TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DIPLOMARBEIT

Evaluation of the Simulated Phase-Locking Stimulation Strategy versus the CIS Strategy for Cochlear Implants

ausgeführt am Institut für

Analysis and Scientific Computing

der Technischen Universität Wien

unter der Anleitung von

Ao.Univ.Prof. Dipl.-Ing. Dr.sc.med. Dr.techn. Dr.rer.nat. Frank **Rattay**

durch

Lucia Navarro de Lara

Vorgartenstrasse 129-143/2/11

1020 Wien

Wien, Mai 2011

KURZFASSUNG

Das Cochlea Implantat ist ein elektronisches Gerät, das unter der Haut implantiert wird und dessen Elektroden, den Gehörnerv in der Cochlea stimulieren. Elektrische Ströme induzieren Aktionspotenziale in den Fasern des Gehörnervs, die zum Gehirn weiter übertragen werden. In dieser Diplomarbeit wird besondere Aufmerksamkeit den Sprachverarbeitungsstrategien geschenkt, da sie die Schlüssel zu einer guten Leistung des cochlea Implantats sind. Das Ziel der Sprachverarbeitung-Einheit ist die Wandlung des Eingangssignale aus dem Mikrophon zu einem Satz von Impulse, die vom Nervensystem richtig interpretiert werden kann. Eine mögliche Methode für diese Transformation, und eine sehr erfolgreiche, ist das Reproduzieren der zerstörten oder fehlenden physiologischen Funktionen, die vom Neuroprothese überbrücken werden müssen.

Die verschiedenen Sprachverarbeitungsstrategien werden hier in der zeitliche Reihenfolge, wie sie entwickelt und vom CI verwendet wurden, präsentiert. Die Weiterentwicklungen der neuen Strategien im Bezug auf die bisherigen Strategien und deren Nachteile und Schwachen werden hier auch behandelt. Eine ganz neue Sprachverarbeitung Strategie entwickelt bei Chen et al. 2009, die so genannte Simulated Phase-Locking Stimulation (SPLS), wird auch behandelt. Ergebnisse aus den Simulationen des CIS und SPLS Strategien im MATLAB bei der Veränderung verschiedenen Parameter, wie Frequenzbereich, Anzahl von Kanälen, Filtertyp oder Bandbreite Art, werden in MATLAB dargestellt und besprochen. Um beiden Strategien vergleichen zu können, wurden Testdaten aus dem H-LAD Test (Heidelberg Lautdifferenzierungstest) für beiden Strategien (CIS and SPLS) akustisch simuliert. Die simulierten Testsdaten wurden verwendet um den Test bei 5 gesunden Hörenden durchzuführen. Das Ergebnis des „simulierten“ H-LAD Test wird im letzten Teil der Diplomarbeit beigefugt und besprochen. Verbesserungspotenzialen werden im letzten Teil kurz beschrieben.

ABSTRACT

The cochlear implant (CI) is an electronic device implanted under the skin with electrodes positioned in the cochlea to stimulate the auditory nerve. Electrical currents induce action potentials in the auditory nerve fibers and these are transmitted to the brain. In this master thesis, a special attention will be done to the speech processing strategy, which is considered the key module for an efficient performance of such implants. The goal of the speech processing unit is to transform the input signal from the microphone into a set of stimuli that can be interpreted by the brain. One approach for this transformation, and a very successful one, is to reproduce the damaged or missing physiological functions that have to be bypassed by the neuroprosthesis. The best way to achieve an excellent performance would be to mimic the exact physiological mechanism. However some important details of the hearing mechanism are still being discussed.

In this thesis different speech processing strategies are presented in a chronological way, as developed and used by the CI. Their enhancements respect to the previous strategies and their handicaps and weakness are discussed as well. Finally, a novel speech strategy developed by Chen et al. 2009, the Simulated Phase-Locking Stimulation (SPLS), is also be outlined. Results of the simulations done with MATLAB of the Continuous Interleaved Sampling (CIS) and the SPLS strategies using different important parameters, such as frequency range, number of channels, type of filter and kind of bandwidth, are discussed. In order to compare both strategies, test data from the H-LAD (Heidelberg Lautdifferenzierungstest) are acoustically simulated with both strategies (CIS and SPLS). The obtained simulations were used to realize the test in 5 different normal listeners. The results of the H-LAD (Heidelberg Lautdifferenzierungstest) using the simulations are presented in the last part of the thesis. Further improvements are presented in the conclusion.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to express special thanks to my little boy Luis and to my older children David and Paloma. They just accepted that their mother is a person who cannot stay at home and that likes her work, something not always well accepted by the "modern" society. Luis was from the first day on such an easy baby that I could find some time to do a lot of things for the master and for this project. I also want to thank Prof. Rattay for accepting me as "Diplomandin" and having patience in difficult times, when I just didn't find the time to go on due to different reasons. Thanks also for answering all my questions before I almost asked them and for being there always, whatever and whenever I needed. I also would like to thank my father, who always supports me in all my projects, believing in me.

GLOSSARY

ACE	Advanced Combination Encoder
AN	Auditory Nerve
BM	Basilar Membrane
BP	Bipolar Electrodes Configuration
CA	Compressed Analogue
CI	Cochlear Implants
CID	Central Institute for the Deaf
CIS	Continuous Interleaved Sampling
ERB	Equivalent Rectangular Bandwidth.
F0	Fundamental Frequency
F1, F2 and F3	First 3 Formants
FS	Fein Structure Information
H-LAD	Heidelberger Lautdifferensierungstest
HiRes®	High Resolution Strategy
IHC	Inner Hair cells are
MP	Monopolar Electrode Configuration
OC	Organ of Coti
OHC	Outer Hair Cells.
pTP	partial Tripolar electrode configuration
RTI	Research Triangle Institute

SM	Scala Media
SPEAK	Spectral Peak Strategy
SPLS	Simulated Phase-Locking Stimulation
ST	Scala Tympani
SV	Scala Vestibuli
TP	Tripolar electrode configuration

Chapter 1: The auditory system and its prosthesis: the cochlear implant

Before a detailed exposition about cochlear implants can be done, some important aspects of normal hearing should be first briefly discussed. Our goal here is not to present an exhaustive description of the anatomy and physiology of the auditory system. This can be read somewhere else and is not the ambition of this work. However, in order to understand the evolution of the speech strategies of the cochlear implants in the last decades, the most important theories that try to explain how the information of an acoustic signal is transmitted or coded to be understood by the brain should be outlined and explained. These should help us to understand the limitations of some strategies and why the applied research is going to another kind of speech processing approaches.

Aspects of normal hearing

Sound waves travel through air and reach the tympanic membrane causing vibrations that move the three small bones of the middle ear. This produces a piston-like movement of the third and last bone in the chain, the stapes. This little bone is attached to the oval window, a flexible membrane in the body shell of the cochlea. The movements of this membrane induce pressure oscillations in the cochlear fluids, which in turn initiate a travelling wave of displacement along the basilar membrane (BM) (see Figure 1). The basilar membrane is a highly specialized structure that divides the cochlea along its length. This membrane has important mechanical properties. At the base, near the stapes and oval window, it is narrow and stiff. At the other end, near the apex, it is wide and flexible. These properties give rise to the travelling wave and to points of maximal response according to the frequency or frequencies of the pressure oscillations. High frequencies produce maxima near the base of the cochlea and low frequencies produce maxima near the apex (see Figure 3).

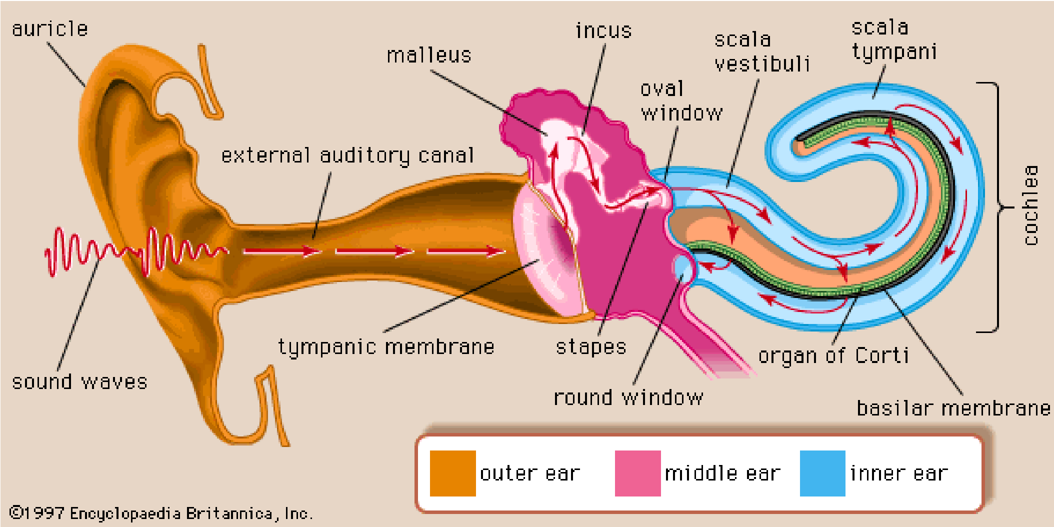


Figure 1 Anatomy and physiology of the ear. Encyclopædia Britannica, Inc. [see website link]

Motion of the BM is perceived by the inner hair cells (IHCs) in the cochlea, which are attached to the top of the membrane in a matrix of cells called the organ of Corti. Each hair cell has fine rods of protein, called stereocilia, coming out from one end. When the BM moves at the location of a hair cell, the rods are deflected as they were articulated at their bases (see Figure 2). Such deflections in one direction increase the release of chemical transmitter substance at the other end of the cells, while deflections in the other direction inhibit the release. According to the direction of the deflections, the variations in the concentration of the chemical transmitter substance may inhibit or increase discharge activity in neighboring neurons. That means, events at the BM are reflected by changes in neural activity. These changes are transmitted to the brain via the auditory nerve, which is the collection of all neurons that innervate the cochlea.

How the sound is really transferred to the brain has been tried to be explained using two important theories, the "place theory" and the "temporal theory". The "place theory" was introduced by Helmholtz in 1863 [Helmholtz, 1863]. For his theory he assumed that:

1. The BM consists of segments with varying stiffness.

2. Frequency selectivity is based on the resonance of a corresponding part

3. Each part is connected with a nerve fiber.

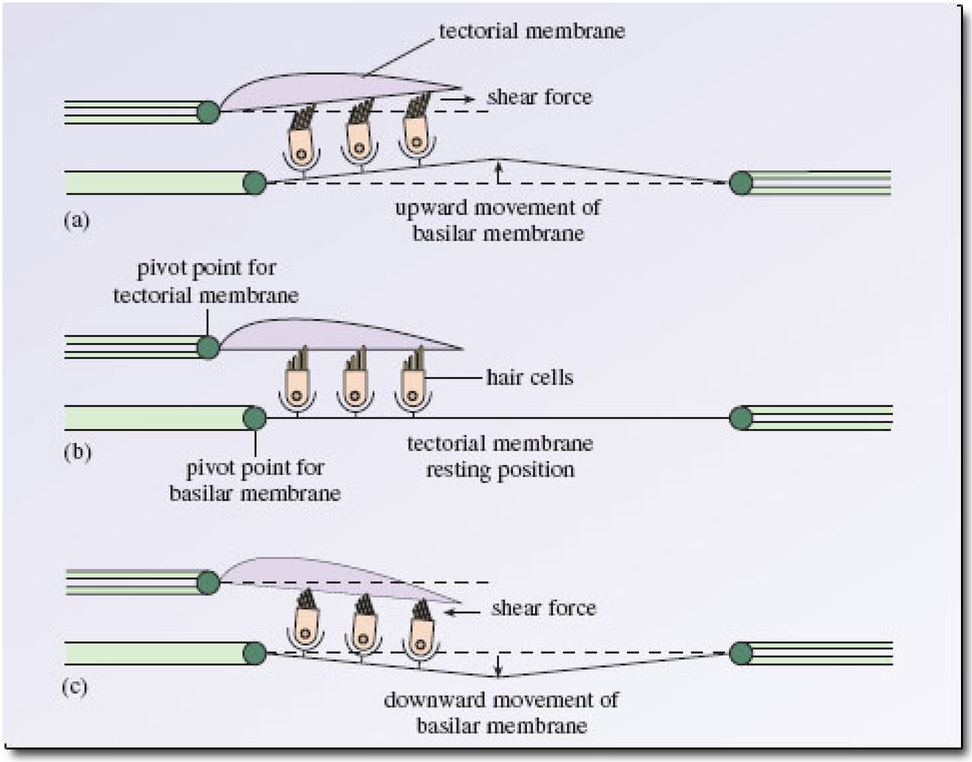


Figure 2 Schematic diagrams of shearing forces created between the hair cells and the tectorial membrane as a result of basilar membrane displacement. Figure from openlearn website [see website link].

In other words, cochlear nerve fibers preserve the frequency selectivity found along the BM. Fibers on the outside of the auditory nerve bundle (those that innervate the basal hair cells) have higher characteristic frequencies than the fibers towards the middle of the nerve bundle (those that innervate the apex of the cochlea). Thus, each place or location within the nerve responds best to a particular frequency. This arrangement is also known as the tonotopic organization.

