated findings with current CI technology.

Chapter 4 ADAPTATION OF THE ‘STARR’ TEST FOR ADULT ITALIAN POPULATION: Speech Test for Realistic Estimate in Real Life Listening Conditions

4.1 Introduction

Everyday communication environment due to changes in both speech and noise levels usually brings challenges to listeners, in particular to people with auditory prostheses in situations such as group conversation where the recipient may have to deal with someone who uses quite a large vocal effort and is placed close to the recipient’s microphone and someone who uses much less vocal effort and is located further from the recipient’s microphone [51]. In the attempt to develop a speech perception test that was representative of adverse listening situations and sensitive to differences in performance between various settings in auditory prostheses, Boyle et al. [78] and Haumann et al. [79] introduced a new test approach which was based on measurement of speech perception using a roving-level adaptive method where the presentation level of both speech and noise signals varied across sentences. The two factors that resulted in the greatest variation to listeners were change in presentation level and modification of signal-to-noise ratio (SNR). Whilst the sentences were presented at a roving level, the noise was adapted automatically to obtain SNR at which the subject reaches the 50% correct level referred to as the Speech Reception Threshold (SRT). The test required the listener to understand speech in the presence of competing noise and to do this for an unpredictable presentation level that should cover the constantly changing range of levels that might be encountered in most everyday life situations. At the initial application of this test, Boyle et al. [78] studied difference in performance between two types of signal processing in 6 cochlear implant (CI) recipients using a crossover design and they found significant differences after one month of experience. Subsequently, Haumann et al. [79] investigated the effect of processor models in 55 CI users using a German test material based on HSM sentences [80]. The participants were divided into five groups according to their CI processor model and all groups were matched for demographic factors and traditional speech perception test scores. The groups showed significant outcome differences when tested with roving-levels across sentences (the mean SRTs ranged from -1 to +6.4 dB), although they performed similarly on HSM sentence test using a fixed speech presentation level (at 65 dB SPL) and a fixed CCITT noise (+10 dB SNR).

These findings suggested that a test using roving levels could reveal differences between processor designs that were not shown when a fixed presentation level and fixed SNR were used.

Later on, Boyle et al. [51] created the Sentence Test with Adaptive Randomized Roving level (STARR) test which made use of the IEEE sentences [81] by adapting them into British-English. The recordings were done by two native speakers, one male and one female, where three consecutive lists of IEEE sentences were combined to one STARR test list which resulted in a total of 25 lists, each with 30 items. The STARR test was applied to adults with normal hearing (NH) as well as to a group of adult CI recipients and the outcomes supported in particular a noticeable difficulty of CI users under these challenging test conditions, showing that for normal hearing subjects the effect of roving was minimal whereas for CI users it was much greater [51].

The present study aimed to introduce an Italian version of the roving-level adaptive test method. For this purpose, the STARR test was adapted into Italian, normative data were collected for an adult population, and interlist-variability as well as learning effects were investigated.

4.2 Materials and Method

4.2.1 Participants

A group of 32 NH adults (15 female and 17 male) reporting no otologic history and no hearing complaints participated in the study. Their ages varied between 18 and 53 years (mean=32yrs, SD=11). All had hearing thresholds ≤ 20 dB HL for frequencies between 250-8000 Hz on both ears (mean=10 dB HL, SD=6). This study was approved by the Local Ethical Committee and subjects' consent was given freely.

4.2.2 Italian STARR test design

The Italian STARR test made use of sentences from the standard Italian speech recognition test developed by Cutugno et al. [82]. The original corpus consisted of 200 meaningful sentences which were organized into 10 test lists, each of 20 sentences. The sentences were selected based on lexical and morpho-syntactic characteristics in order to make them more easily accessible to a heterogeneous group, coming from different regions of Italy [83]. Independent, short, simple everyday sentences were used. 100 phonemically balanced bisyllabic words [84] were included within the sentences; criteria regarding the frequency of use in contemporary Italian were followed [85,86].

The recordings were done in a double-walled sound room with additional sound absorption material on the walls and ceiling (complied with ANSI S3.1-1999) [87]. The speaker was a male native Italian phonetician using clear conversational speech with natural rate and natural vocal effort without emphasizing any key words. During recordings the speaker was monitored by another phonetician in order to control pronunciation errors. Sentences misread or pronounced incorrectly were repeated. Recordings were made directly to digital format with a Neumann condenser microphone (TLM 193) which was directly connected to a Tascam DAT recorder with 16-bit resolution at 44100 Hz. Following the recording session, the contents of the DAT were transferred digitally to a computer. All signals were analysed using KAY 4300 Computerized Speech Laboratory (CSL) system. Equalization of the levels, when necessary, was done using a software for digital processing of audio signals (Sound Forge 4.0, Sonic Foundry). For each sentence, the integral value of RMS expressed in dBV was calculated and subsequently all sentences were equalized considering ±1 dB as normal range in respect to the average value. Sentences were separated by seven-second intervals.

In the Italian STARR test, sentence material recorded by Cutugno et al. [82] was transferred digitally to a PC and edited using the CoolEdit 2000 sound-editing software. Each sentence was isolated and adjusted in level such that the total root-mean-square power was 20 dB below full scale for a.wav file, i.e. -20 dBV. In pilot testing of STARR test, it was found that a set of 15 sentences provided a practical balance for a clinical test: sufficient sentences were delivered to allow convergence around an SNR range that represented the abilities of a subject to understand speech in the presence of competing noise, while avoiding an overly long test that would induce fatigue [88]. Therefore, 150 sentences were selected from the original corpus of 200 sentences and each test list consisted of 15 sentences (5 sentences were removed from each test list of Cutugno *et al.* [82] maintaining their list boundaries). The sentences typically consisted of 5 words with 9 to 13 syllables (median = 5 words across all sentence lists as well as for individual lists, range 3-7). Each sentence was used only once during the whole procedure in order to avoid repetitions in case of outcome comparisons using different test lists during same sessions. Three key words (noun, verb, adverb or adjective) were allocated for each sentence and speech-shaped noise was used as competition. Figure 8 illustrates both the spectrum of the first 10 sentences from list 1 and that of the noise. The speech spectrum showed maximum energy between 100 and 500 Hz followed by a drop with an average slope of 10 dB per octave up to 8000 Hz. The noise had a spectrum resembling long-term spectrum of the speech test material and was based on recommendations of the Institute of Hearing Research (IHR) in the UK. The spectrum of the noise used by the IHR - produced by shaping a white noise and used for various IHR speech tests - was flat from 100 Hz to 1000 Hz and then dropped with an average slope of 11 dB per octave to 8000 Hz.

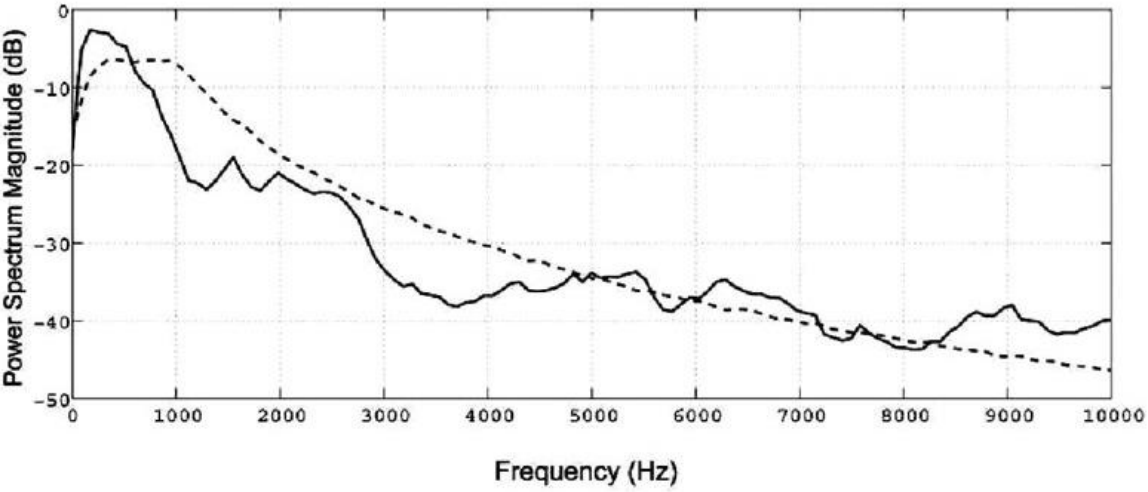


Figure 8 The spectrum of the speech (continuous line) and that of the noise (dotted line). The speech spectrum showed maximum energy between 100 and 500 Hz and then dropped with an average slope of 10 dB per octave up to 8000 Hz. The noise spectrum was approximately flat from 100 Hz to 1000 Hz and then dropped with an average slope of 11 dB per octave up to 8000 Hz

The STARR software was written in Visual Basic and provided the clinician with a graphical user interface to both deliver and score the test. Three presentation levels were used within each test list: 50, 65 and 80 dB SPL. This range was selected in line with the levels that were typically explored in hearing instrument research. Once the clinician selected a test list to be presented, the software randomly selected both a sentence and a presentation level at which to begin the procedure. This process continued with any sentence from the list being presented only once and an equal number of presentations being made at each presentation level, i.e. 5 at 50, 5 at 65 and 5 at 80 dB SPL. Each time any test list was presented to a new subject, a different presentation order of the sentences was likely to occur and any given sentence was likely to be presented at a different presentation level.

4.2.3 Procedure

All participants underwent testing in a sound-proofed booth of 2x2 metres. The stimulus was presented via the PC and a preamplifier connected directly to a single loudspeaker. Both sentence material and noise came from 0° azimuth with loudspeaker at 1m from participant’s head. Before each test session, calibration was performed. Each sentence was matched in terms of RMS energy. All gains needed to apply the desired sentence presentation level and the SNR were automatically calculated by the program software. This meant that only a single point calibration was necessary. A 10 second duration noise burst, centred around 1 kHz, was used for calibration. The burst was arranged to be 11 dB higher than the sentence RMS level. By delivering the calibration noise burst such that a level of 91 dB SPL was observed on a sound level meter, placed where the subject's head would be during the test, the correct sentence levels would be automatically produced by the software. For example, a gain of -11, -26 or -41 dB was applied to a sentence to produce maximum, middle or minimum presentation levels of 80, 65 and 50 dB SPL respectively. In order to verify inter-list variability, all participants were tested using all 10 test lists. List order was counterbalanced by staggering the allocation of test lists to participants. No practice list was used in order to investigate any learning effects. The test was started after explaining the task to participants and warning them about the possibility of sometimes hearing quite loud stimuli. The upper sound pressure limit was set to 91 dB SPL resulting in a -10 dB SNR for the highest sentence presentation level of 80 dB SPL. Beyond this limit, the program automatically decreased both the sentence and noise levels, maintaining the required SNR while avoiding uncomfortable stimuli. However, this happened very rarely in practice as it was extremely difficult to understand the speech at the -10 dB SNR, especially at such high presentation levels. Participants were told that they could ask for a break whenever they needed but none of them requested it, since a test session including hearing assessment never exceeded 45 minutes. After presentation of a sentence, participants were asked to repeat it as accurately as possible, while explaining that not every word needed to be correct and encouraging them to guess when not sure.

For scoring, single key words were highlighted in white on the screen becoming green to indicate words repeated correctly. The clinician could either click on single key words or, alternatively, select the "all" or "none" options to indicate participant's response. The SNR was +20 dB initially and varied adaptively according to the participant's response. Where the minimum specified number of key words was correctly identified, typically 2 out of 3, the sentence was considered correct and the SNR used for the next sentence was made more adverse. Where insufficient key words were correctly identified, the sentence was considered incorrect and the SNR for the next sentence to be presented would become more favourable. The initial step size was 10 dB, dropping to 5 dB after the first reversal of the adaptive track and dropping again to the final step size of 2.5 dB after a further reversal. The SNR was varied by adjusting the noise level while keeping the speech level at 50, 65 and 80 dB SPL and using the same SNR for all three levels. The SRT was computed by averaging the SNRs for the last nine (7th to 15th) sentences, along with the SNR at which a next (16th) sentence would have been presented. Clicking on the OK button finalized scoring for individual sentences and the next sentence was presented by the software until the end of the test list where the SRT value was displayed automatically.

4.3 Results

4.3.1 Normative Data

A mean SRT was calculated averaging the SRTs for each participant across all test lists. The mean SRTs for individual participants ranged between -5.8 to -9.7 dB SNR. Subsequently, an overall mean SRT was produced averaging the SRTs across all participants (Mean SRT= -8.4 dB SNR, SD=0.9).

4.3.2 Inter-List Variability

To assess the variation in difficulty between the test lists and to calculate deviations from the participant's mean, the SRTs across all test lists were averaged for each participant and the SRT score for each test list was subtracted from the average SRT for that participant. The deviations for each test list were averaged across all participants in order to calculate correction factors for each list. Figure 12 represents the deviation of SRT values from the overall mean for each test list. The deviations for lists 1, 3, 4, 6, 8 were reasonably small (<0.5 dB). The largest deviations were 0.9 dB for list 7 and 1 dB for list 10. The SD of the deviations was 0.7 dB.

To assess the variation in difficulty between the test lists and to calculate deviations from the participant's mean, the SRTs across all test lists were averaged for each participant and the SRT score for each test list was subtracted from the average SRT for that participant. The deviations for each test list were averaged across all participants in order to calculate correction factors for each list. Figure 12 represents the deviation of SRT values from the overall mean for each test list. The deviations for lists 1, 3, 4, 6, 8 were reasonably small (<0.5 dB). The largest deviations were 0.9 dB for list 7 and 1 dB for list 10. The SD of the deviations was 0.7 dB.

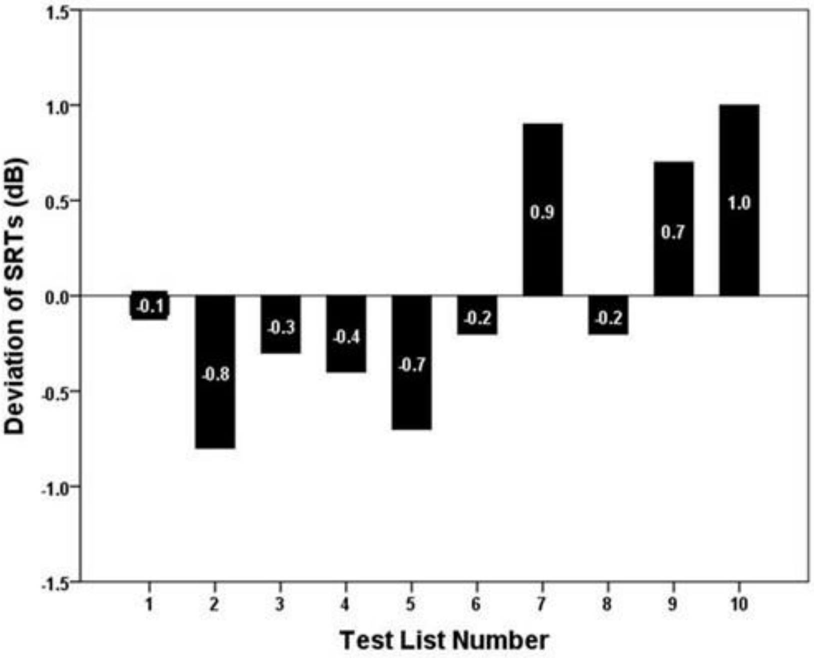


Figure 9 Deviation of SRT values from overall mean for each test list. SRT, speech reception threshold.

The SRT obtained using a given test list was corrected by subtracting the deviation for that list in order to compensate for differences in list difficulty. The reliability of SRT estimates was assessed by calculating the SD across all corrected SRTs (of all test lists) for each participant and this SD was averaged across all participants. This value was compared to the average SD without applying correction factors in order to evaluate the effect of the application of correction for each test list. The average SD across all corrected SRTs was 1 dB in comparison to the average SD of 1.2 dB without applying correction factors (uncorrected SRTs).

To determine the meaningful difference in SRT value for an outcome comparison (e.g. when comparing two different CI processors) by using one single test list under each listening condition, the SD of the difference in SRT between two lists was √2 times SD. A difference was considered meaningful if it was more than 2 SD; therefore, the difference in SRT was calculated as 2√2 times SD. Therefore, if an outcome comparison is done by using one single test list under two listening conditions, a difference of 2.8 dB in SRTs would be considered meaningful and this value would be 3.4 dB in case of uncorrected SRTs. If two test lists were to be used per condition, the corresponding values would be 2 dB and 2.4 dB for corrected and uncorrected SRTs respectively.

Statistical analysis was carried out using Statistical Package for Social Sciences (SPSS) version 19.0 (Chicago, IL, USA). One-way repeated measures analysis of variance (ANOVA), with Greenhouse-Geisser correction, was conducted by using paired t-test procedures with Bonferroni correction for multiple comparisons in order to investigate the statistical differences between the test lists with uncorrected and corrected SRTs. A significant main effect of the test lists was found for uncorrected SRTs [F(6.16, 191.19)=12.17,p=0.000] where post hoc tests indicated significant differences of lists 7, 9 and 10 to the rest of the test lists (p<0.005). However, no significant main effect of test list was found with corrected SRTs [F(6.16, 191.19)= 0.13, p=0.993].

4.3.3 Learning Effects

To evaluate learning effects, the SRTs were averaged according to the order of testing without considering the list number. The SRTs averaged across test lists that are administered in the same serial position (first, second, etc.) are illustrated in Figure 13. The mean SRTs for test lists 1 to 10 changed within a range of 0.7 dB. A one-way repeated-measures ANOVA did not show any significant main effect of the list presentation order [F(9, 279)=1.60,p=0.114].