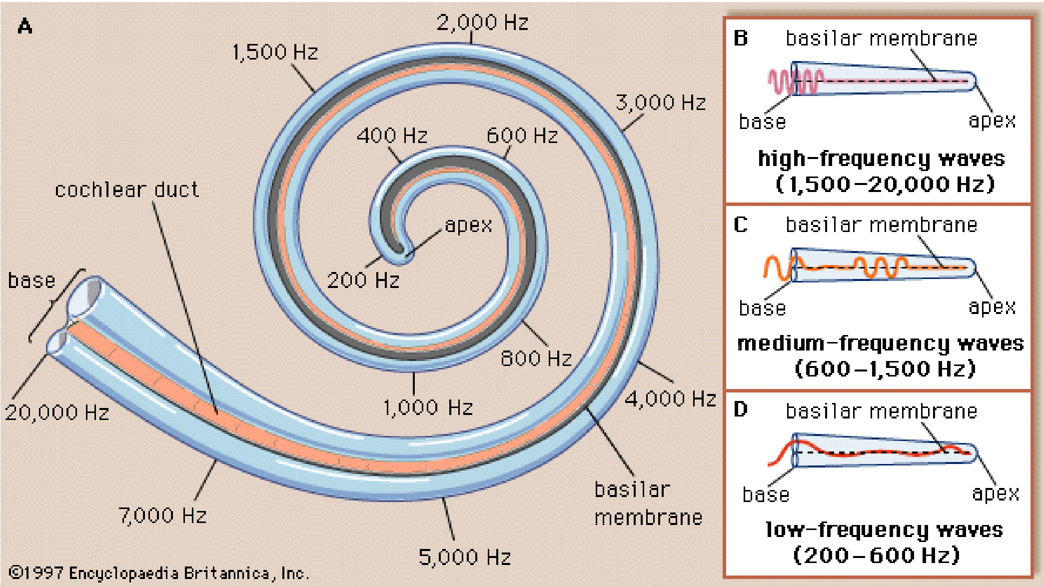


Figure 3 Model showing the distribution of frequencies along the basilar membrane of the cochlea. Encyclopædia Britannica, Inc [see website link]

The "temporal theory" was first mention in Rutherford's work, who postulated that each hair cell in the cochlea responds to every tone coming into the ear [Rutherford, 1886]. This theory proposes that in response to a pure tone, the vibration of the BM matches the input frequency. It is also suggested that the temporal pattern of the BM vibration is reproduced in the firing pattern of the neuron. This is also known as a phase-locked response, since the response appears to be locked to a certain point (e.g. the peak) of the stimulus (see Figure 4a). In such a case, the response pattern of the nerve fiber would accurately reflect the frequency of the sound wave. That why this theory is also called the frequency code. However, this could only be valid for low frequencies. Due to the absolute refractory period of about 1 ms of the neurons, they cannot fire much faster than about 1000 action potentials per second. This realization led Wever and Bray [Wever, 1954] to propose the operation of the volley principle. Figure 4b illustrates their proposal.

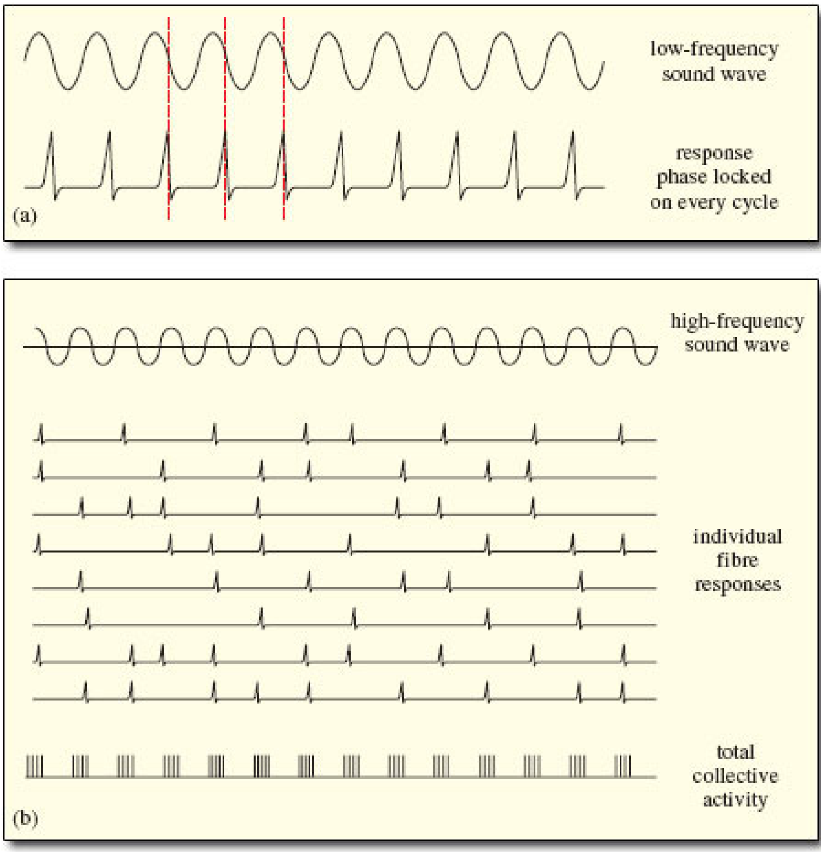


Figure 4(a) Phase-lock principle. (b) Volley principle. Figure from openlearn website [see website link]

When frequency of the sounds is too high for a single fiber to fire on every cycle, according to the volley principle, each fiber only fires at a certain point of the cycle and does not respond to each cycle. None of the individual fibers reproduces the pattern of the wave, but the combined response is sufficient to reproduce the frequency of the incoming signal. This principle can explain the phase locking mechanism to signals with frequencies up to 5 kHz. Above this level, the variability inherent in neural firing is too great for such fine patterns to be resolved.

This is the old controversy whether the inner ear uses the place coding principle or the temporal structure of the spiking pattern. Actually, it is accepted that acoustic signals are represented in the auditory nerve both by place-rate information and by the fine structure of the time differences in the firing pattern. For those parts of human speech which are quite aperiodic the place-rate information is very poor because the mechanical tuning process is slow and requires time. In such cases, the time structure of the firing pattern seems to carry much higher significant information than assumed before the presentation of computer simulation studies on inner ear mechanism done by Rattay and Lutter in 1997 [Rattay et al., 1997]. In these studies the importance of the temporal information for quasiperiodic parts of speech could be proved as well.

Cochlear Implants

Cochlear implants are the most successful neuronal prosthesis of the past forty years. The first devices were developed and implanted in the early seventies. They just proved to the scientific community that electrical stimulation on the hearing nerves could result in meaningful sound perception.

The developed devices just provided a sound sensation and were used as aid for lip-reading. In the 80s, the multiple channels for speech processing (using filter bank strategies) and multiple sites of stimulation in the cochlea (due to arrays of electrodes) achieved a much better level of speech perception than their predecessor, which had a single-channel strategy and single-site stimulation (see Figure 5). At the end of the eighties, new and better speech processing strategies and the multielectrode technology perform further improvements in intelligibility of the perceived sound.

The principal cause of deafness or hearing loss is the damage or complete destruction of the sensory hair cells. The sensory hells are very sensible structures which are exposed to several distresses, such as genetic defects, infectious diseases, exposure to very loud sounds, secondary effects of some drugs and aging. When the damaged or destructed structures are the OHC's, the hearing thresholds are elevated and the frequency resolution is degraded.

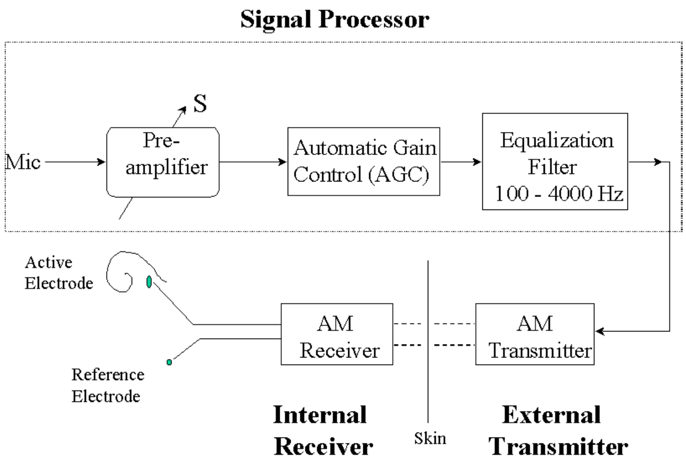


Figure 5 Block diagram of the Vienna/3M Cochlear Implant, one of the first implants developed at the Technical University of Vienna in the early 1980s. From Loizou, 1998.

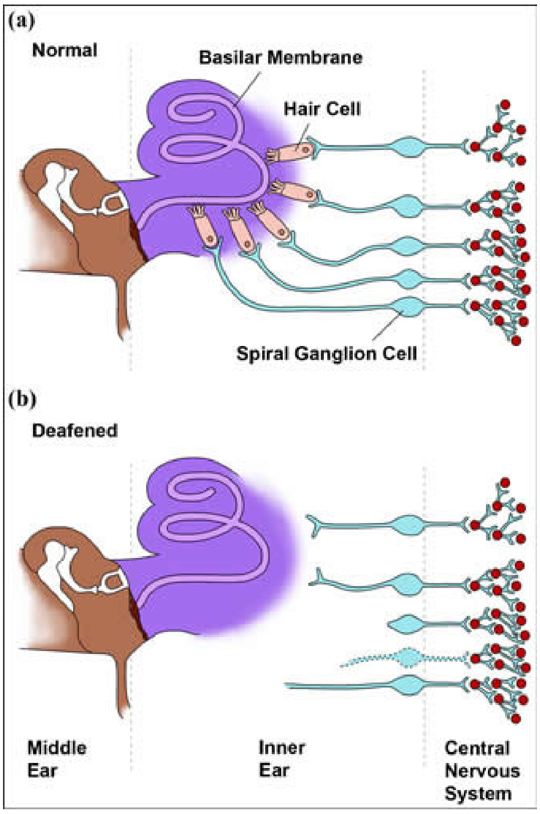


Figure 6 Healthy and Deafness Ear Comparison. Figure from Dorman MF, Wilson BS. The design and function of cochlear implants. From Dorman and Wilson 2004.

When the affected structures are the IHC's, more severe damages occur up to total deafness. In a deafened cochlea, the IHC's are not able to stimulate the connection between the peripheral nervous system and the central nervous system.

This is the function that has to overtake the CI. In Figure 6a simple schematic picture shows the difference between a healthy and a deafened cochlea.

A cochlear implant makes sense in patients with enough auditory nerve fibers left for stimulation in the near of the electrodes. The Figure 6b shows the case where all the IHC's were damaged. Usually, a small number of sensitive hair cells remain left, but not enough to stimulate the ganglion cells in order to produce neuronal stimulation. Because of the absence of stimulation, the peripheral parts of the neurons, between cell bodies in the spiral ganglion and the terminals parts inside the organ of Corti, undergo "retrograde degeneration" and end their function. It has been observed that fortunately, the cell bodies are more robust than the peripheral parts and many of them survive, even after prolonged deafness. These parts, concretely the nodes of Ranvier near to them, are the sites for the electrical stimulation that have to be reached by the electrodes of the CI.

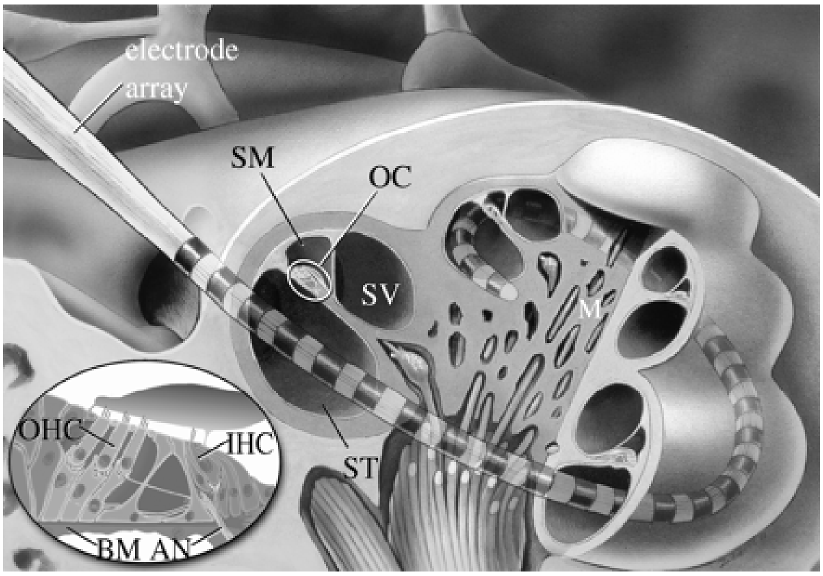


Figure 7 Detail of an array of electrodes implanted in the cochlea in ST. Figure from Graeme, 2006.

The aim of a cochlear implant system is to reproduce the "tonotopic" encoding of the inner ear. The direct stimulation of the auditory nerve is produced by currents delivered through electrodes placed in the Scala Tympani (ST), one of the three fluid-filled chambers along the cochlea (see Figure 7). The different electrodes of the implanted array stimulate different subpopulations of neurons, depending of the frequency content of the received sound. That means, the electrodes basally situated will be stimulated in the presence of high-frequency sounds. Otherwise, the most apical electrodes will be activated when sound with lower frequencies is being received.

The loudness of a perceived sound depends on the number of activated nerve fibers and their rates of firing. If a sound is loud, a large number of nerve fibers should be stimulated. Otherwise, by a soft sound, only a small number of nerve fibers will be activated. The number of stimulated fibers is a function of the amplitude of the stimulus currents at the corresponding electrodes. That means that the loudness of a sound can be therefore controlled by changing the amplitude of the stimulus current on the electrode.