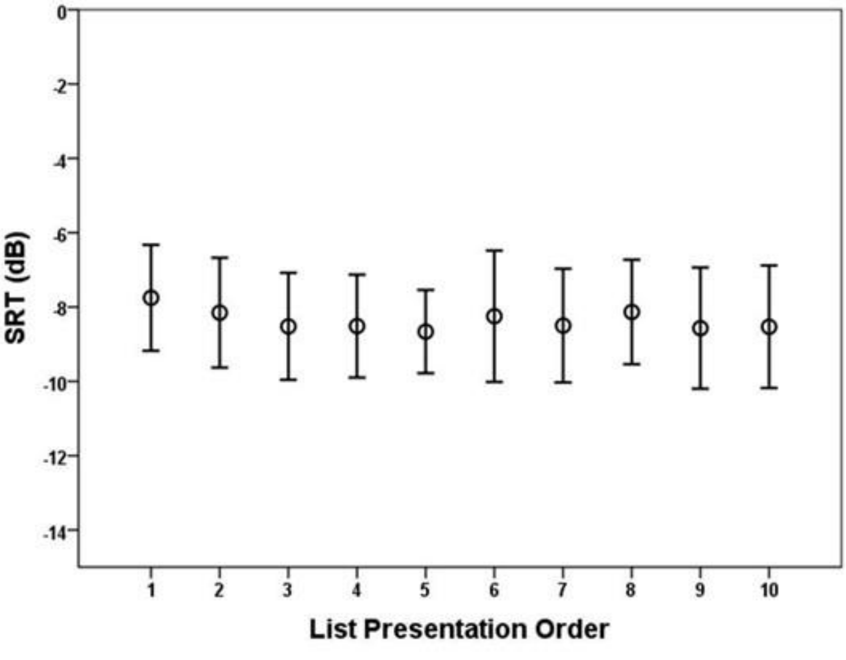


Figure 10 The SRTs averaged across test lists considering list presentation order. Error bars show ±1 SD. SRT, speech reception threshold.

4.4 Discussion

Improvements in prostheses such as hearing aids and cochlear implants are very crucial to enabling the patient to reach her/his best potential. In parallel to such progress, there is an increasing need for new speech assessment tools that mimic challenging real life listening conditions where background noise is usually present and speech level varies according to vocal capacity and distance of the speaker. A recent attempt to meet this need has been done by Boyle et al. [51] by introducing the STARR test which is based on adaptive randomized roving levels across sentences and the present study is the Italian adaptation of the original STARR test. This test was found particularly sensitive to lower level speech and is believed to provide a better estimate of improvements in technical settings of auditory prostheses [51].

Outcomes for STARR in a British-English NH population [50] were based on two speakers- male/female- whereas the Italian version had a male speaker only as in Cutugno et al. [82]. The mean SRT in Italian NH population (-8.4 dB, range -5.8 to -9.7 dB) was similar but slightly lower than the mean SRTs for the male (-6.1 dB, range -2.8 to –9.3 dB) and the female (-5.7 dB, range -0.8 to -10.3 dB) speakers as well as the mean across them (-5.9 dB, range -2.1 to -9.6 dB) studied with the original version of the STARR test. Such outcome differences have been found in other multilingual speech perception tests such as the Oldenburg Sentence Test (OLSA)/Matrix Sentence Test [89] and the Hearing in Noise Test (HINT) [90] and can be attributed to language and speaker dependent factors [91].

A key aim of this work was to verify and strengthen the clinical reliability of the Italian STARR test. For this purpose, the variations in difficulty between the test lists were assessed. The variability of mean SRTs across test lists for NH subjects was ≤ 1 dB for all sentence lists. The findings indicated that the SD of the deviations was 0.7 dB. This value was 0.6 dB in the case of Boyle et al. [51] and both outcomes were highly consistent. Although the deviations for the majority of the Italian STARR test lists were reasonably small (<0.5 dB), they were higher than in multilingual matrix tests (range 0.13 to 0.2 dB) [92]. Some test lists (7, 9 and 10) had considerably larger deviations and showed statistically significant differences with the rest of the test lists. However, the SRT obtained through a certain list was corrected by subtracting the deviation for that list from the raw SRT and the statistically significant differences between test lists were eliminated with this procedure.

On the other hand, the average SD across all corrected SRTs was 1 dB in comparison to the average SD of 1.2 dB without applying correction factors. The benefit of applying correction factors was remarkable. Outcome comparisons using one single test list under two different listening conditions in a within subject design considered a difference of 2.8 dB to be meaningful in corrected SRTs versus 3.4 dB in case of uncorrected SRTs. The corresponding values became 2.0 dB and 2.4 dB for corrected and uncorrected SRTs respectively when two test lists were used to assess each condition. In the light of these findings, the authors suggest using two test lists per condition and applying correction factors in order to ensure more reliable outcomes. Furthermore, in order to avoid a reduction in the number of available test lists while maintaining reasonably low interlist-variability, it would seem worthwhile to increase the number of items for each test list as in the original version of the STARR test by adding 15 sentences recorded by the female voice to each list. This would lead the SRT to be computed by averaging the SNRs for a total of 20 sentences instead of 10 in the present Italian version and would help the variability to decrease as the number of independent items scored in the task would be increased [93]. Alternatively, equalisation of sentence intelligibility, which is a standard reliability optimization procedure, could be applied [89,90,91,94]. In the Italian STARR test, the question of accounting for differences afterwards was preferred as in the original version in order to maintain low inter-list variability; applying correction factors was required to compensate for differences in difficulty across test lists so that they effectively provide consistent outcomes and can be used interchangeably.

Learning effects were investigated in that no practice list was used prior to testing and the SRTs were compared only considering order of list presentation. Similarly with Boyle et al. [51], the mean SRT for the first list presented was slightly higher than the next lists and the mean SRTs (lists 1-10) covered a range of only 0.7 dB. Moreover, statistical analysis confirmed no significant main effect of order of list presentation. Nevertheless, it should be considered that the SRT calculation of the STARR test is based on averaging the SNRs for the last nine sentences along with the SNR at which a next sentence would have been presented and the experience through initial sentences may have contributed to learning of the NH participants. However, it seems reasonable to expect that people with auditory prostheses may not compensate as quickly as NH people. Therefore, it would be better to administer one practice list before a test session with STARR in order to minimise learning effects in populations with hearing impairment. This recommendation can also be supported with the findings of other speech recognition tests such as the Canadian Francophone HINT [95] and the Italian OLSA Test [94] that have found significant differences between testing trials suggesting a possible practice effect in NH populations.

It has been shown that speech intelligibility can be affected by phonemic content, word familiarity, sentence length, RMS levels and sentence difficulty [90,96]. Hence, the creation of Italian STARR lists, based on the original material provided by Cutugno et al. [82], took into consideration aspects such as word familiarity, sentence length, syntactic structure, equalization of RMS level and minimum number of items per test list. However, an important limitation of the Italian STARR test was that the phonemic distribution of the entire original sentence set was not matched although phonemically balanced bisyllabic words [84] were equally distributed throughout the test lists. Moreover, further studies will be required to investigate use in people with auditory prostheses who have greater difficulty in understanding speech, especially at lower levels and in the presence of noise. Such outcomes may provide useful information for optimizing fitting procedures, patient performance and contributing to cross-language studies.

4.5 Conclusion

The present study showed findings for the Italian STARR test which was based on a roving-level adaptive method. Outcomes were in line with previous research in NH population and the variability of mean SRTs across lists was relatively small (≤ 1 dB for all test lists). The benefit of applying correction factors was basic to improval of reliability. Statistical analysis showed no significant learning effects.

* Material covered in Chapter 4 has previously been published in Dincer D’Alessandro H, Ballantyne D, De Seta E, Musacchio A, Mancini P (2016). Adaptation of the STARR test for an adult Italian population: A speech test for a realistic estimate in real-life listening conditions. Int. J Audiol 55:262-267.

Chapter 5 TEMPORAL FINE STRUCTRE PROCESSING, PITCH AND SPEECH PERCEPTION IN COCHLEAR IMPLANT RECIPIENTS

5.1 Introduction

The envelope and Temporal Fine Structure (TFS) are known to be two important acoustic cues for speech intelligibility [97]. The envelope is an amplitude modulated carrier with relatively slow modulations over time that are superimposed on a more rapidly varying TFS which is a frequency modulated carrier [81]. The envelope traditionally has been considered as the most important carrier of information for speech signals. However, envelope cues alone are insufficient to permit the spatial separation of multiple voices and therefore, it is believed that TFS information may be crucial for speech perception in the presence of noise, especially for the ability to listen in fluctuating background sounds [79,81]. Present-day cochlear implant (CI) design allows conveying mainly envelope information in different frequency bands whilst TFS information is mostly discarded and this fact is thought to contribute to the big difficulty of CI recipients when listening in the presence of noise while most of them show high level of speech understanding in quiet [51,81,98].

TFS processing has proved to contribute to pitch perception which is important for speech perception [80,81]. Neurophysiological studies in animals have shown that pitch can be encoded by two auditory mechanisms: place coding that is based on tonotopic excitation where pitch cues are conveyed through spatial alteration of nerve fibres and phase locking which is a time-based mechanism that locks onto the TFS of the signal [99]. Place coding is more dominant in conveying pitch for high frequency (HF) signals whereas low frequency (LF) coding is dominated by phase locking and TFS processing is assumed to be reflected in the phase locking patterns within the auditory nerve [100].

Two recent pitch perception tests, A§E Harmonic Intonation (HI) and Disharmonic Intonation (DI), are expected to be indicative of the capability of the ear to use its phase locking mechanism and therefore, are believed to depend on the availability of TFS cues since both tests focus on the spectral discrimination of the auditory system in the LF domain linked in particular to Fundamental Frequency (F0) [101,102]. Previous HI/DI outcomes in CI recipients revealed abnormal pitch perception in comparison to normal hearing (NH) subjects [76,101,103]. Furthermore, Vaerenberg et al. [104] studied in 6 adult CI users outcomes for pitch and speech perception comparing electric stimulation only to electro-acoustic stimulation (EAS). Differences for speech perception performance in quiet were not statistically significant. However, results indicated a remarkable improvement in DI with an EAS processor compared to electrical stimulation only, just as speech perception outcomes at +10 dB SNR showed statistically significant differences. When testing speech perception in noise, their study has used a method based on a fixed Signal-to-Noise Ratio (SNR) where speech-shaped noise fixed at 65 dB SPL was presented with varying speech signal (SNRs at +10, +5, 0, -5 dB) using monosyllabic CVC words and performance was measured as a percent-correct score at that SNR.

However, in real world listening conditions, the presence of noise in a fluctuating background together with the variability of speech levels according to vocal capacity as well as to the distance of the speaker brings further challenges to listeners. A new assessment tool, the STARR test [51,79], is based on measurement of speech perception using a roving-level adaptive method. Whilst the sentences are presented at a roving level, the noise is adapted automatically to obtain SNR at which the subject reaches the 50% correct level referred to as the Speech Reception Threshold (SRT). The STARR test in British English language was applied to adults with normal hearing as well as to a group of adult CI recipients and the outcomes supported the big difficulty of CI users under these challenging test conditions [51]. The STARR test has been also adapted into Italian language and normative data as well as inter-list-variability and learning effects were studied previously [105]. However, outcomes for Italian speaking CI users are so far missing.

The above-mentioned background as well as lack of outcomes regarding TFS processing in relation to speech perception skills, especially when both speech and noise signals are fluctuation just as in realistic listening conditions lead us to conduct the present study with the aim to investigate in adult CI recipients:

- The LF pitch perception outcomes linked to TFS processing capacities, in a large group;
- The Italian STARR outcomes in relation to other speech perception tests;
- The interaction between TFS processing and speech perception outcomes in particular those of the Italian STARR;
- The effects of variables such as age, the duration of profound deafness before implantation, the duration of CI experience and hearing thresholds;

- The correlations between TFS processing and speech perception outcomes in relation to the amount of bimodal benefit.
-

5.2 Materials and Method

5.2.1 Participants

The participants in the study were 43 post-lingually deafened adult CI recipients (17F, 26M), aged 18-83 years (mean=57yrs, SD=17.2) at the time of testing. All were consistent CI users with at least 3 months of CI experience (range 3 to 208mths, mean=52.7mths, SD=50.9). The mean duration of profound deafness before implantation was 75 months (range 3 to 360mths, SD=98.6) and this value was longer than 10yrs for 37% of the group. The study group consisted of 23 unilateral, 6 bilateral and 14 bimodal listeners. All bilateral CI recipients were implanted simultaneously with the same CI model on both ears. There was one exception who had a different model on the contralateral ear 3,5yrs after the first implantation (a Cochlear Freedom device on one ear and a MedEl Concerto device on the contralateral ear).

Assessment regarded a total of 49 CI ears where 15 were implanted with Advanced Bionics devices [14 ears implanted with 90K device and fitted either with HiRes-S (n=8), HiRes-S with Fidelity 120 (n=5) or HiRes-P with Fidelity 120 (n=1) plus 1 recipient implanted with C1 device and fitted with CIS]; 30 were implanted with Med-El devices [all implanted with the Concerto and fitted either with FS4 (n=20) or FS4-p (n=10)]; 4 were implanted with Cochlear Freedom devices [all implanted with CI24RE and fitted with ACE]. The number of active electrodes all ranged from 12 to 22 in the group. The bandwidth for the most apical electrode varied between 250 to 416 Hz for HiRes users, from 238 to 442 Hz for Fidelity 120 users, 198 to 325 Hz for FS4 users, 188 to 313 Hz for ACE users and 350 to 494 Hz for the CIS user.

All unilateral and bilateral CI listeners showed no degree of LF residual hearing in both ears that may have interfered with pitch perception outcomes (hearing thresholds ≥ 85 dBHL for frequencies 125-1000 Hz), whilst bimodal listeners did have residual hearing on side with contralateral hearing aid (HA). All participants had CI thresholds that were ≤ 40 dB HL (mean= 34.5 dB HL, SD= 5.7) for octave frequencies between 125-8000Hz. For bimodal listeners, mean unaided threshold on contralateral ear (125-4000 Hz) was 83.8 dB HL (range 59.2 to 112.5 dB HL, SD=13.8) and mean aided threshold was 63.8 dB HL (range 37.5 to 81.7 dB HL, SD=13.3).

This study was approved by the Local Ethical Committee and subjects' consent was given freely.

5.2.2 Procedure

5.2.2.a Fitting

CI programs for individual recipients were controlled prior to testing. Since existing HA programs were to be used during testing, all bimodal listeners were asked to visit their HA providers shortly before their appointment in our centre. Furthermore, following a regular CI fitting session, most comfortable levels were verified in live-speech when listening together with contralateral HA in order to avoid any discomfort due to a loudness summation effect

5.2.2.b Hearing Thresholds

Unaided pure tone thresholds were recorded via an Aurical audiometer and TDH39 headphones in a standard sound-proofed booth at octave frequencies between 125-8000 Hz using a warble tone as were aided thresholds in Free Field through a loudspeaker placed at 0^o azimuth at 1m distance from the subject's head.

5.2.2.c HI/DI Performance

A§E - HI/DI tests were used separately in each ear to evaluate LF pitch perception skills that are linked to TFS processing in CI users whilst performances were measured additionally in the CI plus HA listening condition for bimodal users. CI only condition in bimodal listeners was performed by occluding the contralateral ear with an ear foam plug plus a circum-aural headphone to avoid a potential contribution of the non-implanted ear [73].

HI/DI tests were based on a discrimination task between two consecutive complex tones: one intonating and one non-intonating. The non-intonating stimulus represented a harmonic complex signal of a F0 at 200 Hz and its 3 higher harmonics presented at levels lower than F0 (-6 dB at 2F0, -12 dB at 3F0 and -18 dB at 4F0). In both tests, the non-intonating sound was contrasted by an intonating sound. In the HI test the intonating sound was characterized with a frequency sweep of F0 together with all harmonics [from NF0 to $N(F0+\Delta F)$, N ranging from 1 to 4] whereas in the DI test a sweep of only F0 [F0 to $F0+\Delta F$] was used resulting in a disharmonic intonation. For both tests, the sweep, that was linear, lasted 120ms and was introduced at 330ms after the start of the signal. Total duration of each stimulus was 600ms and the two consecutive stimuli were separated with a 500ms inter-stimulus interval. White noise was added to the stimuli (SNR +10.9 dB) so that they sounded more natural and intensity roving (±2 dB) was applied in order to avoid the use of loudness cues [101].

HI/DI test orders were counterbalanced across subjects in order to minimize the learning effects. Two consecutive stimuli were presented to subjects at 70 dB SPL who were asked to discriminate between same or different. Testing was carried out after training for both the stimuli and the test task.

For each subject, the Just Noticeable Difference (JND) - the smallest ΔF that the subject could discriminate - was calculated by the software using an adaptive staircase procedure where ΔF (41 Hz at test start, range 0 to 214 Hz, 0 Hz means no change between two signals) was increased for an incorrect response and decreased for a correct one until reaching an estimate of the 50%-point on a subject's psychometric curve. If it could not be found within 100 trials, JND was set to 220 Hz which was above the maximum ΔF value [76,101]. This adaptive staircase procedure is described in detail in a study by Vaerenberg et al. [106].

5.2.2.d Speech Perception

Similarly, speech perception performance was evaluated ear by ear for all participants and additionally in bimodal listening condition for contralateral HA users. The stimulus was presented via a computer and a preamplifier connected directly to a single loudspeaker. Both sentence material and noise were presented from a loudspeaker at 0° azimuth at 1m from participant's head.

The test battery consisted of standard Italian phonetically balanced bi-syllabic words for an adult population [82]. Speech recognition in quiet (Word Recognition Score- WRS) was tested at 65 dB SPL whilst the performance in noise was evaluated presenting words at +10 and +5 dB SNR ((SNR+10 and SNR+5) with speech-shaped noise at 65 dB SPL.

5.2.2.e Italian STARR

The Italian STARR test made use of sentences from the standard Italian test battery [82]. The material consisted of 10 test lists each containing 15 sentences, all recorded with male voice. For competition CCIT noise was used resembling long-term spectrum of the speech test material. Three presentation levels (50, 65 and 80 dB SPL) were used for sentences and there were 5 presentations at each of these levels within each test list. The details for Italian STARR test design are described in a previous publication [105].

The test was started after explaining the task to participants and warning them about the possibility of sometimes hearing quite loud stimuli. The upper sound pressure limit was set to 91 dB

SPL resulting in a -10 dB SNR for the highest sentence presentation level of 80 dB SPL. Beyond this limit, the program automatically decreased both the sentence and noise levels, maintaining the required SNR while avoiding uncomfortable stimuli. After presentation of a sentence, participants were asked to repeat it as accurately as possible, while explaining that not every word needed to be correct and encouraging them to guess when not sure. For scoring, and for the benefit of the operator, single key words were highlighted in white on the screen becoming green to indicate words repeated correctly.