Elements of a Cochlear Implant

All different Cochlear Implants developed over the last past decades have the same common elements (see Figure 8). The first element that enters in contact with the sound is the **microphone**. It converts the sound pressure signal in an electrical one to be processed by the **speech processor**. The **battery pack**, which is the source of energy for the system, the microphone and the speech processor are packed into a behind-the-ear housing (BTE), like the ones used by the hearing aids. The speech processor has to convert the electrical signal in a set of stimuli to be applied to the array of the electrodes. The way this will be done defines the speech strategy of each implant. An **external transmitter** is responsible for the transmission of power and the stimulus information to the implanted **receiver/stimulator**. The receiver is usually implanted in a flattened portion of skull, posterior to and slightly above the pinna. It decodes the RF-Signal sent by the external transmitter and conducts the stimulus Information to an **array of electrodes** through a cable. The array of active electrodes is inserted into scala tympani (ST) through round window membrane or through larger drilled opening in bony shell of cochlea (cochleostomy) near round window. The **reference electrode** (or 'ground' electrode) is usually implanted as far as possible from the cochlea, usually attached in temporalis muscle. Each of the CI components has to work together as a system to achieve an excellent performance for the patients. The weaknesses or limitations of each of them, for example the dynamic range of a microphone or the bandwidth of the data link for the transcutaneous link, can degenerate the performance of the system substantially.

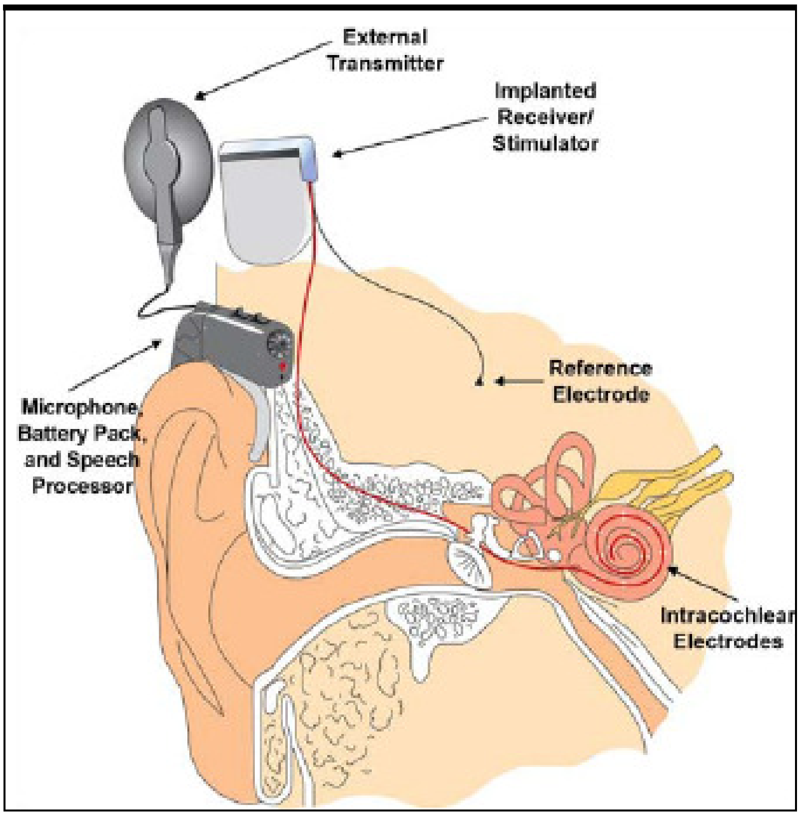


Figure 8. Components of a cochlear implant system. TEMPO+ system of MED-EL Medical Electronics GmbH. Figure from Wilson and Dorman, 2008.

Characteristics of Cochlear Implants

There are several types of cochlear implants that have been developed over the last past years. The differences among them can be summarized as the characteristics of a CI, which are presented here:

- **Electrode design:** The goal of the electrode design is to maximize the number of non overlapping neuron populations that can be stimulated with each single electrode. The design doesn't only define the number of electrodes, but also the spacing between them, the configuration in which will be used (monopolar or bipolar), their placement in the cochlea and the orientation of the electrodes with respect to the excitable tissues.

  The number of electrodes and the spacing between them define in the design of a cochlear implant, the place-frequency resolution of the prosthesis. In principle, a larger amount of electrodes should improve the place resolution for the frequency coding. Unfortunately, the frequency resolution is constrained by the number of surviving neural tissue in each location that should be stimulated and also by the spread of the electric field from adjacent or non-distant electrodes. With the present research, no more than 4 to 8 independent sites were able to be stimulated.

  The first exposed problem cannot be addressed in the design, because it depends on the etiology of the deafness of each patient. For the second problem, the overlaps are unavoidable for electrode placements in the Scala Tympani (ST), because the electrodes are surrounded by the perilymph, which is a highly conductive liquid. The electrodes are also placed relatively far away from the target neural tissue, and because of this, the currents required by the electrodes have to be high enough to reach the desired stimulation. If the electrodes could be placed closer to the target tissue, moving them to the inner wall of the ST, the stimulation amplitudes could be reduced and better spatial specificity stimulation could be achieved. However, if a larger number of stimulation sites should be reached, either new type of electrodes should be developed or new placement for these should be found.

The meaning of monopolar (MP) and bipolar (BP) configuration has to be also explained. In the first one, the active electrode is located far from the reference electrode. The reference electrode acts as a ground for all electrodes. In the bipolar configuration, the active and the reference electrode are place close to each other. In this case it has been shown that the bipolar electrode configuration produces more located stimulation than the monopolar one. Though, the monolopar configuration is being used primarily as it supports performances that are at least as good as the bipolar one and it also requires less current and battery power to reach the thresholds to produce auditory perception.

A relatively novel stimulation strategy is the partial tripolar electrode configuration (pTP). This can generate more spatially focused electric fields than the commonly used monopolar configuration. Focused stimulation strategies are the goal of the CI because it should improve spectral resolution and decrease channel interactions effects. There are other effects that have been evaluated by Goldwyn et al. [Goldwyn et al., 2010]. They have investigated the effects of electrode configuration electrode, electrode-neuron distance, and spiral ganglion neural survival on spatial spread of excitation, threshold levels, and growth of neural activation. Their model has shown 3 important principles:

1. The spatial pattern of neural activation becomes more selective as electrode configuration varies from MP to TP.

2. Configurations that are more localized require higher current levels to reach perceptual threshold, and the threshold current levels are

more sensitive to variations in electrode-neuron distance and loss of auditory neurons.

3. The effects of variations in electrode-neuron distance and neural-survival are not independent. For close electrode-neuron distances, become more spatially localized and consequently have a greater sensitivity to local neuron loss.

New challenges appear with the use of pTP or TP electrode configurations. The large current requirements for the TP configuration and its sensitivity to electrode-neuron distance and loss of auditory neurons makes it not viable to be used as a stand-alone solution. A possibility is suggested in Goldwyn et al., 2010, focused pTP configurations could be used in regions of the cochlea with the most surviving SG neurons, and the MP mode could be used in the vicinity of a neural dead region. For such a multiple configuration, clinicians should be able to evaluate the electrode-neuron interfaces at each channel.

- **Type of stimulation:** There are two types of stimulation, depending on how the stimuli information is presented to the electrodes. The **analog stimulation** presents this information as an electrical analog signal of the acoustic waveform. As an example, in some multi-channels implants, the acoustic signal is filtered by a bank of band pass filters, and the filtered waveforms are the stimuli signals delivered to the electrodes. The disadvantage of this kind of stimulation is the channel interactions due to simultaneous neighbor channel stimulation. This drawback helped the development of a new type of stimulation, the **pulsatile stimulation**. In this case, the stimuli information is presented to the electrodes as a set of narrow pulses. The amplitude of such pulses usually depends on the envelope of the filtered signal. The pulse rate of the stimulation has an influence on the speech recognition performance. Higher pulse rate stimulation achieves better speech recognition than lower pulse rates.

* **Transmission Link:** this is the way the external processor transmits the stimuli information to the electrodes. There are two systems to transmit the signal (see Figure 9):

    a. Through a transcutaneous connection: In this case a RF Link is used. The external processor encodes the stimuli information signal and it is sent as a RF signal to the receiver. The receiver is usually implanted in a flattened or recessed portion of skull (see Figure 8). The transmitting coil is held in place by a magnet. The advantage of such a system is that the skin in the scalp is closed after the operation avoiding possible infections. On the other hand, the disadvantages are that the implanted receiver may fail and a new surgery would be required. Another disadvantage is the presence of magnetic material in the skull, which is incompatible with MRI scanners.

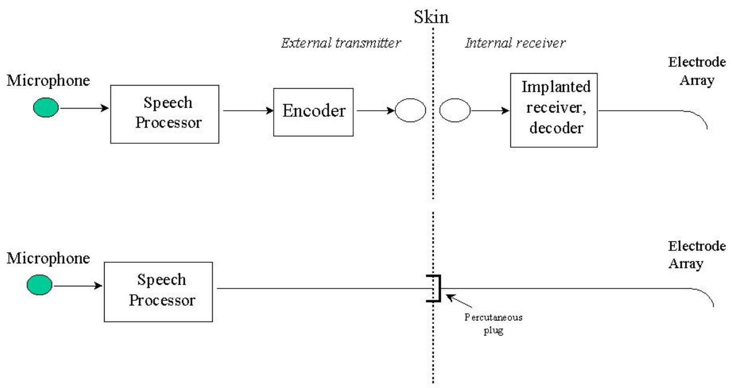


Figure 9 Types of Transmission Link. The first one is an example of a transcutaneous connection and the second one of a percutaneous connection. Picture from Loizou, 1998.

b. Through a percutaneous connection: in this kind of connection, the stimuli information is transmitted to the electrodes directly through plug connections. There are in this case no other implanted electronics than the electrodes. The most important advantage in this link is its flexibility and transparency, which made it ideal for research purposes. However it was shown that using devices with this type of connection infections happened more frequently. It was also observed that the lead wire press on and made groove in the skull. These and some other problems (detailed in Graeme, 2003), made this type of link to disappear in the today used Cochlear Implants.

- **Processing Strategy:** The most important difference among cochlear implants is how the devices transform the wave sound signal into a set of stimuli that can be interpreted by the nervous system, expecting to reach the same interpretation as a healthy ear does. The research and development in the last 30 years of such strategies have achieved very good improvement in the performance of the cochlear implants. As briefly mentioned in the introduction, the development from one-channel to multiple-channel devices improves the neuroprosthesis in such a way that some of the patients could score in test of intelligibility the same results as healthy listeners. The processing strategies are the key elements of cochlear implants and also the main topic of this work. A shortly description of the most important strategies in the history of these prosthesis, the Compressed Analog (CA) and the CIS (Continuous Interleaved Sampling) strategy, will be done. The newest strategies (n-of-m, SPEAK, ACE and HiRes) and the strategy that has to be evaluated, the Simulated Phase-Locked Stimulation (SPLS) will also be mentioned. All these processing strategies will be described in detail in the next chapter.

a. Compress Analog: in this case, the input signal was usually passed to 4 band pass filters and their analogue output waveforms were presented simultaneously to all electrodes. In such a strategy the patients received much more information that couldn't perceive, like temporary details in the stimulation signal. The benefits and drawbacks of the method will be discussed later.

b. Continuous Interleaved Sampling: this strategy developed by Wilson, Finley, Lawson et al. [Wilson et al., 1991], presents short pulses to each electrode in a non overlapping sequence. The amplitudes of such pulses are the envelopes of the band pass filter outputs. The pulses that are used for the stimulation are biphasic balanced pulses. The poor speech recognition in noisy environment as a consequence of the absence of fine structure on the stimuli information is one of the most important weaknesses of this important strategy.

c. n-of-m, Spectral Peak (SPEAK) and advanced combination encoder (ACE): All these strategies are based on the CIS Strategy. The difference to the original method is that there is a channel selection based on the amplitude of the envelopes. The stimulus pulses are only delivered to the electrodes that correspond to the channels with the highest amplitudes of the envelopes.

d. HiResolution® (HiRes) and HiRes Fidelity 120 : The HiRes strategy is also based on the CIS strategy, but with relative higher rates for the stimulation. The HiRes Fidelity 120 is based on the concept of virtual channels, which will be explained in detail in next chapter. This strategy is able to create 120 stimulus sites to achieve a better performance in speech and music perception.

e. Simulated Phase-Locked Stimulation (SPLS): this strategy presented by Chen, Wu, Li and Chi [Chen et al., 2009], was presented in 2009 and is still being evaluated. Like for the CIS Strategy, it also uses a filter bank strategy, with an envelope extraction. The decision about the firing time for the pulse trains will be done through the extraction of the phase information of the filtered band limited signals. Conserving this phase information and using it for the stimulation of the electrodes, should improve the speech recognition, in special for tonal languages like Chinese. The problems that can appear due to simultaneous stimulation of neighbor channels is also one of the topics of this work.

Chapter 2: Speech Processing Strategies

The Speech Processing Strategy is, as remarked in the first chapter, one of the key elements of a cochlear system. Its goal is to transform the input signal from the microphone into a set of stimuli that can be interpreted by the nervous system. One approach for this transformation, and a very successful one, is to reproduce the damaged or missing physiological functions that have to be bypassed by the neuroprosthesis.

But before the most important developed strategies are described in detail, a brief overview about speech signal should be done. What kind of information is contained in the speech signal and which is perceptually important should be discussed in order to understand the goal and motivation of the different speech processing strategies.

Speech Signal

The different information contained in the speech signal should be here shortly analyzed. The most important knowledge needed for the application in the cochlear implants is what kind of information of the speech signal has to be preserved in order to keep the speech intelligible, which means that the patient still understands what is being said.

Acoustic speech can be considered the result from a combination of a sound energy (e.g. the larynx) modulated by a transfer function (filter), which is determined by the shape of the supralaryngeal vocal tract [Fant, 1970]. This model is referenced as the "source-filter theory of speech production", which comes from the experiments of Johannes Müller (1848). Depending on the type of input excitation (source), two types of speech sounds can be produce, voiced or unvoiced. In the case of voiced sounds, the larynx serves as the source of sound energy and it generates quasiperiodic sounds like /a/, /i/, etc. by vibration of the vocal cords. The frequency of the input excitation by voiced sounds is usually called the fundamental frequency (F0). In case of unvoiced sounds like /s/, /t/, f/, etc. the input excitation is noise. This noise is generated either by forcing air through a narrow constriction (like for the sound /f/) or by building air pressure behind an obstruction and then releasing it suddenly (like for the sound /t/).