The clinician could either click on single key words or, alternatively, select the “all” or “none” options to indicate participant’s response. The SNR was +20 dB initially and varied adaptively according to the participant’s response. Where the minimum specified number of key words was correctly identified, typically 2 out of 3, the sentence was considered correct and the SNR used for the next sentence was made more adverse. Where insufficient key words were correctly identified, the sentence was considered incorrect and the SNR for the next sentence to be presented would become more favourable. The initial step size was 10 dB, dropping to 5 dB after the first reversal of the adaptive track and dropping again to the final step size of 2.5 dB after a further reversal. The SNR was varied by adjusting the noise level while keeping the speech level at 50, 65 and 80 dB SPL and using the same SNR for all three levels. The SRT was computed by averaging the SNRs for the last nine sentences, along with the SNR at which a next sentence would have been presented. Clicking on the OK button finalized scoring for individual sentences and the next sentence was presented by the software until the end of the test list where the SRT value was displayed automatically. All participants were tested using two test lists following a practice list which was administered to minimize any learning effects. All the tests were administered on the same day.

5.2.2.f Statistical Analysis

Data analysis was carried out using the Statistical Package for Social Sciences (SPSS) version 19.0 (Chicago, IL, USA). Non-parametric statistical tests were used since one-sample Kolmogorov-Smirnov tests revealed that outcomes were not normally distributed ($p<0.001$). Differences between tests were investigated by Wilcoxon Signed Ranks test as well as differences between CI only and bimodal listening conditions within subjects. Percentage of outcomes that fell within the normal clinical zone (JNDs ≤ 4 Hz for HI and ≤ 10 Hz for DI as per normative data collected by Vaerenberg et al. in adult population) was computed in order to compare outcomes to those of people with normal hearing [101]. For STARR test, the scores obtained from two lists were averaged for each ear/condition and percentage of meaningful STARR score was calculated considering scores that were less than 20 dB SNR. Spearman rank-order correlations were used to investigate the correlations between HI/DI tests and speech perception outcomes as well as to evaluate the effect of variables such as age, duration of CI experience, duration of deafness, unaided/aided PTA thresholds (PTA <1000 Hz, PTA≥1000 Hz, PTA ≤8000 Hz with CI and ≤4000 Hz with HA). The cut-off level for statistical significance was set to 0.05.

5.3 Results

5.3.1 HI/DI Performance

The analysis regarded outcomes of unilateral/bilateral CI users (based on individual ears) plus CI only condition in bimodal listeners (n=49). Pitch perception results between HI and DI tests indicated statistically significant differences (Z= -5.5,, p<0.001): outcomes were better for HI than for DI [median JND= 27.0 Hz (range 1.0 to 220.0 Hz) and 147.0 Hz (range 7.0 to 220.0 Hz) for HI and DI respectively]. 12% of HI outcomes were within the normal zone versus 8% of DI outcomes.

The correlations between HI and DI outcomes were statistically significant (rs=0.68, p<0.001). There was a significant positive correlation between age and JND score for both HI and DI (rs=0.51 and 0.40 respectively, p<0.005). Other factors such as the duration of CI experience and the duration of deafness were not significantly correlated with pitch perception outcomes (p>0.05).

Bimodal users showed better HI and DI outcomes in bimodal listening than CI only condition [median HI JND=39.5 Hz (range 1.0 to 220.0 Hz) versus 6.5 Hz (range 3.0 to 133.0 Hz) and median DI JND=133.5 Hz (range 9.0 to 220.0 Hz) versus 38.5 Hz (range 7.0 to 220.0 Hz) in CI only and bimodal listening conditions respectively, N=14]. Differences were statistically significant for both HI (Z = -2.6, p≤0.001) and DI (Z = -2.9, p<0.005). DI outcomes in bimodal listening condition showed a significant correlation with unaided PTA thresholds for octave frequencies lower than 1000 Hz (rs=0.68, p=0.008). HI/DI outcomes at CI only in comparison to bimodal listening condition are given in Figure 11.

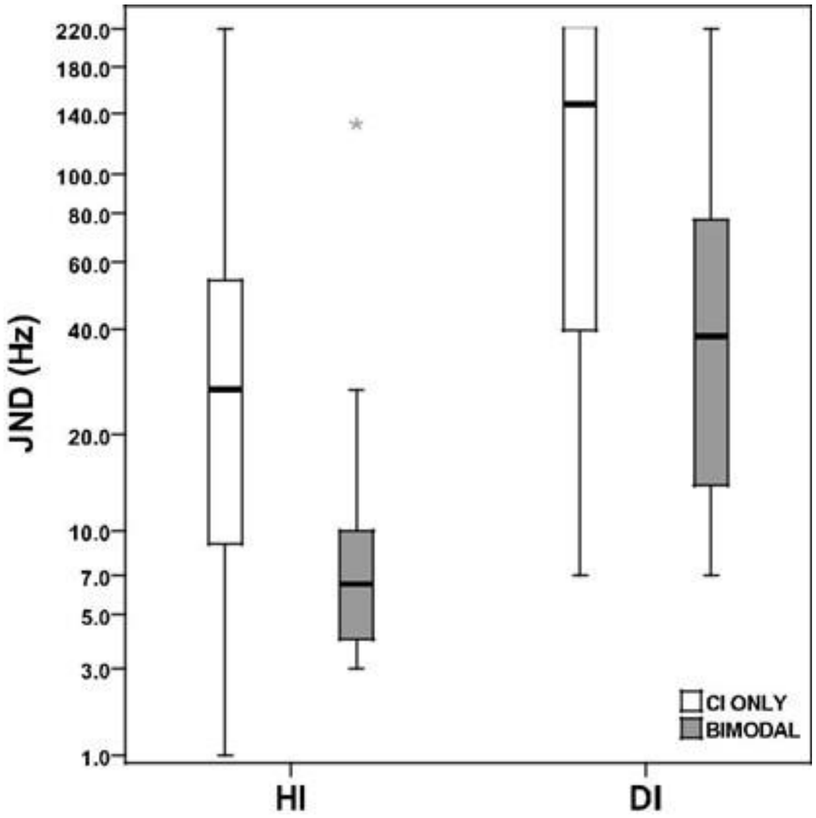


Figure 11. HI/DI outcomes at CI only in comparison with bimodal listening

5.3.2 Italian STARR outcomes in relation to other speech perception tests

All participants were able to complete the test. Median STARR outcome in CI only listeners was 14.8 dB SNR (range -1.8 to 125.0 dB, N=49). 67% of the group had a STARR score that was less than 20 dB SNR. In this subgroup, median STARR score was 8.6 dB SNR (range -1.8 to 19.5, N=33).

Analysis showed statistically significant negative correlations between STARR outcomes and other speech recognition measures (p<0.001). Table 1 illustrates outcomes for individual tests as well as Spearman's rho (rs) and p values regarding the correlations with STARR.

	Median	**Range**	**RS**	**p**
WRS in quiet	82%	22 to 100%	-0.65	0.000
SNR+10	35%	0 to 83%	-0.52	0.000
SNR+5	12%	0 to 75%	-0.61	0.000

TABLE 1. Outcomes for individual speech perception tests as well as Spearman's rho and p values relative to Italian STARR test.

The duration of deafness and CI thresholds had a significant effect on the STARR performance whereas other factors such as age and CI experience were not significantly correlated (p>0.05).

Bimodal users showed better speech recognition outcomes in bimodal listening than in CI only condition (median WRS in quiet = 76% vs 85, median SNR +10 = 35 vs 40%, median SNR +5 = 5 vs 22% and median STARR = 17.3 vs 8.1 dBSNR in CI only and bimodal listening condition respectively, N=14). Results are shown in Figure 12.

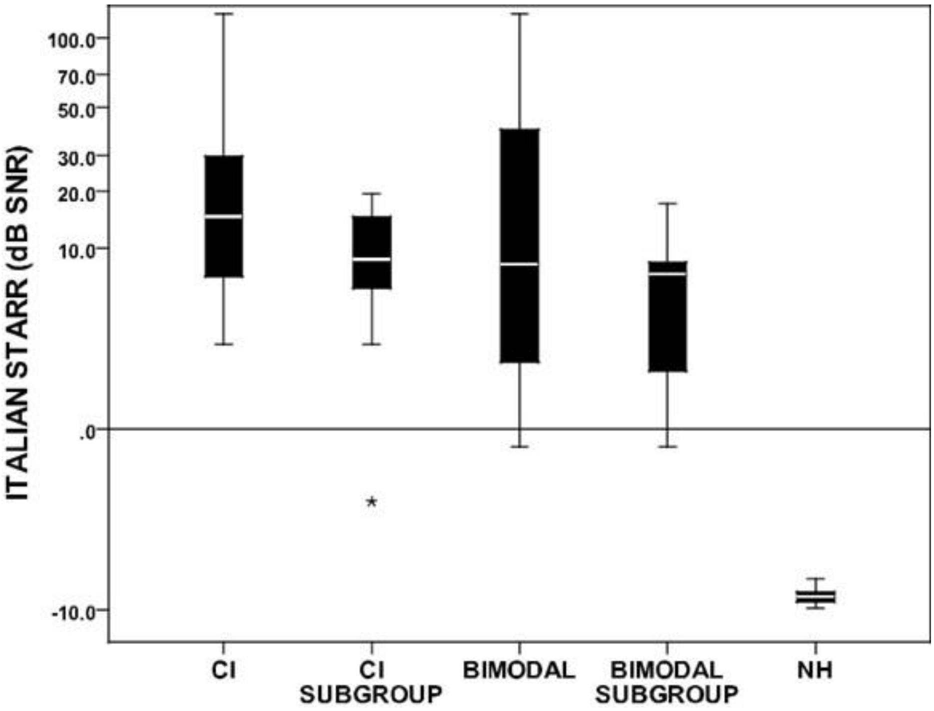


Figure 12 Italian STARR outcomes. Data for normal hearing (NH) people is received from Dincer et al (105)

Differences were statistically significant for speech perception in noise whilst they were not significant for speech understanding in quiet (Z= -2.9, p<0.005 in STARR). The details of analysis are given in Table 2. When listening with CI only, 9 out of 14 listeners (65%) had a STARR score that was less than 20 dB SNR instead of 10 (71%) when listening bimodally. 7 out of these 10 subjects had a meaningful amount of bimodal benefit [greater than 2.4 dB SNR improvement as indicated by Dincer D'Alessandro et. al. [105] when using two test lists per condition]. On the other hand, one subject had an improvement of 2.2 dB SNR whilst the other two had a deterioration of 1.5 and 3.0 dB SNR when listening bimodally. STARR outcomes did not show any significant correlations with unaided nor aided PTA thresholds for HA ear (p>0.05).

	CI only Listening		Bimodal Listening		Differences	
	Median	**Range**	**Median**	**Range**	**z**	**p**
HI (Hz)	39.5	1.0 to 220.0	6.5	3.0 to 133.0	**-2.6**	0.001
DI (Hz)	133.5	9.0 to 220.0	38.5	7.0 to 22.0.	**-2.9**	0.003
STARR (dB SNR)	17.3	-1.8 to 125.0	8.1	-0.3 to 125.0	**-2.9**	0.004
WRS (%)	76	30 to 100	85	50 to 100	-1.5	0.130
SNR+10 (%)	35	0 to 60	40	0 to 98	**-2.0**	0.050
SNR+5 (%)	5	0 to 40	22	0 to 75	**-2.0**	0.050

Table 2. Within-subject comparisons for CI vs bimodal listening conditions. Statistically significant differences in bold

5.3.3 HI/DI in relation to STARR and other speech perception tests

For all participants including CI plus HA listening condition in bimodal users (N=63), analysis of correlations between HI/DI outcomes and Italian STARR showed statistically significant positive correlations (rs=0.40, p<0.005 for HI and rs=0.30, p<0.05 for DI). Figure 13 shows the correlations between HI/DI and Italian STARR outcomes.

On the other hand, HI was significantly correlated with WRS in quiet (rs= -0.46, p<0.001), with WRS+10 (rs= -0.38, p<0.005) whilst DI was not significantly correlated with other speech perception tests. The details are given in Table 3.

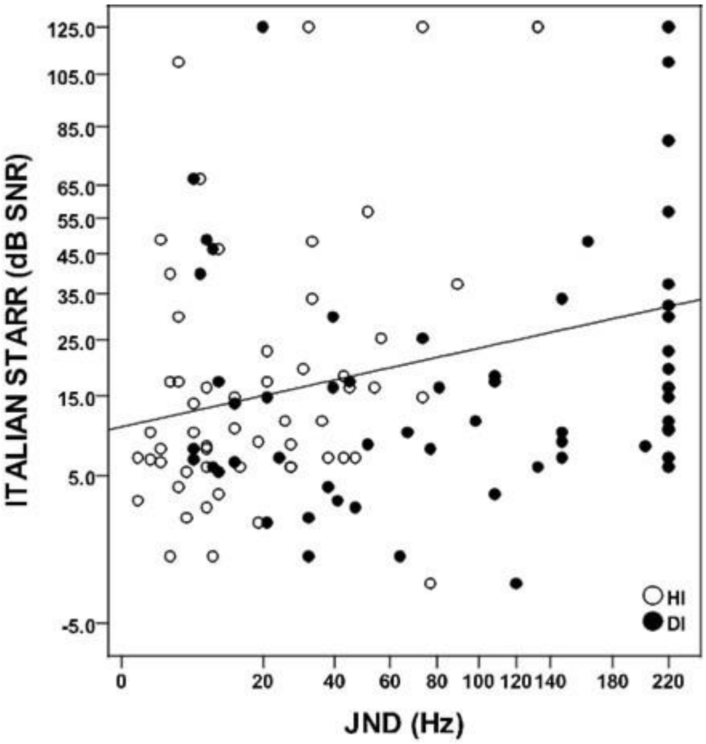


Figure 13. Correlation between HI/DI and Italian STARR outcomes

	HI		DI	
	R_S	p	R_S	p
ITALIAN STARR	**0.40**	**0.002**	**0.30**	**0.036**
WRS	**-0.46**	**0.000**	-0.22	0.077
SNR+10	**-0.38**	**0.003**	-0.18	0.186
SNR+5	-0.26	0.067	-0.16	0.239

Table 3. HI/DI outcomes in relation to Italian STARR and other speech perception tests. Statistically significant correlations are in bold

5.4 Discussion

One of the primary objectives of the present study was to investigate TFS processing that is assumed to be reflected in the LF pitch perception capacities for a larger CI recipient group. For this purpose, A§E Harmonic Intonation and Disharmonic Intonation tests were used. Findings were similar to outcomes studied previously on a smaller group of adult CI recipients [76] where median JNDs were found to be 16.0 and 139.0 Hz for HI and DI respectively. The corresponding values were 27.0 Hz and 147.0 Hz in the present study. Although CI recipients had significantly better scores for HI than for DI, the majority showed abnormal outcomes for both HI and DI tests. Significant correlations between HI/DI outcomes confirmed that both tests targeted LF processing capacities. However, relative performance deterioration for the DI test was re-indicating DI with more differential outcomes on phase locking and TFS processing capacities whereas the HI test, due to the sweep of F0 together with its harmonics, could provide some HF cues in the complex signal and therefore, additional place cues for lower JNDs [101,104]. Furthermore, there was a strong effect of age on both HI/DI outcomes indicating that TFS sensitivity worsened with increasing age. Indeed, it is well-known that aging is accompanied by changes in physiological, psychophysical and psychological domains. As a result of increase in longevity, the effects of aging on the auditory system have received much attention recently [107,108]. Previous studies on TFS sensitivity were carried out in NH populations and results indicated a significant effect of aging both for binaural, and monaural listening tasks like in our study. However, more studies are required to give a precise definition of the role of changes in peripheral sensory processing or in the central auditory system or in cognition and attention associated with aging [100,109].

Another objective was to evaluate the Italian STARR outcomes in adult CI recipients as well as to study the correlations with other speech perception tests that are commonly used in Italian. The STARR test originally existed in British-English and was based on both male and female speakers instead of a male speaker only as in the Italian version. Present findings indicated that median Italian STARR score (14.8 dB SNR) was lower than the British version (22dB, 34 dB and 28 dB SNR for the male, female speakers and their overall mean respectively) studied by Boyle et al (51). However this difference certainly was due to the difference in the measure of central tendency (in the present study, median value had to be used since the data was not normally distributed).

On the other hand, Boyle et al. [51] reported that for SRTs lower than 20 dB SNR, performance during the adaptive track was related in a more orderly way to the SNR whereas higher SNRs did not materially change the performance and the score was determined basically by the ability to understand speech in quiet rather than by the SNR. Therefore, a subgroup of participants with SRTs better than 20 dB was further analysed to investigate performances that were obtained in a meaningful way from the test. Although all participants were able to complete the test, 67% of the study group had an SRT that was considered meaningful. However, it should be considered that 37% of the overall group had a duration of profound deafness longer than 10yrs and its significant effect on STARR performance was among the remarkable findings of the present study. effect on STARR performance was among the remarkable findings of the present study. In the subgroup, the median STARR score was 8.6 dB SNR. In the study of Boyle et al. [51] for British-English population, 88% of the CI users achieved to complete the test whilst the subgroup consisted of only 40% of the overall group. The mean SRT for the male speaker was a bit lower (5.9 dB) in their case whilst the overall mean for both speakers (9.4 dB) was very similar to Italian STARR outcomes. Such differences in multilingual speech perception tests are to be expected and can be attributed to language as well as speaker dependent factors [91].

On the other hand, both the overall success rate and the percentage of subgroup were smaller in the Boyle et al. study [51]. But their group concerned CI recipients who performed more poorly than typical and the duration of profound deafness before implantation was even longer. Moreover, their subjects consisted of CI users with older-generation technology and their processors were not optimally adjusted. On the contrary, in another similar study, Haumann et al. [79] tested 55 German-speaking CI users using roving levels and all their subjects were able to achieve meaningful SRTs. Nevertheless, all these studies including the present one showed abnormal SRTs in CI recipients confirming their difficulty for listening in the presence of noise. Even the better performers (with SRTs lower than 20dB) showed a vast difference in comparison to NH people (17 dB poorer than NH group in italian STARR). Furthermore, the significant effect of hearing thresholds on Italian STARR performance, even in a group that achieved typical target CI thresholds ≤ 40 dB HL, was re-emphasizing the sensitivity of the test to lower level speech which a CI user can face very often during everyday life.