The supralaryngeal vocal tract, which consists of both oral and nasal airways, acts as a time-varying acoustic filter. This filter attenuates the passage of the sound energy at certain frequencies while allowing the passage of others. The formants are the frequencies that are allowed by the vocal tract to pass with the maximal energy. These

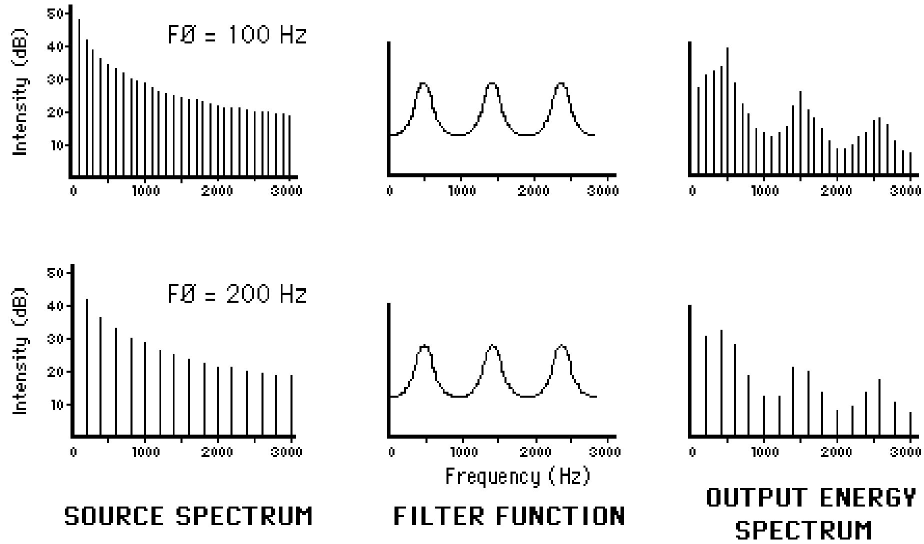


Figure 10 The source spectrum represents the spectrum of typical voiced sound source signal with a fundamental frequency of 100 Hz (top Spectrum) and of 200 Hz (bottom Spectrum). The filter function is in this case for an idealized neutral vowel /´/, with formant frequencies at approximately 500 Hz, 1500 Hz and 2500 Hz. The output energy spectrum shown at the right side is the result of the filter function shown in the center excited by the source spectrum shown at the left. Figure from haskins yale education website [see website link]

are determined by the overall shape, the length and the volume of the vocal tract. They can also be seen as the resonances of the vocal tract. These two concepts, fundamental frequency and formants are depicted in the Figure 10. Usually, the frequency of the first three formants (often called F1, F2 and F3) contain sufficient information for recognition of both vowels and other voiced sounds. Formant movements have also been found extremely important for the perception of unvoiced sounds [Border et al., 1994; Cooper et al., 1952]. Another term that is important in speech processing is pitch. Pitch is the relative perception of the fundamental frequency. Fundamental frequency (F0) describes a physical phenomenon, while pitch describes a perceptual phenomenon.

There are also two other important concepts used in signal processing that should be clarified: the fines structure information (FS) and the envelope of a signal. These two concepts are based on the work of the mathematician David Hilbert, who showed the relationships between the real and the imaginary parts of complex functions, known as the Hilbert Transform Relationships [Oppenheim et al., 1989]. An application of his work is that a band limited signal has the form of a sinusoid that is both amplitude and phase modulated. The sequence that modulates the amplitude is called the envelope and the carrier signal, whose instantaneous phase is also a function of the time, is called the fine structure information.

$$x(t) = A(t)\cos(\phi(t))$$

$A(t)$ is the envelope, $\phi(t)$ the instantaneous frequency and $\cos(\phi(t))$ the Fine Structure information. An example of this decomposition is shown in Figure 11.

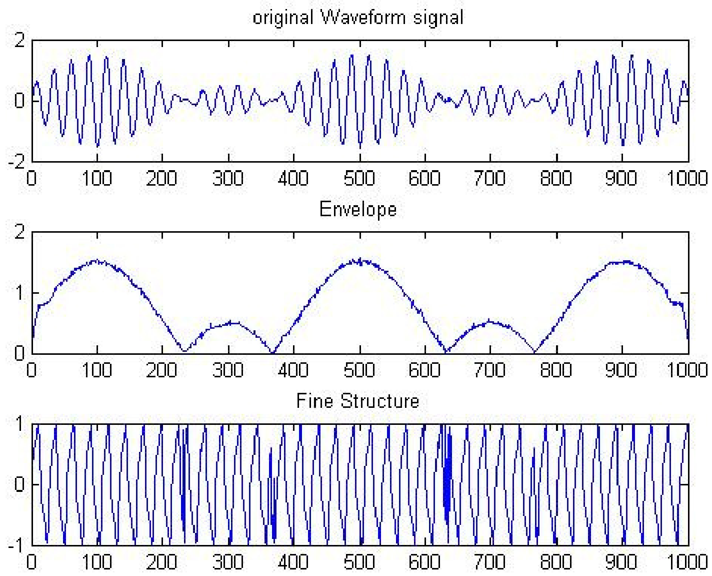


Figure 11 Decomposition of signal using the Hilbert Transformation.

Zachary Smith and coworkers at the Massachusetts Institute of Technology in Boston, Massachusetts, have proven important facts about the importance of the envelope and FS information for speech reception, melody reception and sound lateralization [Smith et al., 2002]. They created 'auditory chimeras', which are synthesized sounds using the envelope of one sound and the fine structure information of another sound. They found out that after hearing an auditory chimera, the correctly identified sound by the listeners with healthy hearing depended on the type or types of sounds used in each pairing and on the number of processing channels (bands). Here a summary of their conclusions:

- Speech was identified by its **envelope information** for **eight** or **more channels**, whereas the **FS information** was more important for **one** or **two channels**.
- Both envelope and FS information contributed to sentence recognition or intermediate number of channels.
- **Melodies** were recognized almost exclusively by their **FS information** for **up to 32 channels**. **Envelope** Cues became dominant at **48 and 64 channels.**
- **Lateralization**[1] of sentences was **difficult** for subjects with **small number of channels**. It improved with increasing number of channels. Lateralization was cued by the **FS information** in all cases.

The summary of their investigations indicates the importance of the FS information for speech reception using no more than 8 processing channels, for music reception and for lateralization of the sounds. It also reveals a possible acoustic basis for the hypothesized "what" and "where" pathway in the auditory cortex. For more information about the 'auditory chimeras', visit the link of the researchers http://research.meei.harvard.edu/chimera/index.html

After this brief description of the important definitions used in speech analysis and synthesis, the most important speech processing strategies will be exposed. In the next subsection a detailed explanation of the most important speech processing strategies that use waveform representation will be presented. Other approaches based on speech features extractions were developed in the middle of the 80's until begin of the 90's. They estimate different features from speech, such as F0 or the formants (F1 and F2) to present stimulation on the electrodes. The calculation of F1 and F2 was used to make the selection of .the electrode to stimulate, and F0 was used as the stimulation rate for voiced segments, while for unvoiced segments an average rate of 100 pps was used. One of the latest strategies based on speech features extraction was the MPEAK. This approach stimulates four electrodes at a rate of F0 pps for voiced sounds, like the F0/F2 and F0/F1/F2 strategies, and at quasi-random

[1] Lateralization means the localization of a sound as near the right or the left ear.

intervals with an average rate of pps for unvoiced sounds. Additional high-frequency information was extracted, using envelope detectors, from the higher frequency bands. This additional information improved the performance in consonant identification and on open-set sentence recognition compared to the first F0/F2 and F0/F1/F2 strategies. However, one major limitation of these strategies was that they tend to make errors in formant extraction, especially in situations where the speech signal is embedded in noise. These types of processing strategy have not been developed further. That is the reason why we have concentrate our efforts in describing the waveforms representation strategies.

Compressed Analogue (CA)

The compressed analogue approach was the first processing strategy developed for multichannel cochlear implants. It was first used in the Ineraid device manufactured by Symbion Inc, Utah and also in the UCSF/Storz, which is already discontinued. In the CA strategy, the signal coming from the microphone with a very high dynamic range is first compressed into a narrow dynamic range for electrically evoked hearing using an Automatic Gain Control (AGC).

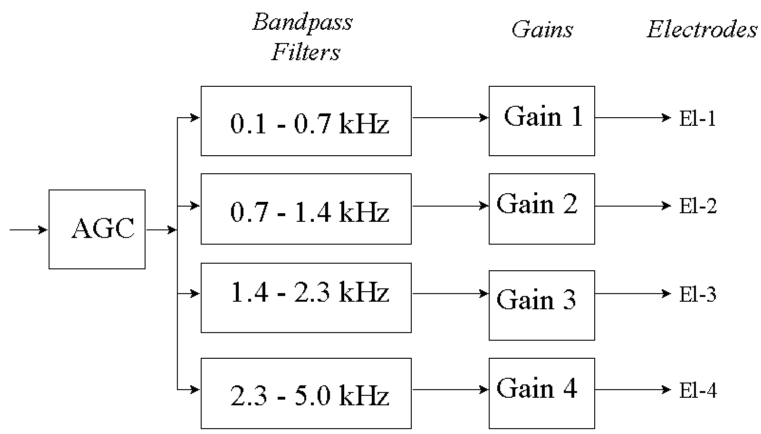


Figure 12 Block diagram of the Continuous Analogue Approach

The output signal of the AGC is filtered using four contiguous band pass filter, which cover the bandwidth from 100 Hz to 5 kHz. The filtered waveforms go through adjustable gain controls and then presented to the corresponding electrode for stimulation. In Figure 12 a block diagram of the CA strategy is presented.

This way of presenting the stimuli in the electrodes has following concerns. First of all, only part of the presented stimuli information can be really perceived by the patients. It is known for example, that most of the patients are not able to perceive changes in the stimulus waveform above 400 Hz (called the pitch saturation limit).

That means that most of the temporal information contained in the stimulus signals is not accessible for most of the patients. The second problem related to this strategy is the fact that simultaneous presentation of stimuli produces significant interactions among channels due to vector summation of the electrical field from each electrode. That means, the neural response to the stimuli from one electrode will be significantly distorted, or even counteracted, by coincident stimuli from other electrodes [White et al., 1984].

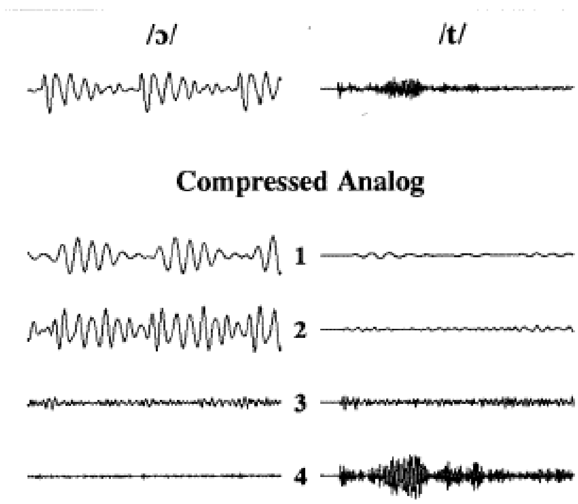


Figure 13 Waveforms produced by simplified CA implementations. The top signals are the preemphasized speech inputs. The left top signal corresponds to a voiced speech signal "aw" and the right one to the unvoiced speech signal "t". The signals below show the stimuli waveforms for the CA strategy. The waveforms are numbered by channels, where channel 1 being the stimuli signal applied to the most apical electrode.

The CA approach, as the first multichannel processing strategy, was very successful because it enabled many patients for the first time to obtain open-speech understanding. A study presented by Dorman et al. [Dorman et al., 1989], showed that for a group of 50 Patients using the Ineraid device, 45% of the patients could identify correctly words in CID sentences, 14% identified correctly monosyllabic words, and 14% of them made the right identification when hearing spondee (two-syllabel words). For that time (end of the 80's) it was the best speech recognition performance ever had.

Continuous Interleaved Sampling (CIS)

The CIS approach was developed by researches at the Research Triangle Institute (RTI) in North Carolina. The motivation of the new approach was to avoid the channel interaction due to simultaneous stimulation using interleaved nonsimultaneous stimuli. Trains of balanced biphasic pulses are delivered to each electrode in a way that only one electrode is stimulated at a time and avoiding any overlaps across the channels. The amplitudes of the pulses are calculated from the envelopes of band pass filter outputs. The CIS Strategy uses a preemphasis filter to attenuate the strong low-frequency components in speech that might mask important high frequency components (in Figure 14 pre-emph. Block). The preemphasis filter is followed usually by five or six channels, although in the figure the block diagram is depicted for an arbitrary number of channels "n". Between 4 and 22 channels (and corresponding stimulus sites) have been used in CIS implementations up to date. Each channel of processing includes stages of band pass filtering (BPF), envelope

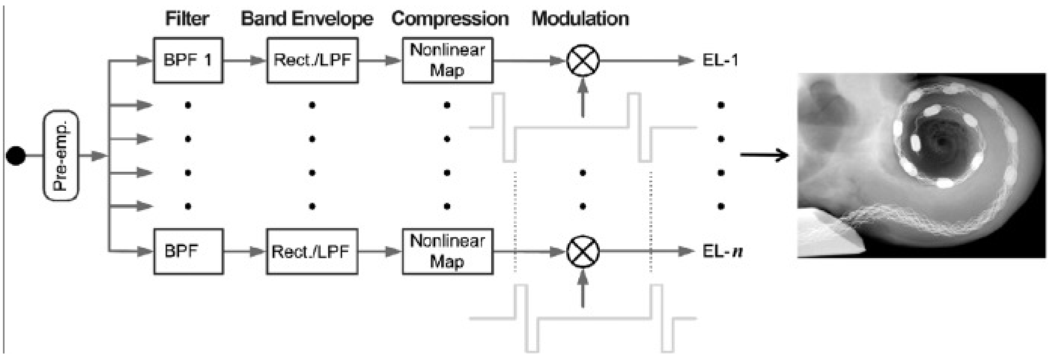


Figure 14 Block diagram of the CIS Strategy

detection, compression and modulation. The envelope detector consists of a rectifier (Rect.) and a low pass filter (LPF) with cutoff frequencies typically between 200 and 400 Hz. Finally, the amplitude of each stimulus is calculated by a logarithmic or power-law transformation of the corresponding channel's envelope at that time. This transformation is done to ensure that the envelopes outputs fit the patient's dynamic range of electrically evoked hearing of that channel. The rate at which the balanced biphasic pulses are delivered to the electrodes is constant and it has been shown to have a major impact on speech recognition.

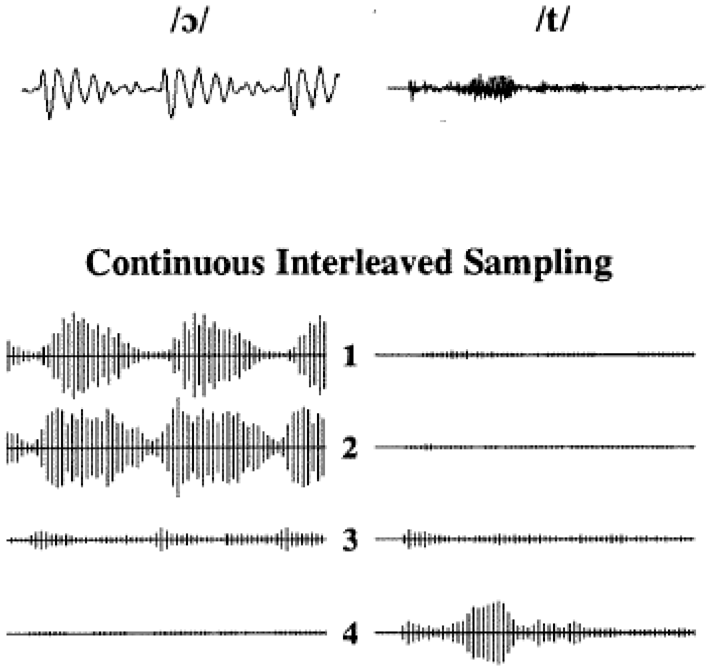


Figure 15 Waveforms produced by simplified CIS implementations. The top signals are the preemphasized speech inputs. The left top signal corresponds to a voiced speech signal "aw" and the right one to the unvoiced speech signal "t". The signals below show the stimuli waveforms for the CIS strategy. The waveforms are numbered by channels, where channel 1 being the stimuli signal applied to the most apical electrode.

There are some parameters in the CIS Strategy that can be varied to obtain a better speech recognition performance for the patients. These parameters are:

1. **Pulse rate and pulse duration**: The pulse rate is defined as the number of pulses per sec (pps) delivered to each electrode. The pulse duration and pulse rate concepts are depicted in Figure 17. Pulse rates as low as 100 pulses/sec and as high as 2500 pulses/sec have been used. The "optimal" pulse rate, as far as speech recognition performance is concerned, varies from patient to patient. Wilson et al. [Wilson et al., 1995] reported that some patients obtain a maximum performance on the 16-consonant recognition task with a pulse

rate of 833 pulses/sec and a pulse duration of 33 μ sec/phase. Other patients obtain small but significant increases in performance as the pulse rate increases from 833 pps to 1365 pps, and from 1365 pps to 2525 pps, using 33 μ sec/phase pulses. It would be expected that better recognition performance would be obtained with higher pulse-rates, since high pulse-rate stimulation can better represent fine temporal variations. However, this was not found to be true for all patients, at least over this tested range of pulse rates.

2. **Stimulation order:** Stimulation order means the order in which the electrodes are stimulated. One possibility is first to stimulate electrode 1, then 2 and so on. In this way, the part of the cochlea that response to low frequencies (apex) is stimulated first and the part that responds to higher frequencies (base) is the last one being stimulated. This apex-to-base order however doesn't maximize the spatial separation between sequentially stimulated electrodes. If the goal is to minimize possible interactions between channels, an alternative way of stimulation order should be chosen. If the electrodes are stimulated in a so called "staggered" order: i.e., in the order six-three-five-two-four-one, the spatial separation between the stimulated electrode is maximized and channel interactions can be minimized. The preferences of stimulation order vary from patient to patient. Some of them prefer the apex-to-base stimulation, because for them the speech sounds more natural.

3. **Compress Function:** The compression of envelope outputs plays an important role in the CIS Strategy. It transforms acoustical amplitudes into electrical ones. This transformation is necessary because the range of the acoustical waveforms is much larger than the implant device's dynamic range. Dynamic range is here defined as the range in electrical amplitude between threshold level (barely audible level) and loudness uncomfortable level. In normal speech, the acoustic amplitudes may vary over a range of 30dB.

However, CI patients may have for various reasons only a 5 dB dynamic range. Because of this, a non-linear compression function has to be used, to fit the acoustical amplitudes into the patient's electrical dynamic ranges.

There are several compress functions being used. One of them is the logarithmic compression function, which has following form:

$$y = A\log(x) + B$$

where x is the acoustic amplitude, A and B are constants and y is the compressed electrical amplitude. Another type of compression functions being used are power-law function, which has following form:

$$y = Ax^p + B \text{ with } p < 1$$

where x is the acoustic amplitude, A and B are constants and y is the compressed electrical amplitude. The advantage of the power-law functions is that the shape (steepness) of the compression can be easily modified by changing the values of the exponent p .The constants A and B are chosen in a way that the input acoustic range $[x_{\min}, x_{\max}]$ is mapped to the electrical dynamic range [THR,MCL], where THR is the threshold level and MCL is the most comfortable level. The values of THR and MCL may vary from electrode to electrode. Such a compress function is depicted in Figure 16.

The values of A and B can be calculated as follows:

$$A = \frac{MCL - THR}{x_{\max}^p - x_{\min}^p} \quad \text{and} \quad B = THR - A^{x_{\min}^p}$$

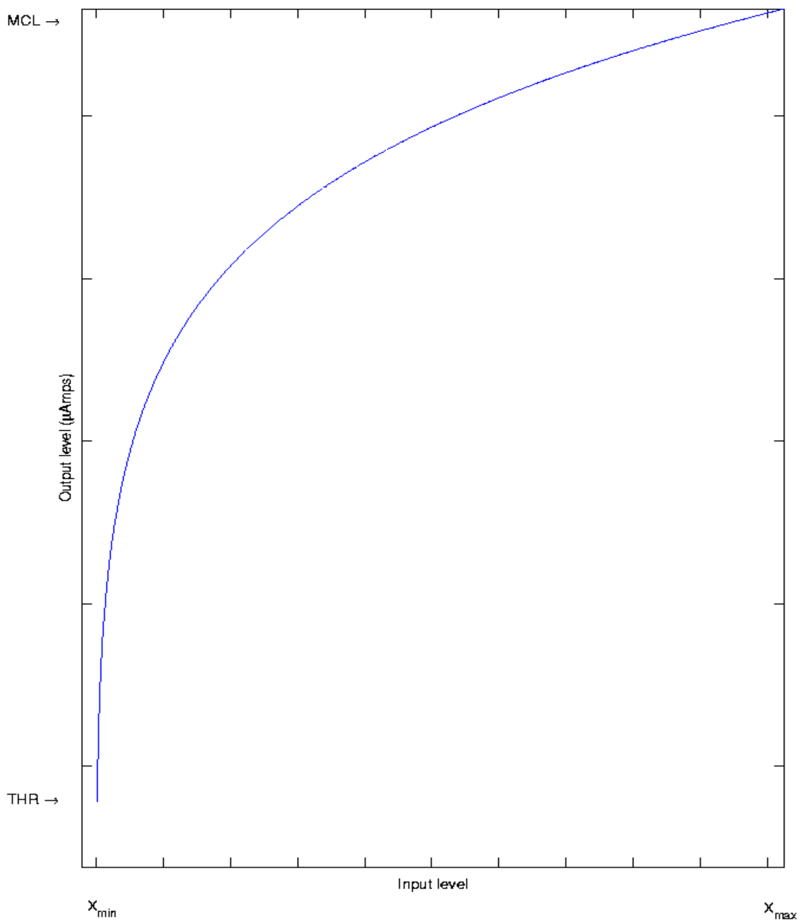


Figure 16 Power law compression function.

The key feature of this new strategy was its relatively high rate of stimulation on each channel. Other pulsatile strategies present sequences of interleaved pulses across electrodes at a rate equal to the estimated fundamental frequency (F0) during voiced speech and at a jittered or fix higher rate during unvoiced speech [Clark et al., 1987, Wilson et al., 1993 II, Wilson et al., 1991 II]. Rates of stimulation of the electrodes usually didn't exceed the 300 pulses per second (p.p.s). In contrast, the CIS strategy generally uses brief pulses and minimal delay, so that rapid variation on speech can be tracked by pulse amplitude variation. The rate of stimulation on each channel usually exceeds 800 pps and is constant, independently of the presence of voiced or unvoiced speech. Because of the use of such a constant high stimulation rate, it is allowed to design the low pass filters for the envelopes detectors with a higher cutoff frequency. For example, if the stimulus rate of 800 pps is used, the low pass filters cutoff frequencies can approach (but not exceed) 400 Hz without introducing aliasing effects in the sampling envelope signal. For a detailed display of how the stimulation signal looks like see Figure 17.

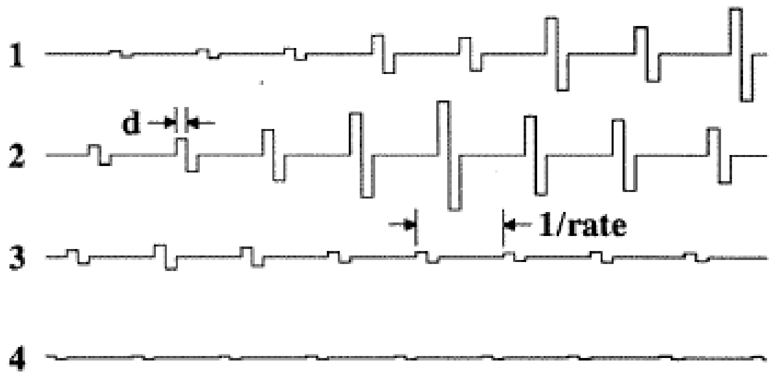


Figure 17 Extended display of the CIS Stimulation waveforms. Pulse duration "d" of the stimulation and period between pulses on each channel "1/rate" is shown. Picture from Wilson et al., 1993.

Several studies [Wilson et al., 1995, Wilson et al., 1991, Boex et al., 1994] were done by RTI and other institutions, comparing the performance between the CA and CIS strategies. In the case of Wilson et al., 1991, the comparisons tests included open-set recognition of words, sentences and paragraphs material. Recognition of words and sentences was evaluated with four tests from the Minimal Auditory Capabilities (MAC) Battery, and recognition of paragraph material was evaluated using connected discourse tracking. The MAC tests included recognition of 50 one-syllable words from Northwestern University Auditory Test 6 (NU-6), 25 two-syllable words (spondees), 100 key words in the Central Institute of the Deaf (CID) sentences of everyday speech, and the final Speech Perception in Noise (SPIN), however no noise was present in the test. For more details of the test development refer to Wilson et al., 1991. The test results are shown in Figure 18. It can be seen that for every test, every subject scored higher or repeated a 100% score using the CIS approach.

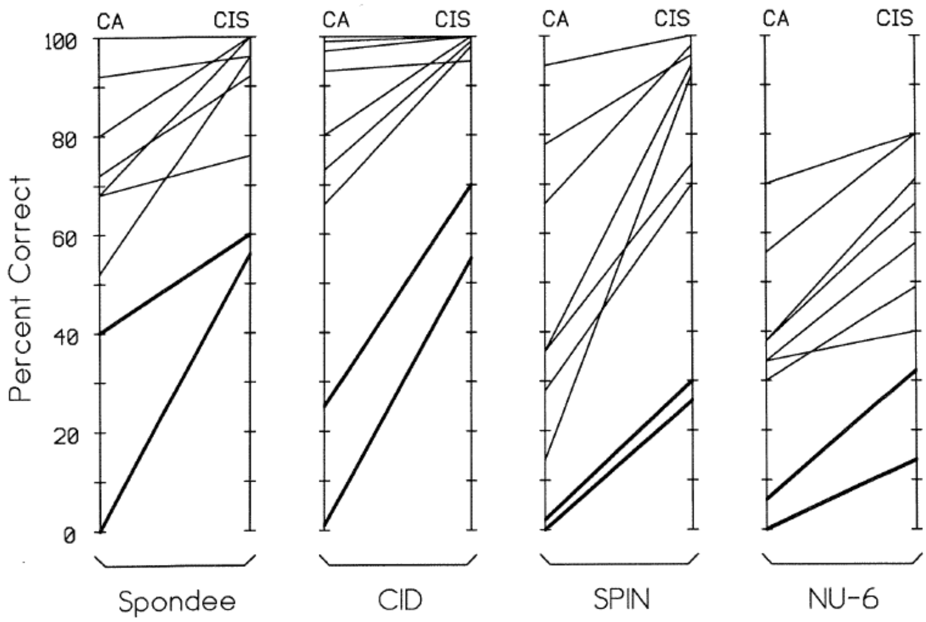


Figure 18 Speech recognition scores for CA and CIS strategies. The lines connect the CA and CIS scores for each subject. Results and picture from Wilson et al., 1991.