A further objective of this study was to investigate outcomes in contralateral HA users. The previous results by Vaerenberg et al. [104] and Dincer D’Alessandro et al. [103] indicated a bimodal benefit only in DI scores whereas HI scores tended not to change between CI only and bimodal listening conditions. However, present findings showed considerable bimodal benefit for both HI and DI outcomes which could be explained by the small number of bimodal listeners in all these studies; nevertheless, the present number is bigger and it is reasonable to expect bimodal benefit for the kind of tasks in both tests. On the other hand, bimodal benefit was remarkable in Italian STARR performance as well; bimodal users showed better STARR outcomes in bimodal listening than CI only condition [median=17.3 versus 8.1 dB SNR in CI only and bimodal listening conditions respectively]. Differences were statistically significant. When listening with CI only, 9 out of 14 listeners (65%) had a STARR score that was less than 20 dB SNR instead of 10 (71%) when listening bimodally. 7 out of these 10 subjects had a meaningful amount of bimodal benefit (greater than 2.4 dB SNR improvement as indicated by Dincer D'Alessandro et al. [105] when using two test lists per condition). Similarly, bimodal benefit had a significant effect on word recognition in noise whereas scores for listening in quiet did not show statistically significant differences. However, overall outcomes in bimodal listeners were still worse than those of NH counterparts. Although speech perception outcomes including those for STARR did not show any significant correlations with unaided or aided PTA thresholds for the HA ear, DI outcomes in bimodal listening condition showed a strong effect of unaided PTA thresholds for octave frequencies lower than 1000 Hz suggesting that bimodal listeners do benefit from the additional LF cues and by remaining phase-locking capacities of residual hearing and in that way the representation of TFS for CI users was improved [79].

Finally, this study aimed to analyze the outcome comparisons between pitch and speech perception. The idea was to get a better understanding of the link between LF pitch perception and the processing of TFS in CI users as well as implications for speech perception performance in particular when using a test that attempts to better represent everyday listening situations. It was previously shown that speech recognition scores in quiet were not significantly correlated with HI/DI outcomes in paediatric population [103]. Furthermore, as mentioned before, outcomes in EAS processor users revealed significant differences between electric only and EAS listening modes in speech in noise results as well as a remarkable improvement in DI performance [104]. Although the strength was moderate, present findings showed that HI was significantly correlated with WRS in quiet, with SNR+10 and with Italian STARR outcomes, whilst DI was significantly correlated only with Italian STARR scores indicating again DI to provide more differential outcomes on phase locking and TFS processing capacities since TFS cues are emphasized in difficult listening tasks such as listening in the dips which is defined as detecting a signal in a fluctuating background [81].

The small sample size for bimodal and bilateral listeners was an important limitation of the present study. Bilateral CI users were tested one ear at a time (mono-lateral listening only) in order to avoid longer testing during the same day but especially due to the limited number of test lists that were available for Italian STARR. Actually, it would be interesting to study outcome comparisons in relation to bilateral benefit as well. Moreover, the group size did not allow us to study any possible impact of the device type, in particular the sound coding strategy, on the performance. Furthermore, it would be useful to study clinical usefulness of HI/DI tests as a predictor of music perception and appreciation in CI users since the availability of LF pitch and TFS cues are even more dominant for music perception.(110)

5.5 Conclusions

CI recipients usually complain about their difficulties for speech understanding in noisy environments. One important aspect that contributes to this fact is recognized as poor TFS processing of cochlear implants. This thesis aimed to investigate the following three topics. First, the LF pitch perception skills that are believed to be linked to TFS processing, were studied in a group of pediatric and adult CI recipients. Secondly, the STARR test which attempts to mimic challenging real life listening conditions was adapted in order to introduce the test into Italian. Finally, the interactions between TFS processing, pitch and speech perception outcomes as well as the effects of demographic factors and the amount of bimodal benefit were investigated.

5.5.1 Findings suggest the following:

1) HI/DI outcomes evaluating LF pitch perception were found to be abnormal in the majority of both paediatric and adult CI recipients, confirming poor TFS processing capacities of cochlear implants. However, DI performance was considerably better in children than in adult CI recipients.
2) HI/DI outcomes showed a significant positive correlation: subjects with higher JNDs on HI tended to have higher JNDs on DI test as well. However, performance was significantly worse for the DI test. This finding indicated the DI test as providing more differential LF pitch perception outcomes in that it reflected phase locking and TFS processing capacities of the ear, whereas HI test provided information of its place coding capacity as well.
3) HI/DI tests were clinically applicable in children of 5 years and older. Chronological age had a significant effect on DI performance. NH children under the age of 8.5 years showed larger inter-subject-variability; however, the majority of them showed outcomes that were considered normal at adult-level.

4) Similarly, age in adult CI recipients, showed a strong effect on both HI/DI outcomes indicating that TFS sensitivity worsened with increasing age.

5) Contralateral HA users had remarkable bimodal benefit on both HI/DI tests. Moreover, DI outcomes in bimodal listening condition showed a strong effect of unaided PTA thresholds for octave frequencies lower than 1000 Hz suggesting bimodal listeners to benefit from the additional LF cues and by remaining phase-locking capacities of residual hearing. However, the outcomes in bimodal listeners were still worse than those of NH counterparts

6) Findings in NH listeners suggested that the Italian STARR test could be a promising supplement to existing speech assessment tools. The average SRT for NH and CI recipient subjects was consistent with SRTs reported for sentence testing by other researchers. The variability of mean SRTs across test lists was relatively small. Statistical analysis showed no significant learning effects. The outcomes for Italian STARR test showed statistically significant correlations with those for standard word recognition test in Italian.

7) STARR outcomes showed abnormal SRTs in CI recipients confirming their difficulty of listening in the presence of noise. Even the better performers had a vast difference in comparison to NH people.

8) The success rate for Italian STARR test was excellent; all subjects managed to complete the test. But only 67% of the study group had an SRT that was considered meaningful.

9) The duration of profound deafness had a significant effect on STARR performance

10) The significant effect of CI thresholds on Italian STARR performance re-emphasized the sensitivity of the test to lower level speech which a CI user can face very often during everyday life.

11) Similarly, with LF pitch perception outcomes, bimodal users had a significant benefit on speech perception in noise whereas scores for listening in quiet did not show statistically significant differences. However, outcomes in bimodal listeners were still worse than those of NH counterparts.

12) HI was significantly correlated with WRS in quiet, with SNR+10 and Italian STARR outcomes whilst DI was significantly correlated only with Italian STARR scores indicating again DI to provide more differential outcomes on phase locking and TFS processing

capacities since TFS cues are emphasized in difficult listening tasks such as listening in the dips which is defined as detecting a signal in a fluctuating background.

*Material covered in Chapter 5 is awaiting publication in *Ear and Hearing*: Dincer D'Alessandro H, Ballantyne D, *et. al.* (2017): Temporal Fine Structure Processing, Pitch and Speech Perception in Cochlear Implant Recipients. *Ear and Hearing* (Submitted February 2017, accepted for revision May 2017, accepted for publication October 2017)

Chapter 6 MUSIC PERCEPTION AND APPRECIATION IN CI RECIPIENTS

6.1 Introduction

Nowadays, most people with modern cochlear implant systems can understand speech using the device alone and – depending on certain variables (age, neuron degeneration, previous device use etc) - even in the presence of noise. In the past years however, increasing research has concentrated on the users' perception of non-speech sounds, especially music. (111)

There are two main dimensions of music perception that have been analysed most in CI users, Discrimination and Music appreciation. Discrimination abilities are measured by a variety of psychophysical or electrophysiological approaches (112,113,114). Music appreciation depicts a complex process a complex process of subjective estimation of the perceived quality of the music. It is usually analysed using questionnaires and rating systems.

Much of the published research on how CI users perceive music is based on the assumption that music can be characterized as an organized sequence of sounds that have a number of fundamental features, including Rhythm, Melody and Timbre. Additional attributes of sounds, such as Harmony and overall Loudness, also contribute to the structure of Music. Beyond these objective characteristics of sounds, there are diverse phenomena that are important in the experience of listening to music. These include subjective quality, mood and situational context.

The most significant findings in this field have established the following aspects: (1) on average, implant users perceive rhythm as well as NH subjects; (2) recognition of melodies especially without rhythmic or verbal cues, is poor; (3) perception of timbre, as seen in identification of musical instrument sounds, is generally unsatisfactory (115); (4) Implant users tend to rate the quality of musical sounds as less pleasant than NH listeners; (5) Pitch perception might be improved by improved designs of speech processors that use both temporal and spatial patterns of electric stimulation more effectively; (6) for those CI users who have usable residual contralateral acoustic hearing, at least for low-frequency sounds, perception of music is likely to be better with combined acoustic and electric stimulation (bimodal).

6.1.1 Discrimination Abilities

6.1.1.a Perception of Rhythm:

Temporal patterns in musical sounds that impart a distinctive rhythm generally occur at the approximate frequency range of 0.2 to 20 Hz. Musical ‘dynamics’ are usually associated with variations in loudness below this range, whereas higher frequency components of acoustic signals carry Pitch information which can be fundamental in CI users. Since perception of musical rhythm has been found to be similar to that of NH subjects it can be assumed that this function is related to the perception of the duration of sounds and the gaps between sounds. To perceive rhythm pattern adequately in most types of music, temporal resolution required for either duration or gaps is in the order of tens of milliseconds (ms) (116).

6.1.1.b Perception of Melody: What enables people to recognise melodies?

First, there is a question of which tunes are sufficiently familiar to a listener such that he/she would be able to name them on hearing them which depends on a range of highly variable factors, including the individual’s musical training and listening experience, the social culture within which this experience has been gained and the person’s memory of the tunes. This has also been highlighted in a recent paper by Bruns et al (119). This corroborates a previous study where Moore and Shannon claim that since hearing experience prior to implantation can have a major impact on speech comprehension post CI, then the ability to process and enjoy music may also be influenced as well (117). Electric signal transfer by the CI is generally limited in fine-temporal and fine-spectral as well as dynamic range resolution compared to NH (118). Besides the technical constraints, anatomical changes due to auditory deprivation and hearing experience pre-implantation will lead to different individual hearing conditions which it is difficult to compare in this study.

Second is the question of Melodic pattern recognition. The task of discriminating between different pitch contours is related to melody identification, but is generally more difficult because of the reduced number of auditory cues available in the test material. A typical melodic pattern recognition test is based on the same principles mentioned in the pitch tests (DI/HI) used in the previous study (see Chapter 5). That is, listeners are asked to label two pitch sequences as the same or different. As rhythm is identical and no coincidental verbal cues are present, discrimination relies on the listener’s ability to perceive a pattern of changes in pitch. In this context, neither the absolute nor relative pitch of each note needs to be recognized, hence the detection of an overall itch contour, such as the perception of a generally rising or falling pitch may be sufficient for a listener to discriminate sequences.

In conclusion, users of implant systems typically have great difficult recognizing melodies, even when the tunes are familiar. In fact, if distinctive rhythm patterns are noticeable in the tunes it is this that seem to provide most of the information that CI recipients use when they identify melodies. Hence, we can conclude that one of the most serious problems that confronts CI users is that pitch information is still conveyed very poorly.

6.1.1.c Perception of Timbre

Timbre can be described as 'tone colour', that is the quality that characterizes differences in tone that are apparent when musical tones are played with the same pitch and loudness on several different instruments. The principal properties of the acoustic signals are the frequency spectrum and the amplitude envelopes of sounds, including changes in these attributes over time. It has been shown that perception of Timbre has generally been reported as much poorer for implant users than for NH subjects. Since CI wearers only use auditory cues, they cannot readily identify the instrument itself but can sometimes discriminate between two instruments when the differences in the temporal envelope of the sounds are obvious owing to harmonics, eg distinguishing between the sound of a flute and a drum. This suggests that information concerning spectral shape is represented only crudely in the electric stimuli. In fact, most studies on the perception of Timbre in CI users seem to have focused on the ability of listeners to identify or discriminate the sounds of different musical instruments, even though in Filipo et al- (115), failure to do this task was attributed to the lack of musical education in most of the participants and not to technical characteristics of Implants.

6.1.1.d Appraisal and Appreciation of Music

In order to get an idea of Appraisal and Appreciation of Music in CI recipients, apart from the strictly acoustic and psychoacoustic aspects of the stimuli, researchers usually make use of appraisal ratings that indicate the subjective pleasantness of sounds. The tendency would seem to suggest that appreciation is necessarily lower for implant users than with normally hearing listeners. These methods are based on rating scales going from bad (0) to very good (10) and take into account various aspects of the listening experience. When data are correlated with the numerous variables present in this sort of study, maybe researchers will be able to offer solutions to CI users.

6.1.1.e Understanding of Music with Cochlear Implants

Processing of musical meaning is a third dimension of music perception which, to date has received much less attention. Recently, in 2016, Bruns et al, (117) carried out an interesting study on the understanding of music. This was carried out via the examination of semantic concepts as elicited by entire musical pieces with complex sound properties (as opposed to past ERP studies which focused mainly on discrimination processes) , being measured by event-related potentials (ERP), ie an objective, multi-dimensional online assessment tool. In fact, the semantic processing of a lot of different stimuli can be studied by investigating a late component in the ERP, namely the N400 (120).

Outcomes implied that despite the degraded hearing impression with CI, cortical plasticity enables post CI users to access musical semantic concepts, which were built up during time of NH. By contrast, the influence of prelingual hearing impairment most likely distorts this initial concept formation.

Finally, even though electric signal transfer by the CI is generally limited in fine-temporal and fine-spectral as well as dynamic range resolution (119) compared to the NH , more recent cochlear implant technology such as current steering, virtual channels, increased spectral resolution and Temporal Fine Structure should enable the cochlear implant user to improve perception and appreciation of Music. This however depends on numerous subjective variables such as age, deprivation, device use etc.

6.2 Materials and Method

6.2.1 Study Group

The final study group consisted of 22 adult subjects with a mean age of 65.6 years (range 46 to 92 yrs). Average Hearing deprivation was 20.8 months.

In this group, 8 subjects were using Advanced Bionics (AB) implants and 14 MedEl devices; 13 were in monolateral mode, whereas 9 were in bimodal use (CI + contralateral Hearing Aid). Average time of CI use was 55.9 months.

Test Room and Instrumentation: Testing was carried out in a silent room 2 x 2metres, where the stimulus was presented via a PC and a preamplifier connected directly to a single loudspeaker at 0° azimuth.

All subjects underwent 3 separate testing sessions: Pitch tests (HI/DI); Speech in noise tests; Music Trials

6.2.1 Pitch Protocol

Pitch is the subjective sensation of the frequency of sound which goes on a scale from low to high. In the instance of complex sounds it is Pitch that contributes both to the perception of Language (Intonation, Accent) and Music (Melody). In deaf adults the ability to appreciate this aspect of sound is impaired. It is fundamental to note that Pitch refers mainly to Low Frequency (temporal) contents of sound and depends mainly on acoustic nerve processing.

The ability to perceive changes in Pitch are assessed by two tests in a protocol known as Auditory Speech Sound Evaluation, ie ASSE or A§E (121).: Harmonic Intonation (HI) and Disharmonic Intonation (DI). In these tests, the stimulus is a complex harmonic tone with fundamental frequency (F0) and 3 upper harmonics (2F0, 3F0, 4F0). The intensity of each harmonic is 6dB under the primary component. In order to offer a greater sensation of naturalness, white noise (SNR+10,9) was added

For both tests 2 stimuli are presented: in DI without intonation; in HI with intonation where the sound used is a sweep which makes use of all the harmonics (including F0) from NF0 to N(F0+deltaF) with N variable from 1 to 4. In DI, sounds are presented in a sweep of F0 (from F0 to F0+deltaF), whereas the upper harmonics are fixed at their initial frequency. This induces a sensation of Disharmony or Dissonance.

Trials were carried out in the silent room and all stimuli were presented in Free Field at an intensity of 70dB. Subjects are familiarized with the procedure, prior to formal testing. Basically, they are required to indicate whether the 2 sounds are similar or not. It is the software which calculates the Just Noticeable Difference (JND) for Pitch Discrimination, using an adaptive procedure which assesses the point in which the patient's psychometric function reaches 50%.

6.2.2 Speech Noise Test Protocol

Materials used in this research included 2 validated Speech Noise Protocols: STARR and OLSA.

STARR or Sentence Test with Adaptive Randomized Roving

Sentences were presented in randomized order at 3 different levels (50, 65 and 80 dB) to a background of speech noise where the SNR varied on the grounds of the patients' responses. The full procedure which is fairly complex is an attempt to mimic real life listening conditions and is described in full detail in Chapter 4. Scoring is carried out automatically by software and is based on SNR dB for correct number of correct words in each sentence.

OLSA or Matrix Sentence Test in Noise for the Italian Language

The speech material used in this test consists of syntactically fixed but semantically unpredictable sentences with a total number of 50 words from which test sentences are randomly combined. Words were selected regarding the following criteria: Word frequency, Semantic neutrality at word and sentence level, Grammatical correctness of all possible sentence combinations, Balanced number of syllables within word groups (words of 2-3 syllables), Phoneme distribution representing the Italian language. A masking noise with the same frequency spectrum as the speech material was generated by superimposing 300 resynthesized sentences containing the whole range of speech material (122).

In this protocol testing was carried out under Open set conditions, and it is the operator who decided on correct responses. Noise level was fixed at 65dB and sentences were presented at randomized SNRs on the grounds of the patients' responses. Scores were expressed in SNR and Intelligibility Slope functions.

6.2.4 Appreciation of Music – AB questionnaire

Since the perception of Music is an exquisitely subjective aspect of listening, this protocol is based on 3 CDs and a questionnaire – introduced by Advanced Bionics for use in clinical practice. The CDs contain 3 tracks of Music and 2 of Speech. In order to assess the patient's ability to analyse the stimulus with the CI, he/she is presented with 3 different types of music of varying difficulty:

- *Classical Music* (Symphonic=complex) 'Dynamic';
- *Jazz* (3 instruments-Trumpet, Wire brush on drum, double Bass - Medium) 'Mellow';
- *Soul* (single female voice – Simple) 'Sweet'

Each of the 3 tracks lasted 1 minute, and were presented in randomized order to the patient. After each single track the patients were required to fill in a questionnaire in order to express their appreciation of various aspects of the specific track to which they had just listened. The questionnaire was based on 6 subjective aspects relative to the Appreciation of Music: Clarity, Pleasantness, Naturalness, Overall Quality, Booming, Metallic. This is assessed on an analogical scale from 0 to 10 where worst conditions are expressed by low numbers and best conditions by high numbers.

This was followed by simple questions such as Instrumental or Vocal, if vocal Male or Female, Loudness and Rhythm. These were not included in statistical analysis.

6.3 Results

6.3.1. Descriptive Analysis

For descriptive analysis, Data was divided into 5 basic modes: Daily Mode (all 22 subjects), CI monolateral (13 Subjects), Bimodal (9 subjects), Bimodal CI only, Bimodal HA only. Testing was carried out regularly for each mode.

Box Plots were drawn up for each of the aspects taken into consideration (Pitch, Speech/Noise Tests and Music) and median values are given..

6.3.1 Pitch

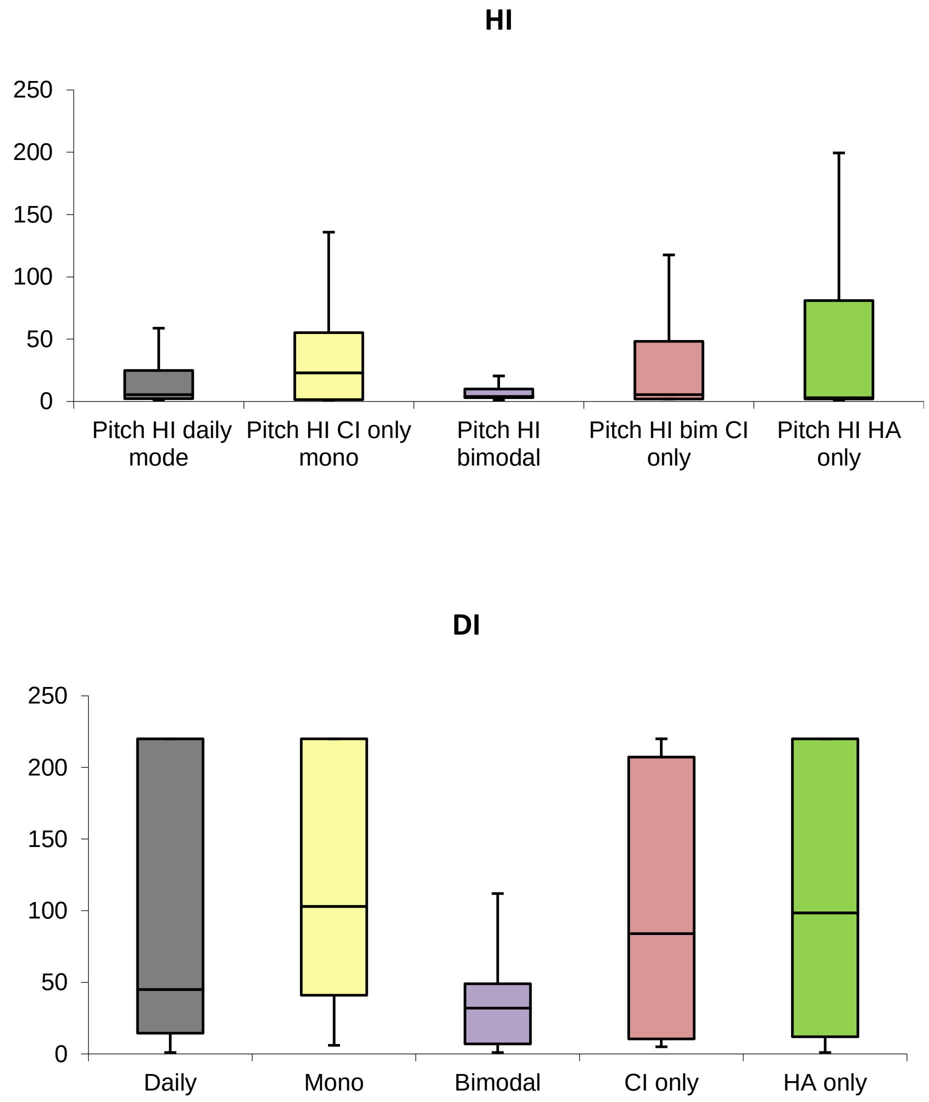
. *Pitch: Harmonic Intonation* (HI) vs *Disharmonic Intonation* (DI).

- Performance was significantly worse for the DI test (HI=3-23 vs DI=32-103). This can be attributed to level of difficulty. This finding is consistent with a previous study (see chapter 5) which indicated how the DI test provided more differential LF pitch perception outcomes in that it reflected phase locking and TFS processing capacities of the ear, whereas HI test provided information of its place coding capacity as well

- Contralateral HA users showed remarkable bimodal benefit on both HI/DI tests. Moreover, DI outcomes in bimodal listening condition showed a strong effect of unaided PTA thresholds for octave frequencies lower than 1000 Hz suggesting bimodal listeners to benefit from the additional LF cues and by remaining phase-locking capacities of residual hearing. However, the outcomes in bimodal listeners were still worse than those of NH counterparts

Figure 14 Box Plot

Pitch



Median Values

HI	**5,5**	**23**	**4**	**5,5**	**3**
DI	**45**	**103**	**32**	**84**	**98,5**

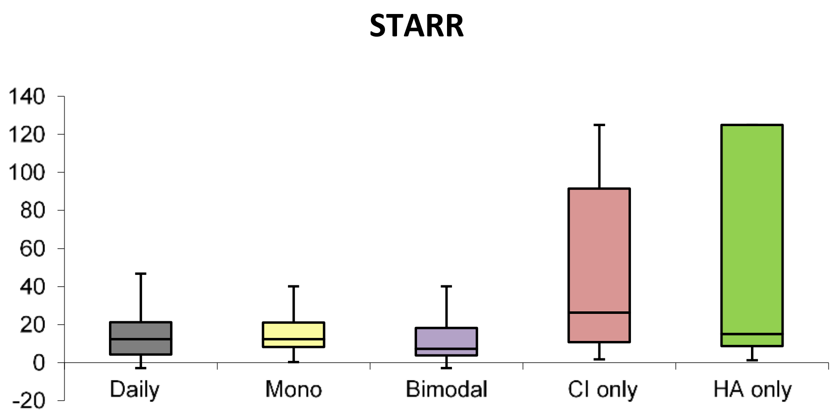
6.3.2 Speech in Noise

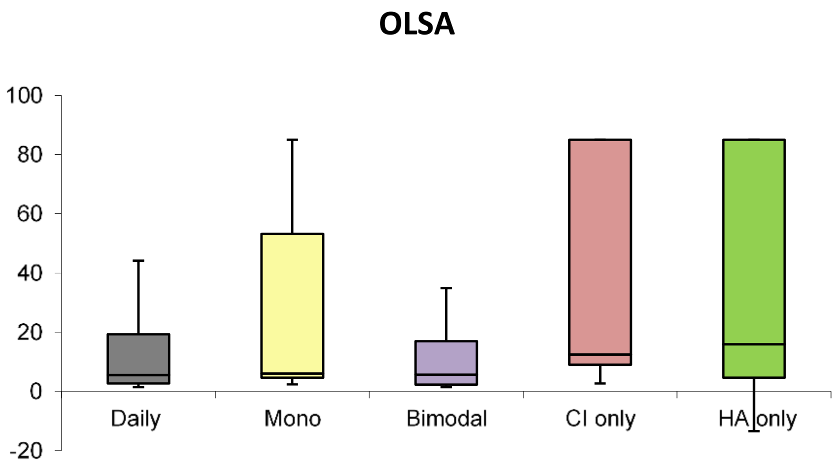
Two different procedures were used for clinical assessment of the ability of subjects to process speech in noise under real life listening conditions: STARR and Italian Matrix test (OLSA). Median values for the 5 groups taken into consideration are expressed in SNR dB where the lower the score the better is the result.

- Findings for STARR ranged from 7.3 to 26 SNR dB which is within the norm for CI users. All subjects managed to complete the protocol despite the fact that this is the more difficult of the procedures and reflects the continuously changing listening conditions encountered by CI users every day. Furthermore, the significant effect of CI thresholds on performance re-emphasized the sensitivity of the test to lower level speech.
- For OLSA median values ranged from 5.5 to 15.9, here again within the norm. This is a very useful clinical tool in that it less laborious and, even though it is adaptive, it cannot be considered a true roving procedure. Here there was much more dispersion of data.

Figure 15 Box Plot

Segnale/Rumore





Median Values					
STARR	**12,3**	**12,3**	**7,3**	**26,3**	**15**
OLSA	**5,5**	**6,1**	**5,6**	**12,5**	**15,9**

6.3.3 Music

Each of the single criteria in the questionnaire (6) has been examined for the different types of Music

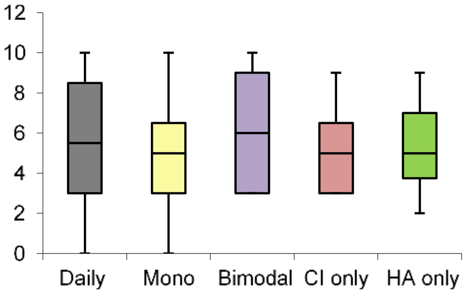
Median Values for the various aspects of Appreciation of Music brought to light the following trends:

- Average values for all 6 criteria ranged from a minimum of 5.8 to a maximum of 7.2,
- When this was broken down into the various types of music, ranges of values for single criteria were as follows: Classical 4-8; Jazz 5 – 9; and Soul 4-8
- ***Classical*** music would seem to be the one subjects appreciated less on the whole (mean 5.8-6.2). This could be attributed to the fact that the music they are asked to listen to is Symphonic, hence very complex in that being a complete orchestra there are numerous instruments each of which plays a different score with different resonance attributes;
- ***Soul*** (single female voice) follows (mean 6.2-7.1), with satisfactory findings (~7) both in Daily mode and Monolateral CI. Values for Bimodal when wearing both devices and with the single CI and HA only, decreased to 6.3 and 6.2 respectively. Here, this could perhaps be attributed to the frequencies of the track which were fairly acute, since it is a single female voice. This could influence appreciation when wearing the hearing aid (acoustic stimulation) owing to phenomena such as recruitment, distortion, lack of sufficient amplification in high frequency range;
- ***Jazz*** was the most appreciated of all 3 tracks (mean 6.7-7.2) even reaching a score of 9 with CI only both in monolateral and bimodal only modes for lack of boominess. In part it could depend on the type of music itself which is considered to be ‘mellow’, with a basic, slow rhythm pattern easily detected by implant wearers. They have more trouble with melody.

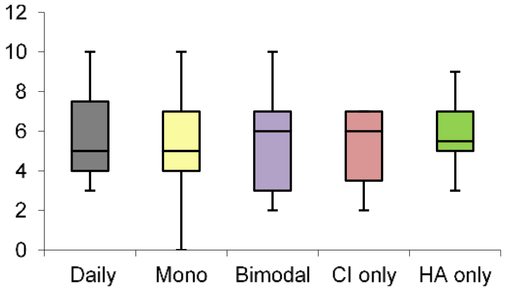
Figure 16 Box Plot

Classical

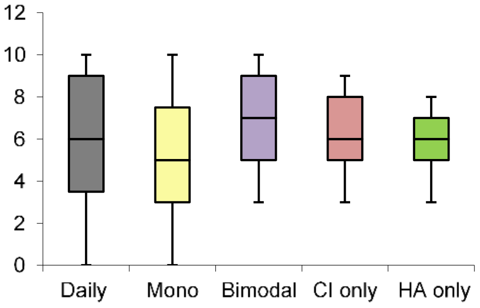
Clarity



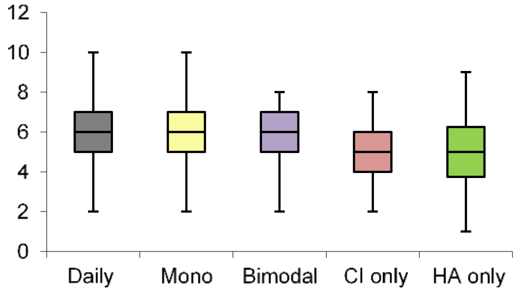
Pleasantness



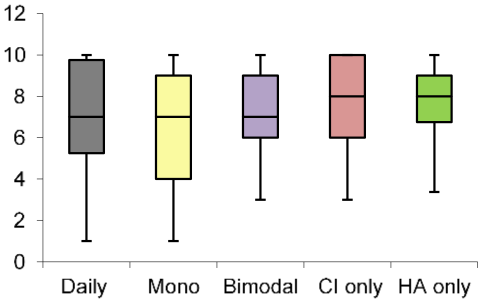
Naturalness



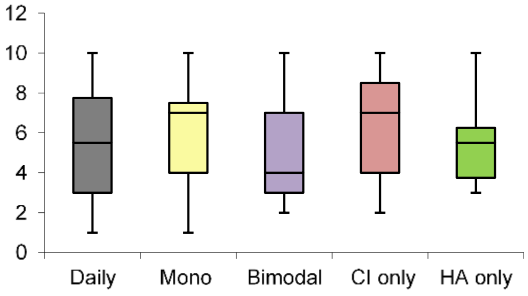
Overall



Booming



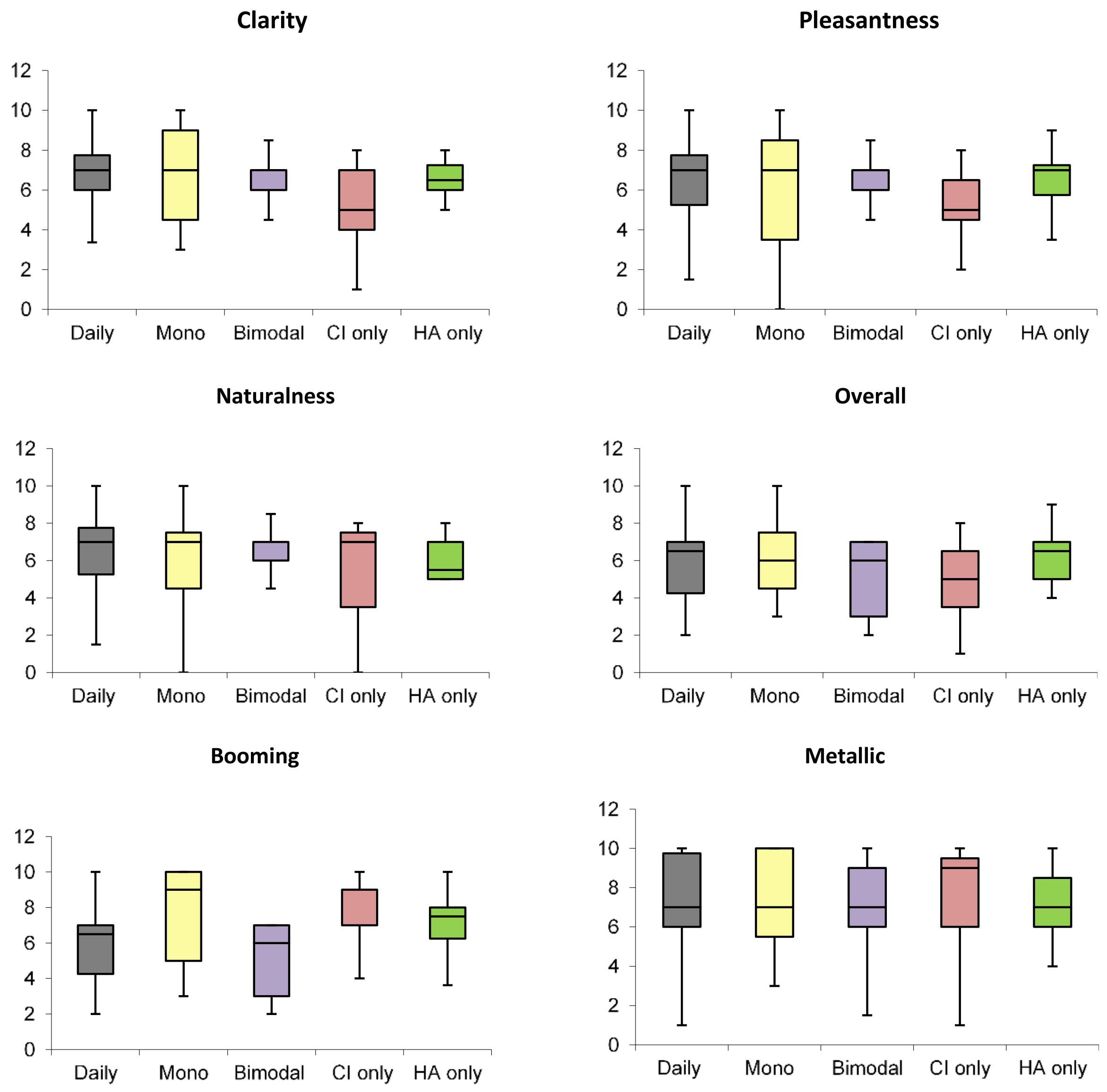
Metallic



Classical	Daily Mode	Mono	Bimodale	CI only	HA only
Clarity	5,5	5	6	5	5
Pleasantness	5	5	6	6	5,5
Naturalness	6	5	7	6	6
Overall	6	6	6	5	5
Booming	5,5	7	4	7	5,5
Metallic	7	7	7	8	8
Media	5.8	5.8	6	6.2	5.8