Several factors could be responsible for the success of the CIS over the CA approach:

1. Reduction of channel interaction due to the use of non simultaneous stimuli.

2. Use of five or six channels rather than four

3. Representation of rapid envelope variations through the use of high pulse rate for the stimulation of electrodes.

As it has been shown, in this processing strategy, envelope signals are derived by rectifying and filtering the outputs of band pass filters. Those signals are used to determine the patterns of stimulation at the electrode array. However, the division in envelope and FS is here not as clearly defined as in the work of Smith et al., 2002. Although only envelope information is presented to the electrodes, some of the fine structure, in special frequencies included in the envelope that range from 200 to 400 Hz, is however presented. That means, substantial FS information is anyway available and may be partially perceived in this low-frequency range. How much FS information is presented in this envelope base strategy is not well defined. The possibility that only a very small quantity of such information is presented in the stimulation pattern, along the findings of Smith et al., 2002, that demonstrated the importance of the FS in speech perception, speech lateralization and music perception, have motivated the development of other strategies to be able to make this information present in the stimulation patterns, so that better speech and music perception can be achieved by the cochlear patients.

n-of-*m*, ACE and SPEAK Strategies

New strategies based on the CIS strategies, have been developed and have also obtained outstanding results. Among these are the *n*-of-*m*, the advanced combination encoder (ACE) and the spectral peak (SPEAK). The *n*-of-*m* strategy was first described in 1988 by Wilson et al. These three strategies have a similar block diagram as the CIS approach, but they use a channel selection scheme, in which the envelopes signal for the different channels are first scanned prior to each frame of stimulation across the intracochlear electrodes to identify the signals with the *n*-highest amplitude from among *m* processing channels. Stimulus pulses are delivered only to the electrodes that correspond to the channels with the highest amplitudes. 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. Another important difference between the *n*-of-*m* and ACE strategy and the SPEAK processing is the stimulation rates. They typically approximate or exceed the 1000pps in each electrode, in the first two ones (the *n*-of-*m* and the ACE) and it reaches only 250 pps pro selected electrode in the second one (SPEAK).

In the implementation of the *n*-of-*m* strategy, *m* is usually a high value. In the ACE approach, for example, *m* may be as high as 22. A block diagram of a possible *n*-of-*m* implementation is shown in Figure 19.

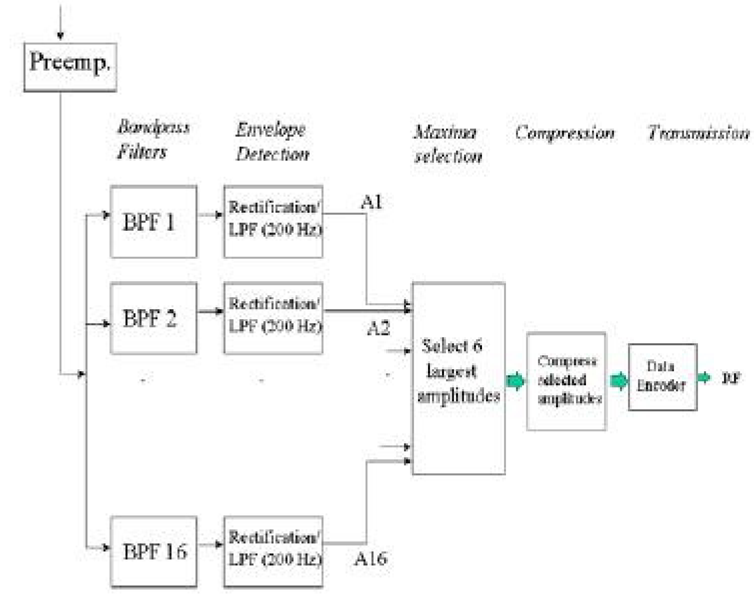


Figure 19 Block diagram of a 6-of-16 Strategy. From http://www.yahyaozturk.com/pages/images/stories/image8.jpg

In a *n*-of-*m* implementation, frequencies for the band pass filters are distributed in a linear-logarithmic map way. That means, that for the frequencies lower than 1000 Hz the band pass filters are distributed along a linear scale. On the other way, for the frequencies higher than 1000 Hz, the band pass filters are distributed along a logarithmic scale. This allocation follows the MEL scale in normal hearing [Zwiker et al., 1990]. With such a frequency allocation, the specification of very narrow bandwidths for the analysis channels with the lowest center frequency is avoided. It also increases the resolution of the filters with intermediate center frequencies, in the region of 500-1500 Hz, by lowering their bandwidths. This mapping may be better than the logarithmic only, for the reasons just mentioned. However, the mapping of the frequencies in the human cochlear described in Greenwood, 1990 and the widths and frequency coordinates of 'critical bandwidths" or "equivalent rectangular bandwidth" [Glasberg et al.,1990] in normal hearing, approximate a logarithmic function, at least for the frequencies above 300 Hz. That means that using a purely logarithmic function achieves a closer imitation of the normal tonotopic pattern, even for a high number of analysis channels.

In the ACE strategy, a linear frequency distribution is used up to 1300 Hz and a logarithmic distribution from that point up to the maximal frequency. For the typical implementations of the ACE processors, *m* ranges from 20 to 22 channels and *n* from 6 to 16.

In acoustic simulations, of *n*-of-20 processors, Dorman et al. [Dorman et al., 2002] reported that the required *n* to get asymptotic performance in speech recognition depends on the type of speech that has to be recognized. They became different results for vowels, consonants or words in sentences. Another parameter that influenced the obtained *n* was the presence of noise. For example, for their normal hearing subjects, an *n* of 9 was required to reach asymptotic performance for recognition of words presented in competition with noise (with a SNR of 0 dB). For the other tests (recognition of vowels, consonants, and so on) asymptotic performance was achieved with lower values of *n*. For all tests, a 10-of-10 implementation produced results that weren't statistically different from the 9-of-20 implementation. The conclusions of these tests are:

1. n should be at least 9 for a 9-of-20 processors.

2. 9 to 10 channels are sufficient for high levels of speech understanding.

Comparisons between *n*-of-*m* (or ACE) versus CIS have indicated roughly equivalent performances for the two strategies [Lawson et al., 1996, Ziese et al., 2000] or better performances for the *n*-of-*m* approach [Kiefer et al., 2001]. The results have varied from subject to subject and also among the different implementations of the strategies using different processors or laboratory hardware and software.

The element that makes these three strategies different to the CIS approach is the channel selection or "spectral peak picking" scheme. See the reference Wilson, 2006 to know more about it. It is designed in part to reduce the density of stimulation, while representing the most important acoustic aspects of the environment. The suppression of the low-level channels and their associated stimuli may reduce the overall level of masking or interference across electrodes and stimulus regions in the cochlea. With the assumption that the channel with low level amplitudes do not contain significant information, an 'unmasking effect" may improve the perception and understanding of the input signal by the patient. In cases where the patient is in a noisy environment with positive Speech to Noise ratios, the selection of the channels with the greatest amplitudes in each frame may emphasize the primary speech signal with respect to the environment noise.

As mention before, the SPEAK strategy uses an adaptive *n*-of-*m* approach, in which *n* may vary from one stimulus frame to the next. The input signal is processed with a bank of 20 band pass filters, with a linear distribution for the low frequencies up to about 1850 Hz, and a logarithmic distribution thereafter. Envelope signals are obtained as in the CIS, *n*-of-*m* and ACE strategies, with an envelope cut-off frequency of 200 Hz. The number of channels selected in each scan (the adaptive *n*) depends on:

1. number of envelope signal greater than a preset noise threshold
2. details of the input signal, such as the distributions of energy over the frequency.

In most of the cases 6 channels are selected. However the number of channels ranges from 1 to a maximum, which can be set as high as 10. Cycles of stimulation include the selected channels and their corresponding electrodes. The stimulation rate can range from 180 to 300 pps. The time required to complete each stimulation cycle varies from frame to frame. It depends on the number of selected channels (*n*) included in the cycle and the pulse amplitudes and duration for each of the electrodes. In general, when only few electrodes are included in a cycle, the stimulation rates for each electrode are relative high. Otherwise, when many electrodes are included, the rate will be reduced. Examples of electrical stimulation patterns for four different sounds using the SPEAK strategy are shown in Figure 20. Five maxima were selected for /s/, while ten maxima were selected for /a/. For sounds with a broader spectrum, a higher number of channels are selected and therefore the stimulated rate is slowed down. For sounds with a limited spectral content, fewer maxima are selected. The stimulation rate in such a case increases to provide more temporal information.

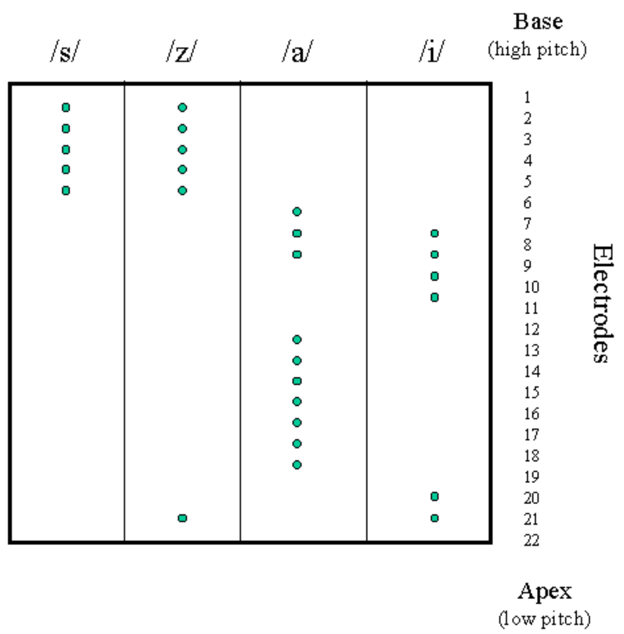


Figure 20 Electrical stimulation patterns for four different sounds using the SPEAK strategy from Loizou, 1998.

A problem that is common to all these approaches is that changes in the rate of stimulation for a given electrode are not perceived as differences in pitch above a "pitch saturation limit" of about 300 Hz for most of patients [Zeng, 2002]. In addition, the difference limens for frequency changes below 300 Hz are generally much worse form implant users than for normal hearers. Because of this, the representation of FS information using temporal code may be limited to 300 Hz or lower and may be highly degraded. Another problem that is present today is that the effective number of sites of stimulation along the length of the electrode array appears to be limited to about 4 to 8 for present placements and design of arrays. That means representation of FS information with a fine–grained adjustment in the site or sites of stimulation may be also highly limited.

HiRes and HIRes 120

The HiRes Strategy is implemented in the Auria and Harmony sound processors from Advanced Bionics LLC. An audio signal is sampled at 17400 Hz and preemphasized by the microphone and then digitalized. Adaptive gain control is also performed digitally using a dual-loop AGC [Stone et al., 1999]. The signal is then split into frequency bands using IIR sixth order Butterworth Filters. The center frequencies of the filters are logarithmically spaced between 350 Hz and 5500 Hz. The last filter is a high pass filter, which bandwidth extends up to the Nyquist frequency. The bandwidths covered by the filters are referred as subbands or frequency bands. Like in the CIS strategy, each frequency band will be associated with an electrode. The subband outputs of the filter bank are used to derive the information that will be sent to the electrodes, like in the methods described before. In this strategy, the filter outputs are half-rectified and averaged. Half rectification means that the negative amplitudes of the signal will be set to 0 at the output of each filter band. These outputs will be averaged over the duration of a stimulus cycle T_s. Finally a "Mapping" block transforms the acoustic values obtained for each frequency band into current amplitudes to modulate the biphasic pulses that will release each electrode. A logarithmic compression function is also used to ensure that the envelope output fit the patient's dynamic range. In each stimulation cycle, HiRes stimulates all M implant electrodes sequentially, to avoid possible channel interactions, as it is done in the discussed strategies (CIS, *n*-of-*m*, ACE and SPEAK). The number of electrodes in the HiRes90K is M=16 and all electrodes are stimulated at the same fixed rate. The maximum channel stimulation rate (CSR) used in the HiRes90K is 2899Hz [Nogueira et al., 2009].

Despite the problems pointed out at the end of the last subsection, new strategies have been developed to increase the transmission of FS information compared with the CIS, *n*-of-*m*, ACE and SPEAK approaches. One of this is a variation of the HiRes, called the HiRes with Fidelity 120™ option (HiRes 120). It is based on the use of "virtual channels" to increase the number of discernable sites beyond the number of physical electrodes. The concept of virtual channels was introduced by Wilson et al [Wilson et al, 1992; Wilson et al., 1993 III; Wilson et al., 1994 and Wilson et al., 1994 II] and has been investigated by others [Poroy et al., 2001; Litvak et al., 2003; Donaldson et al., 2005; Firszt et al., 2007 and Koch et al., 2007]. The term "current steering" is used in these reports to make reference to the same concept. The idea behind the construction of "virtual channels" is illustrated in Figure 21. With virtual channels, adjacent electrodes may be stimulated to shift the perceived pitch in any direction with respect to the percepts evoked by the stimulation of either of the electrodes alone. Results from different test done with implant subjects indicate that pitch can be manipulated using various choices of stimulation and single electrode conditions [Wilson et al., 1993 II]. If, for instance, the most apical electrode an array is stimulated alone (electrode 1, Figure 21 a), subjects reported a low pitch. By stimulation of the next electrode in the array (electrode 2, Figure 21 b) a higher pitch was perceived by the test subjects. An intermediate pitch could be produced by stimulating both electrodes together with the same in-phase pulse (Figure 21 c). It was also found that the pitch perceived by stimulation of a single electrode could also be shifted by presentation of an opposite-polarity pulse (generally smaller) to a neighboring electrode (Figure 21 d).

The possibility of producing perception of pitches others than those elicited by the stimulation of single electrodes showed that there is a way to provide additional distinguishable sites along (and beyond) the length of the electrode array. These additional sites may support a higher number of effective information channels for the implants compared to the implants with stimulation restricted to single electrodes only.