Figure 17 Box Plot

Jazz



Clarity

Pleasantness

Naturalness

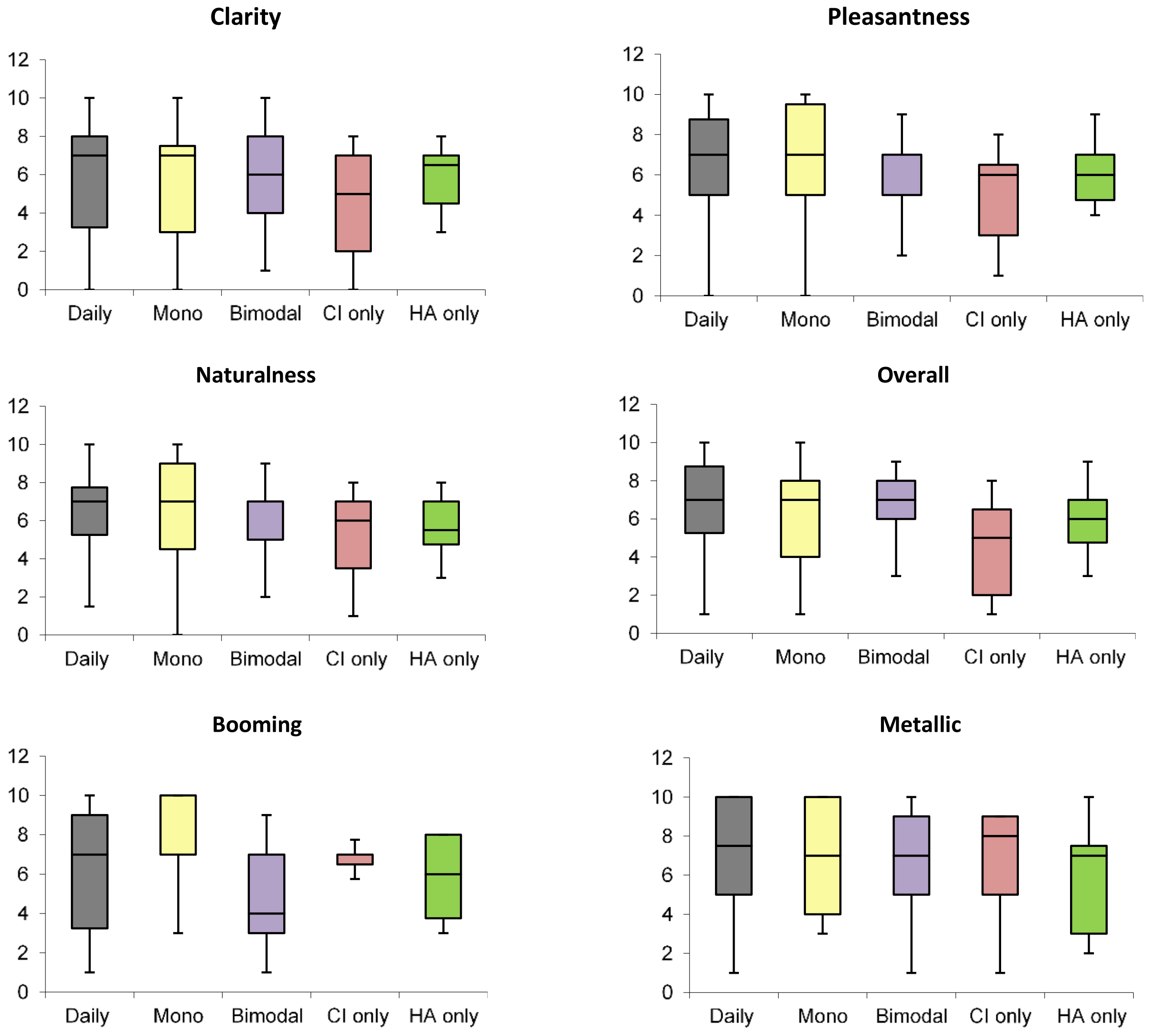
Overall

Booming

Metallic

Jazz	Daily Mode	Mono	Bimodal	CI only	HA only
Clarity	7	7	7	5	6,5
Pleasantness	7	7	7	5	7
Naturalness	7	7	7	7	5,5
Overall Qualit	6,5	6	6	5	6,5
Booming	6,5	9	6	9	7,5
Metallic	7	7	7	9	7
Media	6.8	7.2	6.7	6.7	6.7

Figure 18 Box Plot **Soul**

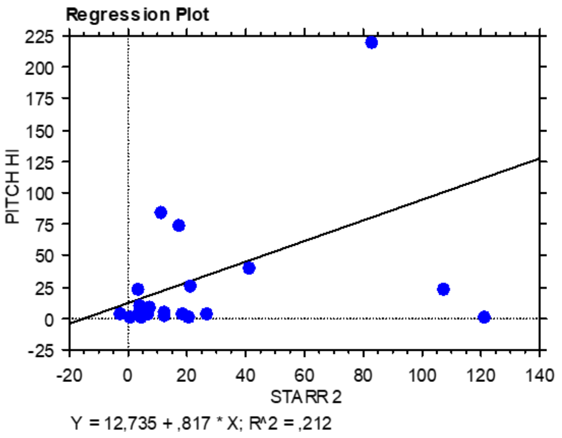


Soul	Daily Mode	Mono	Bimodale	CI only	HA only
Clarity	7	7	6	5	6,5
Pleasantness	7	7	7	6	6
Naturalness	7	7	7	6	5,5
Overall	7	7	7	5	6
Booming	7	7	4	7	6
Metallic	7,5	7	7	8	7
Media	7.1	7	6.3	6.2	6.2

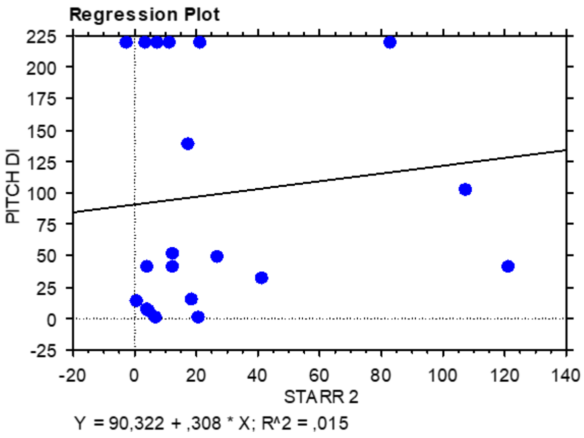
6.4 Results of Regression Analysis

As mentioned previously, the Study group consists of 22 subjects of whom 13 were Monolateral users and 8 bimodal (CI + contralateral HA). Before proceeding with Regression analysis, Student T tests were performed on data, under various conditions (Pitch, Speech/Noise, Music) in order to verify the significance of the difference between findings for the 2 groups. Since findings were negative (ie not significant), the Author proceeded with Regression analysis on the Total Study Group (in Daily Mode).

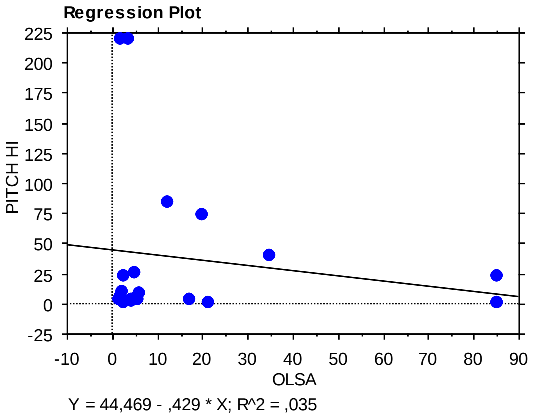
6.4.1 Regression analysis for Speech Noise tests (STARR/OLSA)



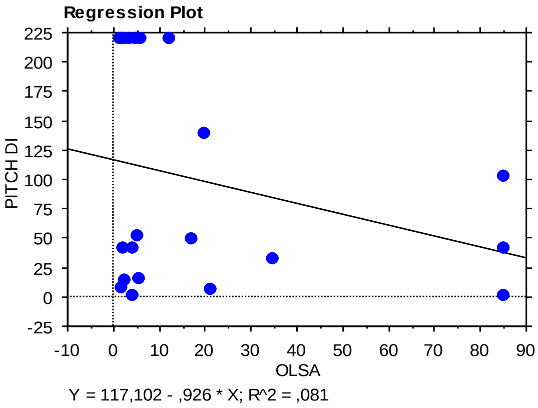
P=0,03



P=0,5



P = 0,4



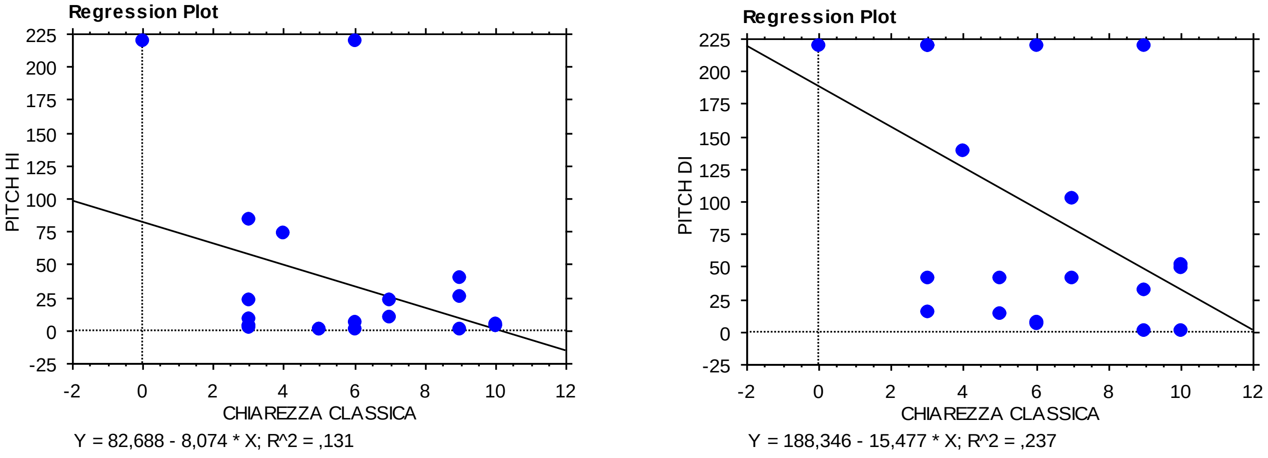
P = 0,2

In this series of analyses – HI vs STARR and OLSA, DI vs STARR and OLSA - an extremely high level of significance is seen uniquely between HI and STARR (P=0,03). This indicates the importance of the ability to process Pitch as shown in HI in relation to the STARR procedure. In fact, of the two S/N test procedures this was the one that proved to be more robust. Furthermore, DI showed no significance for either procedure

6.4.2.Regression analysis for Music

Owing to the complexity of the Questionnaire and the number of Conditions taken into consideration, Regression Analysis was only carried out between Pitch (HI and DI) and the 3 most important criteria: Overall Quality, Pleasantness and Clarity for all 3 Classical, Jazz and Soul tracks.

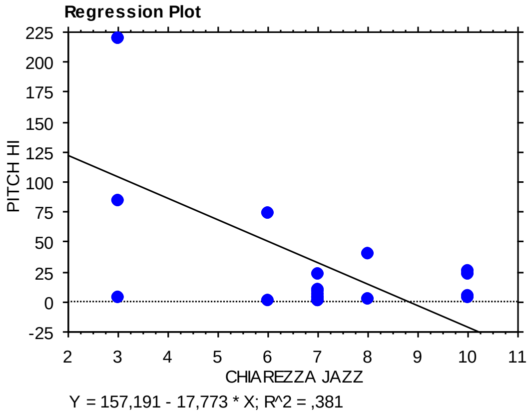
Clarity for Classical



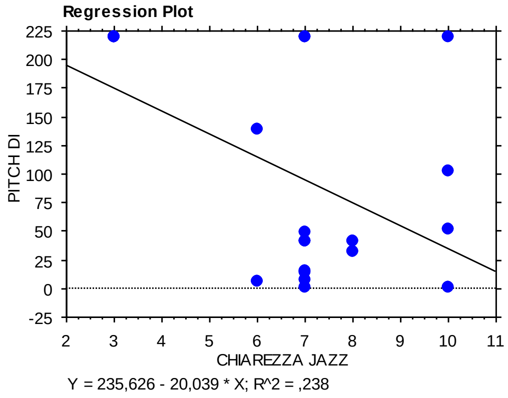
P=0,01

P=0,02

Clarity for Jazz

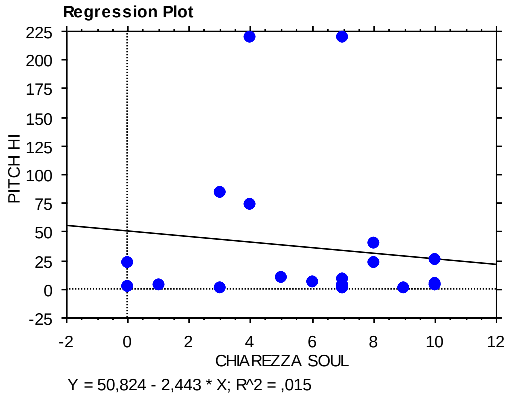


P=0,003

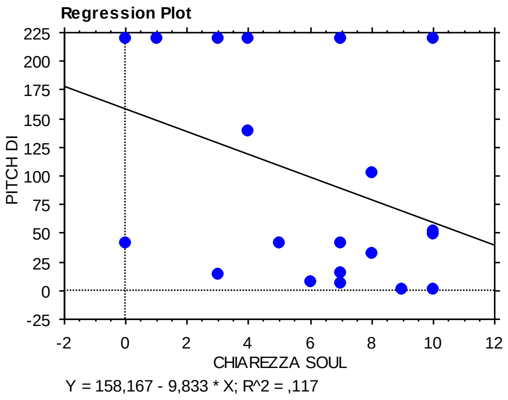


P=0,02

Clarity for Soul

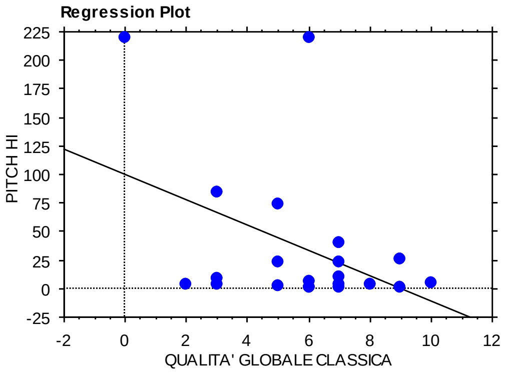


P=0,6

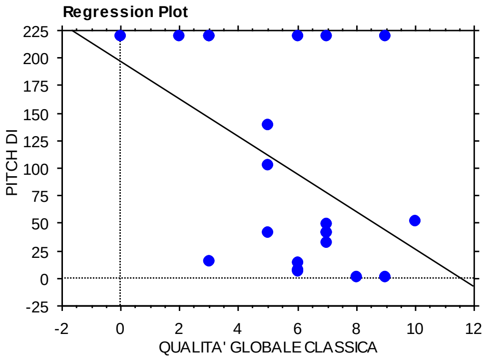


P=0,1

Overall Quality for Classical



Y = 99,537 - 10,998 * X; R^2 = ,175

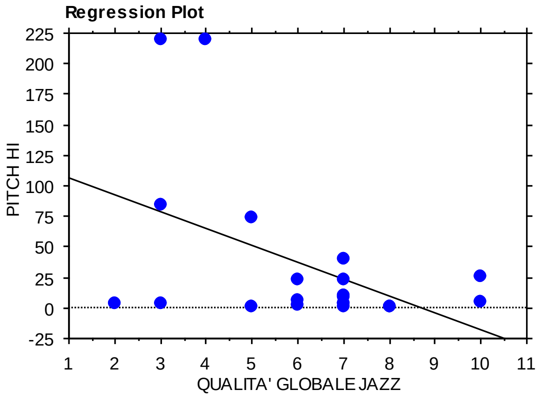


Y = 196,952 - 16,971 * X; R^2 = ,205

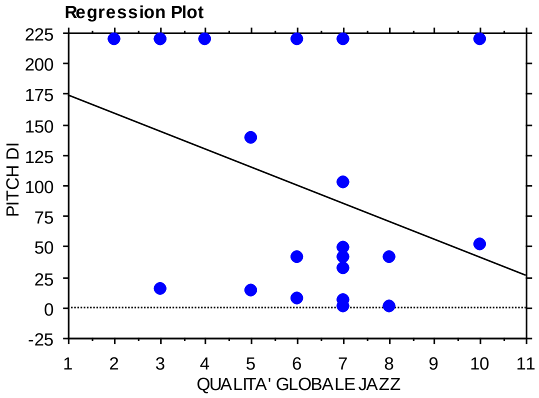
P=0,05

P=0,03

Overall Quality for Jazz



Y = 119,591 - 13,687 * X; R^2 = ,21

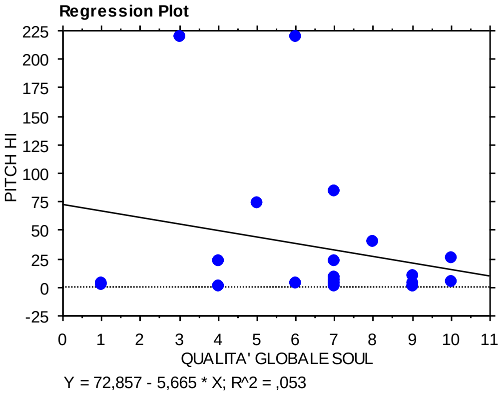


Y = 189,335 - 14,793 * X; R^2 = ,121

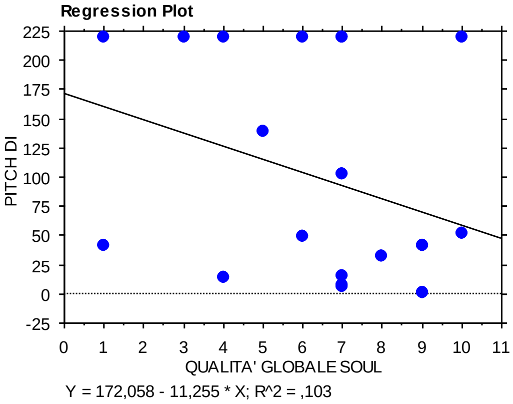
P=0,03

P=0,1

Overall Quality for Soul

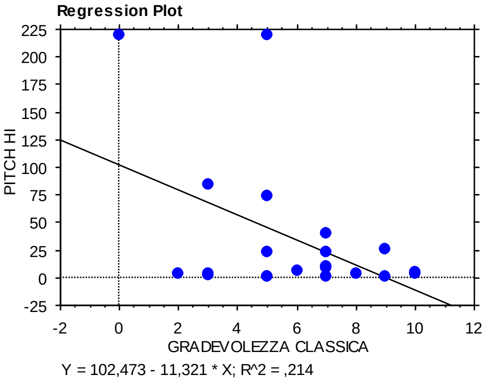


P=0,3

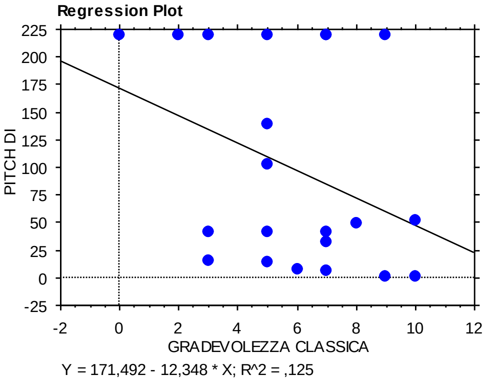


P=0,1

Pleasantness for Classic

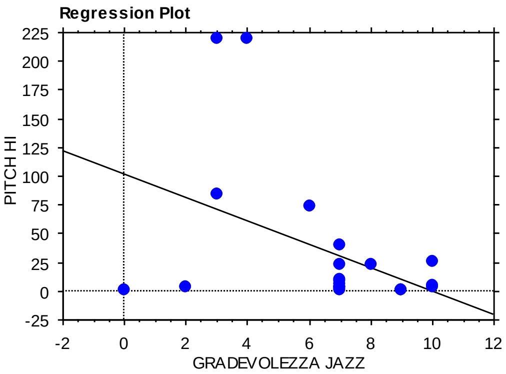


P=0,03



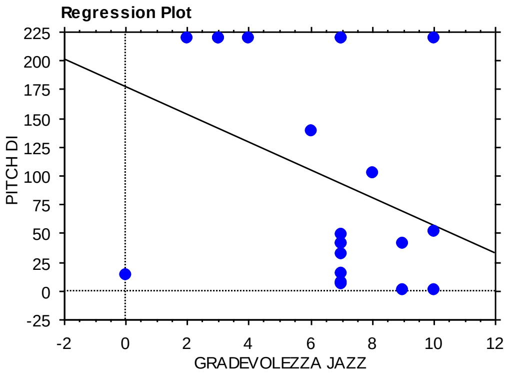
P=0,1

Pleasantness for Jazz



Y = 102,245 - 10,129 * X; R^2 = ,177

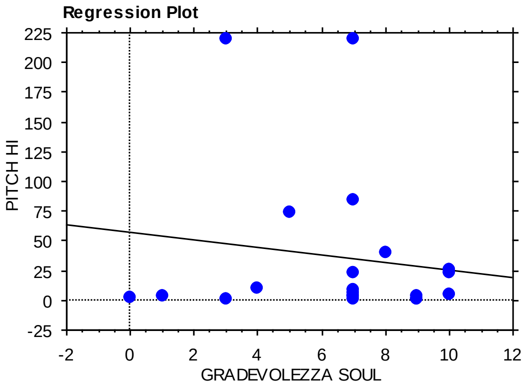
P=0,05



Y = 177,185 - 11,959 * X; R^2 = ,121

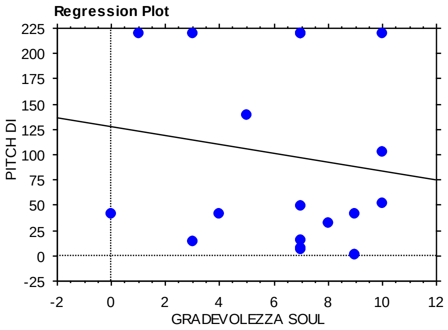
P=0,1

Pleasantness for Soul



Y = 56,986 - 3,191 * X; R^2 = ,02

P=0,5



Y = 127,231 - 4,302 * X; R^2 = ,018

P=0.5

Table 4 Summary of Findings for Regression Analysis

	HI	DI
Overall Quality		
Classical	0,05	0,03
Jazz	0,03	0,1
Soul	0,3	0,1
Pleasantness		
Classical	0,03	0,1
Jazz	0,05	0,1
Soul	0,5	0,5
Clarity		
Classical	0,01	0,02
Jazz	0,003	0,02
Soul	0,6	0,1
OLSA	0,4	0,2
STARR	0,03	0,5

- The test that reflected most the original hypothesis – ie that TFS processing as reflected by Low Frequency Pitch perception was one of the main factors influencing appreciation of Music – proved to be Harmonic Intonation (HI) (ASSE battery)

- In Regression Analysis, a highly significant Inverse correlation was shown between HI and 3 of the 6 criteria taken into consideration (Overall Quality, Pleasantness and Clarity) when listening to both Classical and Jazz: 0,003 – 0,05. This inverse relationship was see constantly (6 conditions)

- DI was only seen to be significant 3 times, ie twice for Classical and Jazz vs Clarity and once for Classical vs Overall Quality.

7. Discussion

Music perception and appreciation in cochlear implant users has been known to be a stumbling block for many years. Clinical research has concentrated on numerous aspects in this field in an attempt to find the right key or ‘link’ between specific peripheral and cortical perception skills and access to the enjoyment of music: for example, Ballantyne et al. (115) showed how, owing to degraded spectral resolution, CI users were unable to perceive any pitch change under a full tone in the context of music; Wright et al (123) failed to show a relationship between music perception skills and enjoyment of music in CI users; Zhang et al (124) in their study on melodic pitch perception - using mismatch negativity (MMN) and Melodic Contour Identification (MCI) – revealed degraded encoding performance in comparison to NH subjects which was attributed to the limited availability of pitch cues provided by the Cochlear Implant as well as deafness-related compromise of the brain substrates. More recently, in 2016 Bruns et al. (117) carried out a study whereby the ability to appreciate the meaning of music was shown in post-lingual adult CI wearers, by using event-related potentials with N400 as a marker.

Despite advances in technology, the ability to perceive music remains limited for many cochlear implant users owing to its **Technological** (disruption of place pitch and rate pitch mechanisms plus dynamic range compression), **Biological** (peripheral, mid-brain and cortical deficits) and **Acoustical** ((temporal and spectral features) constraints.
The limitations of these devices (CI), which have been optimized for speech comprehension, become evident when applied to the appreciation of music, particularly with regards to inadequate spectral resolution, fine-temporal and dynamic range representation (119). Beyond the impoverished information transmitted by the device itself, both peripheral and central auditory nervous system deficits are seen in the presence of sensorineural hearing loss, such as auditory nerve degeneration and abnormal auditory cortex activation
These technological and biological constraints to effective music perception are further compounded by the complexity of the acoustic features of **music** itself that require the perceptual integration of varying rhythmic, melodic, harmonic and timbral elements of the stimuli. CI users not only have difficulty perceiving spectral components individually leading to fundamental disruptions in perception of **Pitch,** melody and harmony, but they also display deficits with higher perceptual tasks required for music perception such as auditory segregation. Despite such limitations improvements in the representation and transmission of the complex acoustic

features of music through technological innovation may offer the potential for significant advancements in cochlear implant-mediated music perception.

Though the main objective of this thesis was to study Music Perception and Appreciation in CI users, in the course of this research it was deemed necessary to carry out preliminary studies which involved the investigation of Temporal Fine Structure processing in CI users as examined by specific Pitch perception tests (HI/DI) which, in turn, could be correlated to outcomes of currently used Speech in Noise tests (STARR/OLSA). All this in an attempt to show up the feasibility of the application of clinical procedures that may underline the problems linked to music perception and appreciation.

Hence, an investigation of TFS processing that is assumed to be reflected in Low Frequency pitch perception capacities, was set up for application to the present study group. For this purpose, the A§E battery was used (102), and specifically two tests of Pitch perception called Harmonic Intonation (HI) and Disharmonic Intonation (DI) as described previously. Although CI recipients had significantly better scores for HI than for DI, the majority showed abnormal outcomes for both HI and DI tests in comparison to NH subjects.

Relative performance deterioration for the DI test was re-indicating DI with more differential outcomes on phase locking and TFS processing capacities, whereas the HI test, due to the sweep of F0 together with its harmonics, could provide some HF cues in the complex signal and therefore, additional place cues for lower JNDs.

In our previous study on this theme we found significant correlations between HI/DI outcomes which confirmed that both tests targeted LF processing capacities. It is these low frequency perception skills that contribute to prosody and timbre in Speech and melody in Music.

Findings from our preliminary study (Chapter 4) suggested that the Italian STARR test could be a promising supplement to existing speech assessment tools, since the average SRT for NH and CI wearers was consistent with SRTs reported for sentence testing by other researchers. The variability of mean SRTs across lists was relatively small. Statistical analysis showed no significant learning effects. The outcomes for Italian STARR test showed statistically significant correlations with those for standard word recognition tests in Italian. In the bimodal group (CI/HA), although speech perception outcomes including those for STARR/OLSA did not show any significant correlations with unaided or aided PTA thresholds for the HA ear, DI outcomes in bimodal listening conditions showed a strong effect of unaided PTA thresholds for octave frequencies lower than 1000 Hz suggesting that bimodal listeners do benefit from the additional LF cues and by remaining phase-locking capacities of residual hearing and in that way the representation of TFS for CI users was improved. However, as can be expected, in this previous study overall outcomes in bimodal listeners were still worse than those of NH counterparts.

Another objective of the present study was to analyse the outcome comparisons between pitch and speech perception for two speech in noise tests (STARR/OLSA) carried out on the study group. The aim was to get a better understanding of the link between LF pitch perception and the processing of TFS in CI users as well as implications for speech perception performance in particular when using a test that attempts to better represent everyday listening situations.

In this study, the success rate for Italian STARR test was excellent. In fact, 67% had SRTs that were considered to be meaningful. All subjects managed to complete the test and a strong positive correlation was also found between HI and STARR (P = 0,03). On the contrary, no correlation was found for OLSA, and not all subjects managed to complete the test which makes it less than feasible for everyday clinical practice. DI was not significantly correlated with either indicating the need for provision of more differential outcomes on phase locking and TFS processing capacities since TFS cues are emphasized in difficult listening tasks such as listening to the dips which is defined as detecting a signal in a fluctuating background.

However, the main core of this study concerned the perception and appreciation of Music in CI users in relation to Pitch perception skills.

In Descriptive Analysis it is interesting to note that almost all median values could be considered sufficient (>5.5) or even optimal (<8) which is indeed encouraging because until fairly recently perception and appreciation of music for CI users was somewhat enigmatic, rather like an emotional mismatch with the subject’s acoustic memory especially in subjects with post-lingual deafness.

Regression Analysis - owing to the complexity of the Questionnaire and the number of Conditions taken into consideration (5 listening modes, 6 subjective quality sensations) - was only carried out between Pitch (HI and DI) and the 3 most important criteria: Overall Quality, Pleasantness and Clarity for all 3 music tracks (Classical, Jazz and Soul) and for the patient's preferred or 'Daily' listening mode

A highly significant Inverse correlation was found between HI and 3 of the 6 criteria taken into consideration (Overall Quality, Pleasantness and Clarity) when listening to both Classical and Jazz: 0,003 – 0,05. This inverse relationship was seen for all 6 conditions but did not necessarily reach significativity. Interestingly, no correlation was found for HI or DI when listening to Soul. It remains to be seen whether this depends on the fact that the input is a single female voice (mid-high frequency, fairly slow rate), hence more simple to process for CI users and not requiring particular pitch skills.

In this study group, the test for Temporal Fine Structure (TFS) processing that reflected most the original hypothesis whereby the ability to perceive Pitch was one of the main factors influencing appreciation of Music proved to be Harmonic Intonation (HI), whereas Disharmonic Intonation was much less effective. To be able to underline the importance of both these aspects linked to subliminal perception (HI/DI), research should continue with a larger study group.

Furthermore, findings were limited by the relatively small number of subjects who managed to complete the whole of this time-consuming study, ie 13 Monolateral, 9 Bimodal, where bilateral CI users were tested one ear at a time. In fact, difference in findings for the two groups were not significant (Student T), but could be if numbers were greater and balanced.

Hence, it would be useful to study the clinical usefulness of HI/DI tests as a predictor of music perception and appreciation in CI users since the availability of LF pitch and TFS cues are even more dominant for music perception

Finally, the improvement of our understanding of the specific deficits in music perception, as demonstrated in CI users, should be based upon the findings reviewed above with a focus on TFS processing and deficits in Pitch perception skills.

8. Conclusions

Harmonic Intonation (HI) and Disharmonic Intonation (DI) are part of the A§E battery for the study of higher perceptive skills, a protocol which can be included in most clinical realities rather than the more complex electrophysiological measures (MCI, MMN) relative to cortical auditory perception functions only available in research clinics. This study has shown a strong correlation particularly between HI and findings for STARR and even Music appreciation.

In this study, Speech perception in Noise was tested by two procedures – STARR and OLSA. The STARR test was the only one that showed a correlation with Temporal Fine Structure processing (Low Frequency Pitch perception). For research purposes, both should be carried out in parallel. OLSA is useful, faster to apply but it has proved to have its drawbacks. Hence STARR would seem to be the procedure of election.

There is an increasing body of research that claims that music training programs, based on brain plasticity and the capacity for auditory learning, could help CI users with current technology to a greater enjoyment of music. This could face the enormous and inexplicable variability in this field between CI recipients, and the drawback of having to resolve technological progress before being able to conceive appreciation of one of the most complex aspects of perception – Music

Whereas Pitch perception skills can be measured by both electrophysiological (MMN, MCI) or clinical (A§E – HI/DI) tools, enjoyment or appreciation of Music are highly subjective and they also lend themselves to scrutiny via questionnaires

Quality of Life questionnaires were also applied during this extensive clinical protocol. Findings were not analysed because not specifically linked to basic hypothesis of this thesis, but will be included in a future study.

Moreover, the group size did not allow us to study any possible impact of the device type on Music perception and appreciation, in particular electrode design and sound coding schemes: HiRes 120 Virtual channels with its increased spectral resonance (Advanced Bionics) vs Fine Structure Processing with accentuation of important Low Frequency input (MedEl)

Hence, improvement in current implant technology must surely contribute to the processing of such complex signal inputs as are those found in all types of Music, perhaps with the development of **Cochlear Implants that specifically target music rather than speech alone?**

9. REFERENCES

1. Flynn S (2003) Cochlear implant systems. In: *Cochlear implants for young children*. McCormick B & Archbold S (Eds) London: Whurr Publishers 11-48.

2. Allum DJ (1996) Basics of cochlear implant systems. In: *Cochlear implant rehabilitation in children and adults.* Allum DJ (Ed) London: Whurr Publishers 1-21.

3. Gelfand SA (2001) Essentials of Audiology second edition. Thieme, New York, USA.

4. Davis H (1962) Advances in the neurophysiology and neuroanatomy of the cochlea. *J Acoust Soc Am* 34:1377-1385.

5. Wilson BS, Dorman MF (2008) Cochlear implants: current designs and future possibilities. *J Rehabil Res Dev* 45:695-730.

6. Stewart L, Adams D (1997) Deafness: Its Implications. In: *Audiology in Education.* McCracken W., Laoide-Kemp S. (Eds) *n*. London: Whurr Publishers 79-106.

7. Hinojosa R, Marion M (1983) Histopathology of profound sensorineural deafness. *Ann N Y Acad Sci* 405:459–84.

8. Moore B.C.J. (1994) Structure and Function of the Ear. In: *An Introduction to the Psychology of Hearing.* London. Academic Press, 3-7

9. Ballantyne J.C.B. (1990) Physiology of Hearing. In: *Handbook of Audiological Techniques.* Ballantyne D. London: Butterworth-Heinemann 46 -50

10. Djourno A, Eyres C (1957) Prosthèse auditive par excitation électrique à distance du nerf sensoriel àl'aid d'un bobinage inclus á demeure. [Auditory prosthesis by means of a distant electrical stimulation of the sensory nerve with the use of an indwelling coil.] *Presse Medicale* 35:14-17.

11. Niparko JK, Wilson BS (2000) History of Cochlear Implants. In: *Cochlear Implants Principles and Practices.* Niparko JK, Iler Kirk K, Mellon NK, Robbins AM, Tucci DL, Wilson BS (Eds). Lippincott Williams & Wilkins, Philadelphia, 103-107.

12. House WF, Berliner K, Crary W, Graham M, Luckey R, Norton N, Selters W, Tobin H, Urban J, Wexler M (1976) Cochlear Implants. *Ann Otol Rhinol Laryngol* Supp 85:1-93

13. House WF, Berliner KI (1986) Safety and efficacy of the House/3M cochlear implant in profoundly deaf adults. *Otolaryngol Clin North Am* 19:275-286.

14. House WF, Berliner KI (1991) Cochlear Implants: from idea to Clinical Practice. In: *Cochlear Implants: A practical Guide.* Cooper H (ed) London: Whurr 9-33

15. Hochmair-Desoyer I, Hochmair E (1996) Implementation of coding strategies in cochlear implant systems and comparative results. *Central East Eur J Otorhinolar Head Neck* 1:33-41.

16. Gantz BJ, Tyler RS, Knutson JF, et al. (1988) Evaluation of five different cochlear implant designs: audiologic assessment and predictors of performance. *Laryngoscope* 10:1100-1106.

17. Cohen NL, Waltzman SB, Fisher SG (1993) The VAH Cochlear Implant Cooperative Study Group: Prospective, randomized study of cochlear implants. The Department of Veterans Affairs Cochlear Implant Study Group. *N Engl J Med* 328:233-237.

18. Doyle PJ, Pijl S, Noel FJ (1991) The cochlear implant: a comparison of single and multichannel results. *J Otolaryngol* 20:204-208

19. Eyles JE, Aleksy WL, Boyle PJ (1995) Performance changes in University College Hospital/ Royal National Institute for the Deaf single-channel cochlear implant users upgraded to the Nucleus 22 channel cochlear implant system. *Ann Otol Rhinol Laryngol Supp* 104:263-265.

20. Shepherd RK, Franz B, Clark GM (1990) The biocompatibility and safety of cochlear prostheses. In: *Cochlear Prosthesis.* Clark GM, Tong YT, Patrick JF (Eds). Melbourne: Churchill Livingstone 69-98.

21. Luxford WM, Brackman DE (1985) The history of cochlear implants. In: *Cochlear Implants.* Gray R (Ed) San Diego: CA: College Hill Press 1-26.