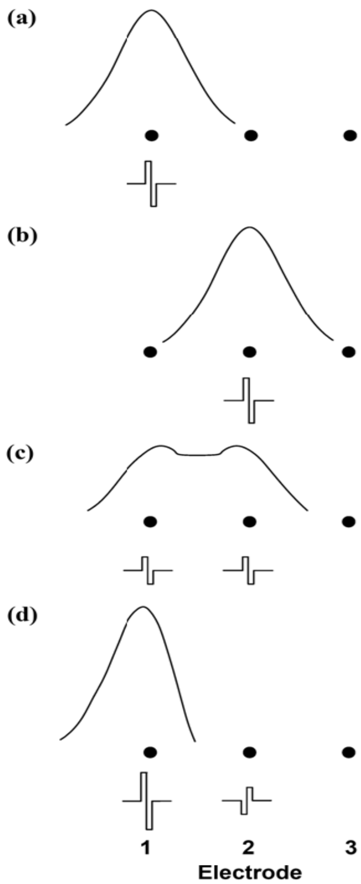


Figure 21 Schematic illustration of neural responses for various conditions of stimulation. In (a)-(b) single electrodes are stimulated. In (c)-(d) multiple electrodes are stimulated. The top curve in each panel is hypothetical sketch of neural responses, as function of position along length of cochlea for given condition of stimulation. Condition is indicated by pulse waveform(s) beneath one or more dots, which represent positions of three adjacent intracochlear electrodes. These different conditions of stimulation elicit distinct pitches for implant patients; see main text for discussion. From Wilson and Dorman 2008.

The concepts of virtual channels can be extended to include a quite high number of stimulus sites and corresponding pitches by using different ratios of currents delivered to two sequential electrodes simultaneously. A possibility is shown in Figure 22. By stimulating electrode 1 alone, stimulus site 1 is produced, while stimulus site 2 is elicited by stimulating simultaneously electrode 1 and 2, with a pulse amplitude of 75% for electrode 1 and 25% for electrode 2. For the stimulus site 3, the stimulation of both electrodes is also required, but in this case with a pulse amplitude of 50 % for both electrodes 1 and 2, and so on. In such a configuration, a total number of 21 stimulus sites and corresponding pitches might be produced for a good subject using only 6 intracochlear electrodes. Other possible arrangements may produce additional pitches. It is important to point out that it is not always possible to achieve such a high number of stimulus sites. This always depends on the patient's degree of ganglions cells survival. By some patients with a very low percentage of ganglion cells survival, only a number of discriminable pitches that was less than the number of physical electrodes were being able to be produced.

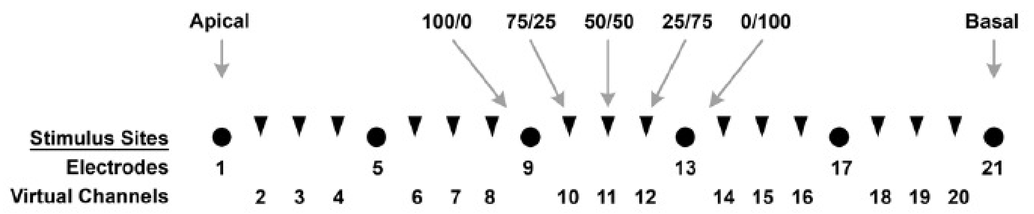


Figure 22 Diagram of stimulus sites used in virtual channel configurations. Filled circles represent sites of stimulation at each of 6 intracochlear electrodes. Inverted triangles present additional sites produced with simultaneous stimulation of adjacent electrodes at indicated ratios of pulse amplitudes for 2 electrodes. From Wilson and Dorman 2008.

In the HiRes 120 sound processing aim to create the optimal location for delivering each spectral band by precisely changing the proportion of current delivered simultaneously to adjacent electrodes in each electrode pair through the implementation of active current steering (virtual channels). There are 16 physical electrodes, which makes up to 15 electrode pairs. Each electrode pair has 8 spectral channels, so 120 stimulus sites can be created. In order to create the information for the 120 spectral subbands, the input signal is analyzed in grater detail than in the standard HiRes approach. First, the incoming signal is analyzed using a 256-bin FFT. After that, a detailed analysis of temporal and spectral information is processed simultaneously. The temporal details are extracted using a Hilbert transformation, while a spectral navigator searches and locates the spectral maximum for each electrode pair across the 120 spectral bands. The estimated frequency of the spectral maximum is used to calculate the rate of the stimulation pulse train and to continuously select the optimal location for delivering stimulation and therefore improving the fidelity of stimulation.

Stimulated Phase-Locking Stimulation (SPLS)

All studies done to evaluate the performance on speech perception of the CIS and the other strategies were done by CI users speaking monotonal languages, such as German or English. For tonal languages, like Chinese, CI patients reported a poor identification of vowels and consonants [Zeng, 1995, Zeng et al., 1998]. Chinese is a tonal language, with 4 tonal patterns. These tonal patterns are defined by the fundamental frequency F0 of voiced speech. In Figure 23, the acoustic spectra and the F0 contours of a set of monosyllabic Chinese syllables, chosen to contrast the four lexical tones, are shown. Changing the tone in the syllable from flat to rising, or to falling and rising, or to falling, changes the meaning of the word. Xu et al.[Xu et al., 2002], studied how parameters of the CIS strategy, such as the cutoff-frequency of envelope detector's low pass filter and the number of channels of the ban pass filter bank, affect tonal recognition. It was shown that these two parameters play an important role in the recognition of the 4 Mandarin tonal patterns.

It has already been mentioned the importance of the transmission of more fine structure information (FS) to achieve a better speech and music perception by CI patients. Motivated by the not so good performance of CIS by Chinese patients, Chen, Wu, Li and Chi presented in Chen et al., 2009 a novel strategy, called Simulated

Phase-Locking Stimulation. This new strategy preserves part of the phase information of the original speech. This should be useful to upgrade the function of a CI device by introducing phase-related modulation on the stimulation pulse intervals. A detailed block diagram is depicted in Figure 24.

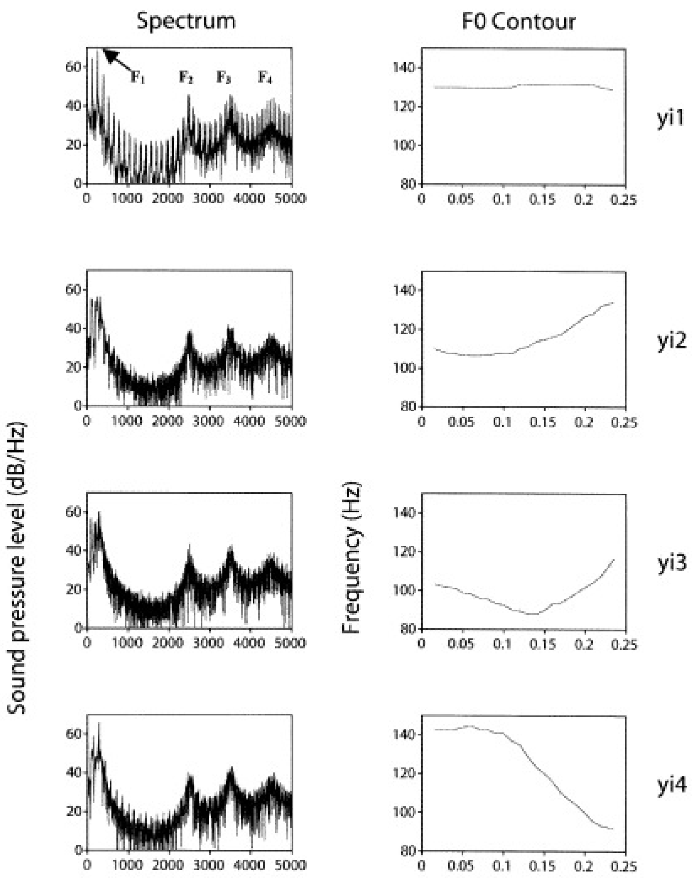


Figure 23 Acoustic spectra (left panels) and F0 contours of a set o monosyllabic Chinese syllables to contrast the four lexical tones (Pynin Roman phonemyc transcription): yi1 “clothing”, yi2 “aunt”, yi3 “chair ” and yi4 “easy” (right panels).From Krishnan et al., 2004.

The wave sound signal captured by the microphone is first pre-emphasized and then decomposed into multiple frequency bands by a filter bank of band-pass filters. To preserve the phase information of the input signal, the bank of band pass-filters should have the property of not distorting the phase. Chen et al. suggest the implementation of this stage using zero-phase filters. A zero-phase filter is a special case of a liner phase filter in which the phase slope is zero. Such filters have a real impulse response $h(n)$ that is even and also satisfy:

$$h(n) = h(-n) \ , \ n \in Z$$

A zero-phase filter cannot be causal (except in the trivial case when the filter is a constant scale factor). However, in many "off-line" applications or applications where a constant delay is permitted, such filters can be used. In the next Chapter zero-phase filters will be described with more detail.

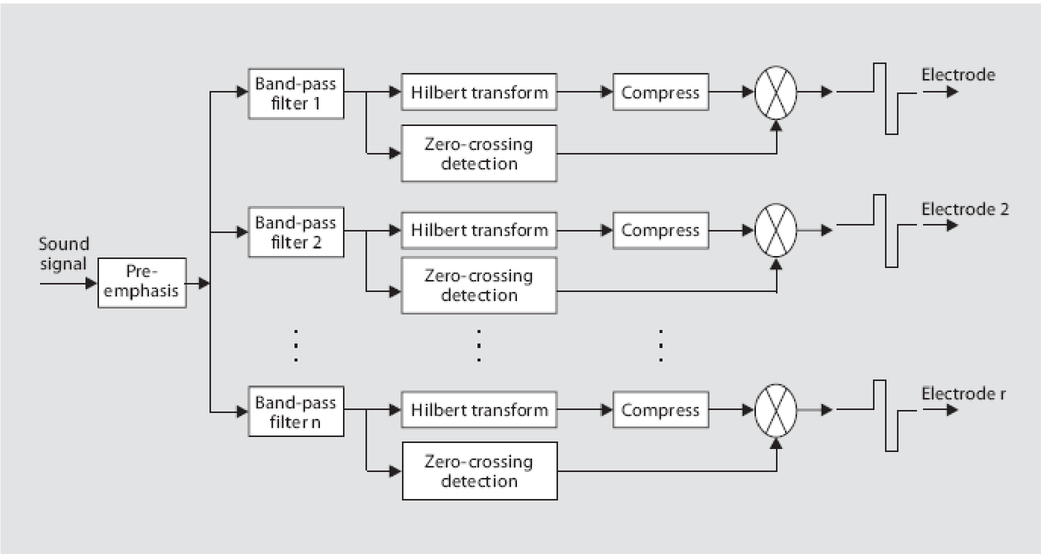


Figure 24 Block diagram of the SPLS approach. It shows how the phase information control the pulse rates. Figure from Chen et al., 2009.

After the filter bank stage, the signal in each band goes through two different signal paths:

1. Envelope extraction: to get the envelope information, the Hilbert transform is being used. The obtained envelope is then logarithmically compressed to an acceptable dynamic range for the CI user. This issue has already been discussed in detail in the CIS strategy. The extracted and compressed envelope will be used to modulate the

amplitude of the pulse trains that will be delivered to the electrode. This pathway is very similar to the one used in the CIS strategy. The only difference is the usage of the Hilbert transform instead of the full rectifier and LPF used in the CIS strategy.

2. Phase detection: to extract the phase information, a "zero-crossing" detector is proposed. It should record every zero-crossing time of the narrow-band signals of each band. This information will be required in order to decide the firing time of pulse trains.

The pulse-firing strategy of the SPLS is a new concept for the speech strategies for CI. It tries to emulate the neural mechanism of human hearing. In the human auditory system, the nerve firing patterns occur at roughly the same phase of the waveform each time. However, differences are also made between low and high frequencies. For the low frequencies, a single auditory nerve fires on every cycle of the stimulus tone. For high frequencies, they don't necessarily fire at every cycle. In the SPLS approach, it is proposed to produce electrical stimulation pulse on each channel when zero-phase events occur on the corresponding band-limited signal. A zero-phase event can be defined as two consecutive zero-crossings on the signal. For a given frequency, whose center frequency is below 1200 Hz, the corresponding electrode starts with a biphasic pulse at every zero-phase event detected in its associated band-pass filtered signal. However, for frequencies above 1200 Hz, the electrode starts with a pulse only once every $[f/1200]$ zero-phase events, where f is the center frequency of the corresponding band-pass, and $[.]$ means the smallest integer bigger than $f/1200$. The amplitude of the pulses are the values extracted in the envelope extraction stage.

Problems related to simultaneous stimulation across electrodes were already exposed at the CIS subsection. They were even part of the motivation for the CIS strategy. As the period between pulses of each channel in the CIS approach is fix, simultaneous firing can be avoided. However, the inflexibility of such firing pattern makes it very difficult to increase the FS information in the stimulus signal. That is the reason for the SPLS strategy, where the pulse rate of each channel is changed according to its phase information, and simultaneous firing between 2 adjacent and no adjacent channels will occur. Some measurements were done by Chen et al. and presented in Chen et al., 2009. They processed a 49-second piece of sound (including male or female English speech, Chinese speech, and a piece of music) with an 8 channels SPLS strategy. A final percentage of 1.9% of simultaneous firing between 2 adjacent channels was calculated. For them, this value is too small to provide additional inhibitory procedures for the strategy. Values to the general simultaneous stimulation (between not adjacent channels) were not presented in the study. This issue, the simultaneous stimulation between adjacent and not adjacent channels delivered by the SPLS strategy, should be evaluated in the simulations that will be presented in next Chapter.