22. Vaerenberg B (2014) Programming Cochlear Implants for Auditory Performance. PhD Dissertation. The University of Antwerp, Antwerp, Belgium.

23. Baumann U, Nobbe A (2004) Pulse rate discrimination with deeply inserted electrode arrays. *Hear Res* 196:49–57.

24. Wilson BS (2000) Cochlear Implant Technology. In: *Cochlear Implants Principles and Practices.* Niparko JK, Iler Kirk K, Mellon NK, Robbins AM, Tucci DL, Wilson BS (Eds). Lippincott Williams & Wilkins, Philadelphia 103-107.

25. Wilson BS (1997) The future of cochlear implants. *Br J Audiol* 31:205-25

26. Kiefer J, Von Ilberg C, Rupprecht V, Hubner-Egener J, Knecht R (2000) Optimized speech understanding with the continuous interleaved sampling speech coding strategy in patients with cochlear implants: Effect of variations in stimulation rate and number of channels. *Ann Otol Rhinol Laryngol* 109:1009–20

27. Friesen LM, Shannon RV, Baskent D, Wang X (2001) Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am* 110:1150–63.

28. Cohen LT, Saunders E, Knight MR, Cowan RS (2006) Psychophysical measures in patients fitted with Contour and straight Nucleus electrode arrays. *Hear Res* 212:160–75.

29. National Institutes of Health Consensus Conference (1995) Cochlear implants in adults and children. *JAMA* 274:1955–61.

30. Wilson BS, Schatzer R, Lopez-Poveda EA, Sun X, Lawson DT, Wolford RD (2005) Two new directions in speech processor design for cochlear implants. *Ear Hear* 26:73S–81S.

31. Niparko JK, Iler Kirk K, Mellon NK, Robbins AM, Tucci DL, Wilson BS (2000) Cochlear Implants Principles and Practices. Lippincott Williams & Wilkins, Philadelphia, USA.

32. Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) Better speech recognition with cochlear implants. *Nature* 352:236– 238

33. Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416:87–90.

34. Donaldson GS, Kreft HA, Litvak L (2005) Place-pitch discrimination of single-versus dual-electrode stimuli by cochlear implant users (L). *J Acoust Soc Am* 118:623–26.

35. Firszt JB, Koch DB, Downing M, Litvak L (2007) Current steering creates additional pitch percepts in adult cochlear implant recipients. *Otol Neurotol* 28:629–36.

36. Koch DB, Downing M, Osberger MJ, Litvak L (2007) Using current steering to increase spectral resolution in CII and HiRes 90K users. *Ear Hear* 28:39S–41S.

37. Wilson BS, Schatzer R, Lopez-Poveda EA (2006) Possibilities for a closer mimicking of normal auditory functions with cochlear implants. In: *Cochlear implants. 2nd ed*. Waltzman SB, Roland JT Jr, (eds). New Your (NY) Thieme 48–56.

38. McDermott HJ, McKay CM (1994) Pitch ranking with nonsimultaneous dual electrode electrical stimulation of the cochlea. *J Acoust Soc Am* 96:155–62

39. Kwon BJ, Van den Honert C (2006) Dual-electrode pitch discrimination with sequential interleaved stimulation by cochlear implant users. *J Acoust Soc Am* 120: EL1–6.

40. Nobbe A, Schleich P, Zierhofer C, Nopp P (2007) Frequency discrimination with sequential or simultaneous stimulation in MED-EL cochlear implants. *Acta otolaryngol* 127:1266-72

41. Hochmair I, Nopp P, Jolly C, Schmidt M, Schösser H, Garnham C, Anderson I (2006) MED-EL Cochlear implants: State of the art and a glimpse into the future. *Trends Amplif* 10:201–19.

42. Arnoldner C, Riss D, Brunner M, Durisin M, Baumgartner WD, Hamzavi JS (2007) Speech and music perception with the new fine structure speech coding strategy: Preliminary results. *Acta Otolaryngol* 127:1298–1303.

43. Green T, Faulkner A, Rosen S (2004) Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J Acoust Soc Am* 116:2298– 2310

44. Riss D, Hamzavi JS, Blineder M, Flak S, Baumgartner WD, Kaider A, Arnoldner C (2016) Effects of Stimulation Rate With the FS4 and HDCIS Coding Strategies in Cochlear Implant Recipients. *Otol Neurotol* 37:882-888.

45. Wouters J, McDermott HJ, Francart T (2015) Sound coding in cochlear implants: from electric pulses to hearing. *IEEE Signal Process Mag* 32:67-80.

46. Lederberg AR, Schick B, Spencer PE (2013) Language and literacy development of deaf and hard-of-hearing children: successes and challenges. *Dev Psychol* 49:15-30.

47. Nicastri M, Filipo R, Ruoppolo G et al. (2014) Inferences and metaphoric comprehension in unilaterally implanted children with adequate formal oral language performance. *Int J Pediatr Otorhinolaryngol* 78:821-827.

48. Mancini P, Dincer D’Alessandro H, Guerzoni L et al. (2015) Adequate formal language performance in unilateral cochlear implanted children: Is it indicative of complete recovery in all linguistic domains? Insights from referential communication *Int J Pediatr Otorhinolaryngol* 79: 598-604

49. Eisenberg LS, Fisher LM, Johnson KC, Ganguly DH, Grace T, Niparko JK. CDaCI Investigative Team (2016) Sentence Recognition in Quiet and Noise. Relationships to Spoken Language. *Otol Neurotol* 37:75-81

50. Park E, Shipp DB, JM Chen, Nedzelski JM, Lin VY (2011) Postlingually deaf adults of all ages derive equal benefits from unilateral multichannel cochlear implant. *J Am Acad Audiol* 22:637-643.

51. Boyle PJ, Nunn TB, O'Connor AF, Moore BCJ (2013) STARR: A speech test for evaluation of the effectiveness of auditory prostheses under realistic conditions. *Ear Hear* 34:203-212.

52. Kong Y-Y, Cruz R, Jones A, Zeng FG (2004) Music perception with temporal cues in acoustic and electric hearing. *Ear Hear* 25:173-185.

53. Hsiao F, Gfeller K (2012) Music perception of cochlear implant recipients with implications for music instruction: a review of literature. *Update Univ S C Dep Music* 30:5–10.

54. Fujita S, Ito J (1999) Ability of Nucleus cochlear implantees to recognize music. *Ann Otol Rhinol Laryngol* 108:634-640.

55. Leal MC, Shin YJ, Laborde M-L, Lugardon S, Andrieu S, Deguine O, Fraysse B (2003) Music perception in adult cochlear implant recipients. *Acta Otolaryngol* 123:826-835

56. Mirza S, Douglas S, Lindsey P, Hildreth T, Hawthorne M (2003) Appreciation of music in adult patients with cochlear implants: a patient questionnaire. *Coch Impl Intl* 4:85-95.

57. McDermott HJ (2004) Music perception with cochlear implants: a review. *Trends Amplif* 8:49-82.

58. Bruns L, Murbe D, Hahne A (2016) Understanding Music with cochlear implants. Scientific Report. *Nature* 6:32026

59. Spahr AJ, Dorman MF (2004) Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices. *Arch Otolaryngol Head Neck Surg* 130:624-628.

60. Spahr AJ, Dorman MF, Loiselle LH (2007) Performance of patients using different cochlear implant systems: effects of input dynamic range. *Ear Hear* 28:260-275.

61. McCreery RW, Venediktov RA, Coleman JJ, Leech HM (2012) An evidence-based systematic review of directional microphones and digital noise reduction hearing aids in school-age children with hearing loss. *Am J Audiol* 21:295-312

62. Cherry EC (1953) Some experiments upon the recognition of speech, with one and with two ears. *J Acoust Soc Am* 25:975-979.

63. Byrne D (1981) Clinical issues and options in binaural hearing aid fitting. *Ear Hear* 2:187-193

64. Veugen LC, Hendrikse MM, van Wanrooij MM, Agterberg MJ, Chalupper J, Mens LH, Snik AF, John van Opstal A (2016) Horizontal sound localization in cochlear implant users with a contralateral hearing aid. *Hear Res* 336:72-82

65. Firszt JB, Reeder RM, Skinner MW (2008) Restoring hearing symmetry with two cochlear implants or one cochlear implant and a contralateral hearing aid. *J Rehabil Res Dev* 45:749-768.

66. Kokkinakis K, Pak N (2014) Binaural advantages in users of bimodal and bilateral cochlear implant devices. *J Acoust Soc Am* 135:EL47-53.

67. Schleich P, Nopp P, and D’Haese P (2004) Head shadow, squelch and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant. *Ear Hear* 25:197-204

68. Livotsky R, Parkinson A, Arcaroli J, Sammeth C (2006) Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study. *Ear Hear* 27:714-731.

69. Buss E, Pillsbury HC, Buchman CA, Pillsbury CH et al. (2008) Multicenter US bilateral MED-EL cochlear implantation study: Speech perception over the first year of use. *Ear Hear* 29:20-32.

70. Sharma A, Dorman MF, Kral A (2005) The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants. *Hear Res* 203:134-43.

71. Mok M, Grayden D, Dowell RC et al. (2006) Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears. J *Speech Lang Hear Res* 49:338-351.

72. Ching TYC, Incerti P, Hill M (2004) Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear Hear* 25:9-21.

73. Dorman MF, Gifford RH, Spahr AJ, McKarns SA (2008) The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiol Neurootol* 13:105-12

74. Van Hoesel RJM (2012) Contrasting benefits from cochlear implants and hearing aids in cochlear implant users. *Hear Res* 288:100-113.

75. Mok M, Galvin KL, Dowell RC et al. (2010) Speech perception benefit for children with a cochlear implant and a hearing aid in opposite ears and children with bilateral cochlear implants. *Audiol Neurotol* 15:44-56.

76. Schauwers K, Coene M, Heeren W, Del Bo L, Pascu A, Vaerenberg B, Govaerts PJ (2012) Perception of pitch changes in hearing impaired adults with aided and unaided hearing loss. *J Hear Sci* 2:OA25-34.

77. Dincer D'Alessandro H, Sennaroğlu G, Yücel E, Belgin E, Mancini P (2015) Binaural Squech and Head Shadow effects in children with unilateral cochlear implants and contralateral heraring aids. *Acta Otorhinolaryngol Ital* 35:343-349

78. Boyle PJ, Büchner A, Stone MA, Lenarz T, Moore BC (2009) Comparison of dual-time-constant and fast-acting automatic gain control (AGC) systems in cochlear implants. *Int J Audiol* 48:211-221.

79. Haumann S, Lenarz T, Büchner A (2010) Speech perception with cochlear implants as measured using a roving-level adaptive test method. ORL *J Otorhinolaryngol Relat Spec* 72:312-318.

80. Hochmair-Desoyer I, Schulz E, Moser L, Schmidt M (1997) The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users. *Am J Otol* 18:S83.

81. O'Neill JJ (1957) Recognition of intelligibility test materials in context and isolation. *J Speech Hear Disord* 22:87-90.

82. Cutugno F, Prosser S, Turrini M (2000) *Audiometria Vocale,* vol. IV. Padova, GN Resound Italia.

83. Voghera M (1992) Sintassi e intonazione nell'italiano Parlato. Il Mulino, Bologna

84. Turrini M, Cutugno F, Maturi P, Prosser S, Albano Leoni F, Arslan E (1993) Nuove parole bisillabiche per audiometria vocale in lingua italiana. *Acta Otorhinol Ital* 13:63-67.

85. De Mauro T (1989) Guida all'uso delle parole. Editori Riuniti, Rome

86. De Mauro T (1989) Guida all'uso delle parole. Editori Riuniti, Rome

87. ANSI/ASA S3.1-1999: Maximum permissible ambient noise levels for audiometric test rooms. Washington, USA: American National Standards Institute

88. Joffo LM, Boyle P (2010) Star2 Validation Working Group: Cochlear Implant Users' Everyday-life performance assessed in clinical practice. *Cochlear Implants Int* 11:52-56.

89. Wagener K, Kühnel V, Kollmeier B (1999) Entwicklung und Evaluation eines Satztests für die deutsche Sprache. 1. Design des Oldenburger Satztests. *Z Audiol/Audiol Acoust,* 38:4-15.

90. Nilsson M, Soli SD, Sullivan JA (1994) Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 95: 1085-1099.

91. Zokoll MA, Hochmuth S, Warzybok A et al. (2013) Speech-in-noise tests for multilingual hearing screening and diagnostics1. *Am J Audiol* 22:175-178

92. Hochmuth S, Brand T, Zokoll MA, Castro FZ, Wardenga N, Kollmeier B 82012) A Spanish matrix sentence test for assessing speech reception thresholds in noise. *Int J Audiol* 1:536-544

93. Thornton AR, Raffin MJ (1978) Speech discrimination scores modeled as a binomialvariable. *J Speech Hear Res* 21:507-518

94. Puglisi GE, Warzybok A, Hochmuth S et al. (2015) An Italian matrix sentence test for the evaluation of speech intelligibility in noise. *Int J Audiol* 15:1-

95. Vaillancourt V, Laroche C, MayerC et al. (2005) Adaptation of the HINT (hearing in noise test) for adult Canadian Francophone populations. *Int J Audiol* 44:358-369

96. ISO 8253-3:2012 Acoustics – Audiometric test methods – Part 3: Speech Audiometry.

97. Chen F, Zhang YT (2008) A novel temporal fine structure-based speech synthesis model for cochlear implant. *Signal Processing* 88:2693-2699.

98. Gifford R, Shallop J, Peterson A (2008) Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol Neurootol* 13:193-205

99. Zeng FG (2002) Temporal pitch in electric hearing. *Hear Res* 174:101–6.

100. Moore BJC, Vickers DA, Mehta A (2012) The effects of age on temporal fine structure sensitivity in monaural and binaural conditions. *Int J Audiol* 51:715-721

101. Vaerenberg B, Pascu A, Del Bo L, Schauwers K, De Ceulaer G, Daemers K, Coene M, Govaerts PJ (2011) Clinical assessment of pitch perception. *Otol Neurotol* 32:736-741

102. Heeren W, Coene M, Vaerenberg B, Avram A, Cardinaletti A, Del Bo L, Pascu A, Volpato F, Govaerts PJ (2012) Development of A§E test battery for assessment of pitch perception in speech. *Cochlear Implants Intl* 13:206-219

103. Dincer D’Alessandro H, Filipo R, Ballantyne D, Attanasio G, Bosco E, Nicastri M, Mancini P (2015) Low-frequency pitch perception in children with cochlear implants in comparison to normal hearing peers. *Eur Arch of Otorhinolaryngol* 272:3115-3122

104. Vaerenberg B, Péan V, Lesbros G, De Ceulaer G, Schauwers K, Daemers K, Gnansia D, Govaerts PJ (2013) Combined electric and acoustic performance with Zebra® speech processor: speech reception, place and temporal coding evaluation. *Cochlear Implants Intl* 14:150-157

105. Dincer D’Alessandro H, Ballantyne D, De Seta E, Musacchio A, Mancini P (2016) Adaptation of the STARR test for adult Italian population: A speech test for a realistic estimate in real-life listening conditions. *Int J Audiol* 55:262-267

106.Vaerenberg B, Heeren W, Govaerts PJ (2013) Managed estimation of psychophysical thresholds. *J Hear Sci* 3:19-31.

107. Pichora-Fuller M (2003) Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *Int J Audiol* 42:S59–67

108. Dorman M, Spahr A, Gifford RH, Cook S, Zhanga T, Loiselle L, Whittingham J, Schramm D (2012) Bilateral and bimodal benefits as a function of age for adults fitted with a cochlear implant. *J Hear Sci* 2:EA37–EA39.

109. Hopkins K, Moore BCJ (2011) The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J Acoust Soc Am* 130:334-349.

110. Heng J, Cantarero G, Elhilali M, Limb CJ (2011) Impaired perception of temporal fine structure and musical timbre in cochlear implant users. *Hear Res* 280:192-200.

111. McDermott H.J., (2004). In Trends in Amplification. Music Perception with Cochlear Implants: A Review- Vol. 8, n°2, p 49-82

112. Gfeller K et al. (2007) Accuracy of cochlear implant recipients on pitch perception, melody recognition and speech perception in noise. *Ear Hear*. 28, 412-423

113. Brockmeier S.J. et al. (2011) The MuSIC perception test. A novel battery for testing music perception in cochlear implant users. *Cochlear Implants Int*. 12, 10-20

114. Sandmann P et al. (2010) Neurophysiological evidence of impaired musical sound perception in cochlear implant users. *Clin. Neuophysiol. Off. J.Int.Fed.Clin.Neurophysiol*. 121, 2070-2082

115. Filipo R, Ballantyne D, Mancini P and D'Elia (2008) *Acta Oto-Laryngologica*, 128: 378-381

116. Shannon RV Psychophysics. In: R.S. Tyler, ed. Cochlear Implants: audiological foundations. (1993) San Diego: Singular, 357-388

117. Bruns et al. (2016) Understanding Music with Cochlear Implants. *Scientific Reports. Nature*

118. Moore, D.R., & Shannon, R., V. (2009) Beyond Cochlear Implants: awakening the deafened brain *Nat. Neuroscience* 12, 686-691.

119. Limb C.J. & Roy A.T.(2014) Technological, biological and acoustic constraints to music perception in cochlear implant users. *Hear. Res*. 308, 13-26

120. Kutas M. & Federmeier K.D. (2011) Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev.Psychol*. 62, 621-647

121.Vaerenberg B, Pascu A, Del Bo L, Schauwers K, De Ceulaer G, Daemers K, Coene M, Govaerts PJ (2011) Clinical assessment of pitch perception. *Otol Neurotol* 32:736–741

122. Puglisi G., Warzbok A., Hochmuth S., Visentin C., Astolfu A., Prodi N., Kollmeier B (2015) An Italian matrix test for the evaluation of speech intelligibility in noise. *Int. J. Audiology,* 54:44-50.

123. Wright R, Uchanski RM (2012). Music perception and appraisal: cochlear implant users and simulated cochlear implant listening. *J.Am.Acad.Audiol., 23 (5) 350-65.*

124. Zhang F., Benson C., Fu Q (2013). Cortical encoding of Pitch contour changes in cochlear implant users: a Mismatch negativity study. *Audiol. Neurotol: 18, 275 288.*