This new strategy has not been yet implemented by any of the three main companies that produce cochlear implants. Real experimental data are until now not available to do an exhaustive evaluation of the method. For the first evaluation of the strategy, an examination of normal hearing listener's responses to acoustic simulation of the strategy was done. Previous studies have confirmed that such an examination is useful for evaluating speech processing strategies [Roggero et al., 1986]. The first psychophysical study has been presented by Chen et al. in Chen et al., 2009. They aim to compare the effectiveness of the CIS and the SPLS strategies, using normal listener's rates in hearing tests using acoustic simulations. For the acoustic simulations, Chinese speech was first processed through a high-pass filter (6dB/octave slope at 1200 Hz) implemented by a 1-order FIR filter to simulate the compensation of high frequencies done by a preemphasis filter. Then, the speech goes through a bank of 8 adjacent band-pass filters. All the band-pass filters were 1/3 octave, 4th order Butterworth filters, and the bandwidths were determined following the equivalent rectangular bandwidth model [Glasberg et al., 1990]. The center frequencies were equally spaced on a logarithmic scale from 215 Hz to 4891 Hz with the successive ratio factor of 1.25. The envelope of each band was extracted by the Hilbert transform, and was used to modulate the amplitude of the sine wave.

For the acoustic simulation of the CIS strategy, the sound signal could be described by the following formula:

$$y(t) = \sum_{i=1}^{N} A_i(t) \cdot \sin(2\pi f_i t + \phi_i), \quad (1)$$

where $y(t)$ is the synthesized signal, subscripts i means frequency band i and there are a total of N bands, $A_i(t)$ means the envelope of the band i, f_i means the center frequency of the band i and finally ϕ_i is the initial phase of the sine wave of the band i. In the work of Chen et al. $N = 8$ and $\phi_i = 0$.

For the SPLS strategy, the formula described is modified into the following:

$$y(t) = \sum_{i=1}^{N} A_i(t) \cdot \sin(2\pi f_i(t) \cdot t + \phi_i), \quad (2)$$

Now $f_i(t)$ reflects the firing rate of the stimulation pulses as a function of time:

$f_i(t) = \dfrac{1}{T_i(t)}$, where $T_i(t)$ is the interval between, what we just defined, two adjacent zero-phase events, or the interval between every second zero-crossing time. The results of the psychophysical experiment are presented in Figure 25 and Figure 26.

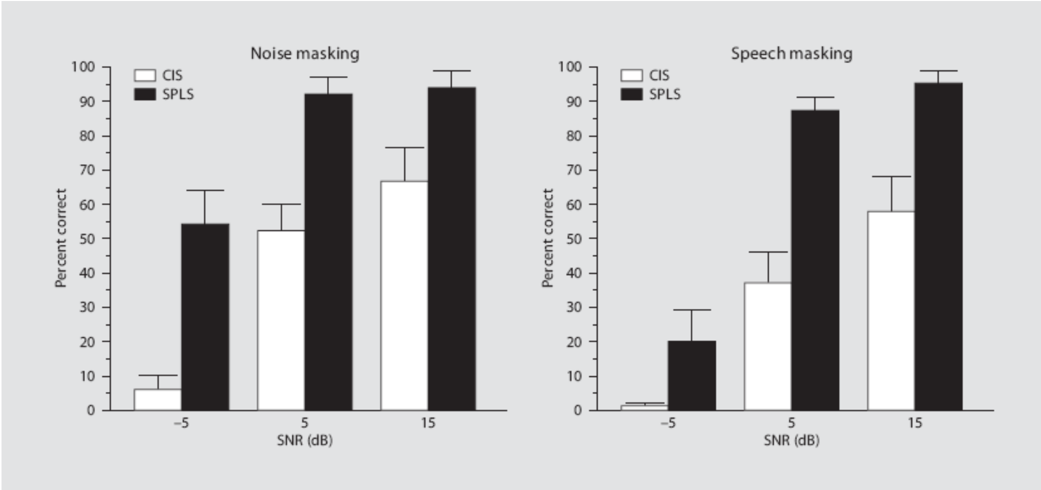


Figure 25 Mean percent of correct identification of key words across 12 subjects as a function of SNR for each of the 2 processing strategies under 2 different masking conditions: steady-spectrum-noise masking and speech masking. Error bars indicate the standard deviation of the mean. From Chen et al., 2009.

There were in the test 3 with-in variables: (1) Masker type (steady-noise, speech), (2) processing strategy (CIS or SPLS) and (3) SNR (-5 dB, 5 dB and 15 dB). Each subject had to hear a list of acoustic simulations of nonsense sentences, which have 3 key components: subject, predicate and object. These 3 key components were also the 3 key words with 2 syllables for each. For more information about the test procedure refer to Chen et al., 2009.

In Figure 25, the left diagram shows that the percentage of correct recognition of the speech processed by SPLS and by CIS changing the level of the SNR and with steady noise as the masking condition. The right diagram shows the same scores but under speech masking conditions. The percentage of correct recognition increases in both cases with the increase of SNR, but the scores for the SPLS processed speech is always much larger than the scores for the CIS processed speech.

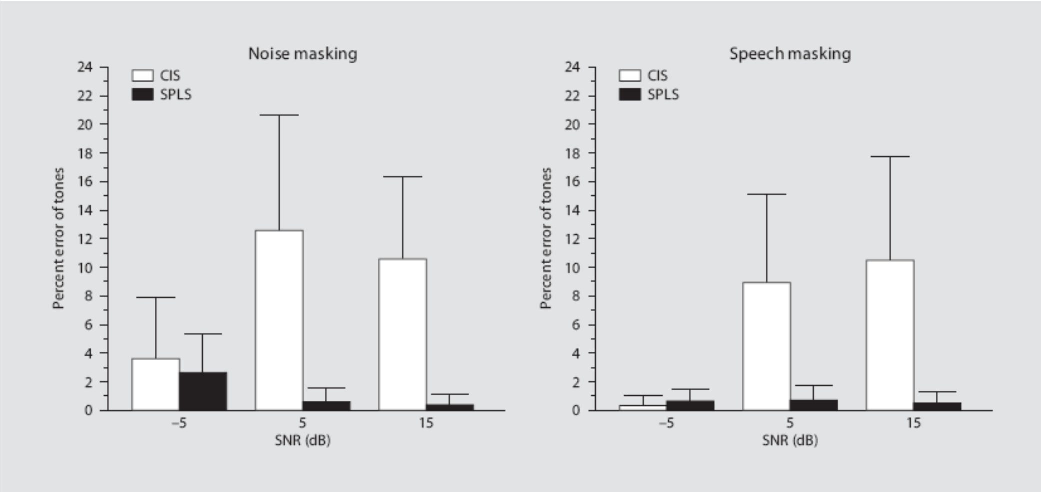


Figure 26 Mean percent error in recognition of tones across 12 subjects as a function of the SNR for each of the 2 processing strategies under 2 different masking conditions: steady-spectrum-noise masking and speech masking. Error bars indicate the standard deviation of the mean. From Chen et al., 2009.

In Figure 26, the results of the "tone error" in sentences repeating across the test subjects is depicted. This test was done in order to evaluate the benefits of the SPLS strategy to the recognition of tones, one of the main problems of the CI users who speak tonal languages and the motivation of this new strategy. Again the left diagram shows the results under steady-noise masking conditions. In this case, the percentage of error in tone recognition is always lower for the SPLS processed speech than that for the CIS processed speech. For very bad noisy conditions (SNR=-5) the difference with the CIS is not significant. An effect that remains unclear is the fact that for very bad conditions, the percentage of error in the tone recognition for the CIS processed speech is lower than for better noisy conditions. The right diagram shows the same scores, but under speech masking conditions. In this case the results obtained are the expected ones. The results show that SPLS strategy cerates acoustic simulations that are not affected by speech masking conditions independently of the SNR.

All the results that were obtained are very promising. The Chinese researches encourage in their paper teams from monotonal languages, like English or German, to keep doing psychophysical studies in their languages to get more simulation data, especially in cases with noise or speech masking environments. Prof. Rattay decided to take this challenge and give me the chance to research in this amazing field.

In the next chapter, a detail explanation of the simulations done in MATLAB will be described. Our intention was not only to get some sounds and compare them, but also to play with some important parameters of both strategies, such as number of channels, type of filters and bandwidths used in the band pass filter bank. Another important issue has to be addressed as well, the simultaneous stimulation. We also wanted to calculate how often this effect can appear in pieces of German Speech.

Chapter 3: Simulations in MATLAB

The goal of this chapter is to present for strategies, the CIS and the SPLS strategies, acoustic simulations varying different possible parameters such as the frequency range, the number of channels, the type of filter or the kind of the filter banks' bandwidth. The acoustic simulations were generated with MATLAB programs, written for this master thesis. The original wave sound is for all simulations the same. It consists on the syllable "tra". It was chosen because it has a consonant part at the beginning, and a vowel part at the end. This syllable is in many monotonal languages present, like English, German or Spanish. The consonant part /t/ can be considered as an unvoiced sound, and because of this, its spectrum is very broad (see spectrogram of the original signal for example in Figure 30). On the other hand, the vowel part /a/ is a voiced sound whose spectrum is concentrated in the low frequency range.

For all simulations, the results will be presented using the same schema. First of all, the transfer functions of the band pass filters calculated by the programs **Simulation_CIS** or **Simualtion_SPLS** will be shown. For the SPLS strategy, these plots are the same as for the CIS strategy; therefore they won't be present in the results of the SPLS section. Then, a plot generated by the programs for each simulation will be depicted. In each of these plots, the original sound wave, the outputs of the band pass filter bank and the acoustic simulated signal are plotted. The acoustic simulations for both cases, the CIS strategy and the SPLS, are calculated using their corresponding formula described in page 49. After that, the wave spectrum of the obtained simulation will always be compared to the original one using the MATLAB software tool for speech analysis called "colea", a free software developed by Loizou Group at the UT Dallas (USA) (see Websites). The last figure for each simulation is the spectrum of the signal evaluated in three different times. It can be considered as a transversal cut of the simulated signal spectrogram, realized in the time points 200 ms, 400 ms and 500 ms respectively. These spectrums are obtained using again the MATLAB software toolbox colea, which calculates them by performing a 20-pole LPC analysis of 10 ms of the speech of the time point selected. The three first formant frequencies are calculated by peak-picking the LPC spectrum. Their values are also written in the plots.

Simulation_CIS and **Simulation_SPLS** also plot the trains of biphasic pulses that would be generated by each electrode. Since the goal of the thesis is the comparison of the sounds generated by both strategies, these plots have not been shown here.

Some other interesting aspects that had to be taken in account for programming the simulation of both strategies, like MATLAB commands or other implementation considerations are also outlined in the following subsections.

Simulation_CIS

Results changing the frequency range:

An important parameter for the CIS strategy is the frequency range that is covered by the electrical stimulation done through the electrodes. Three frequency ranges were used in the simulations. The first frequency range corresponds to the medium electrode used by MED-EL, which covers the distances of 11 mm to 31.9 mm of the cochlea, where the distances are calculated from the apex. It has normally an insertion depth of 24 mm and a total length of 20.9 mm (see Figure 27). Its corresponding frequency range is BW1=[603 Hz, 12986 Hz]. The second frequency range selected corresponds to the short electrode used by MED-EL, which covers the distances of 15,5mm to 32,6 mm of the cochlea, where again the distances are calculated from the apex. It has normally an insertion depth of 15 mm and a length of 13.1 mm (see Figure 27). Its corresponding frequency range is in this case BW2=[2587 Hz 14320 Hz]. This short electrode is used in patients with strong ossification of the cochlea, where a full insertion is not possible. The third frequency range [215 Hz 4891 Hz] is the one used by Chen et al. [Chen et al;, 2009] for their simulations. It would cover the distances 5,5 mm to 24.95 mm of the cochlea, where distances are calculated from the apex.

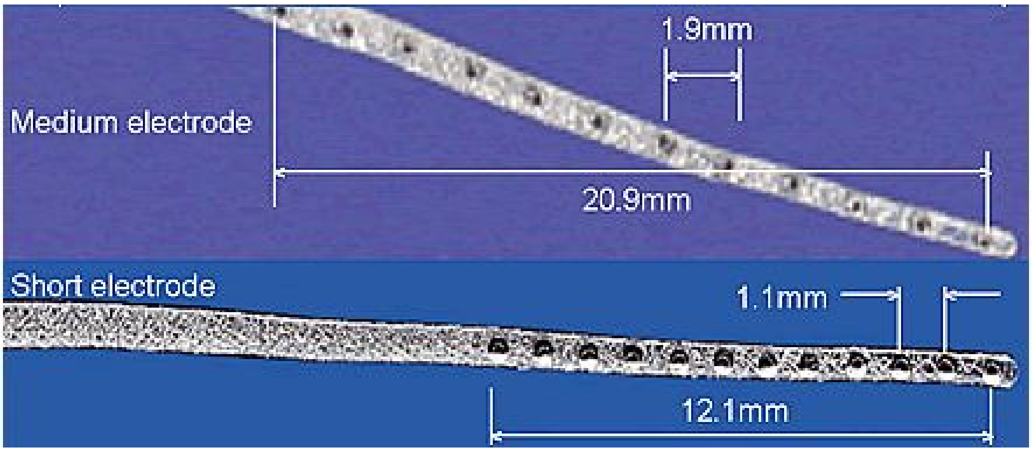


Figure 27 MED-EL Medium and short electrodes.

In order to be able to examine the effects of the frequency range in the acoustic simulation, the other three parameters were chosen to be the same. The three simulations were done choosing six channels and FIR filters with uniform bandwidth.

The transfer functions of the band pass bank filters for each frequency range BW1, BW2 and BW3 are shown in Figure 28, Figure 32 and Figure 36 respectively. Then, the results of the simulation for each frequency ranges are depicted in Figure 29, Figure 33 and Figure 37 respectively.

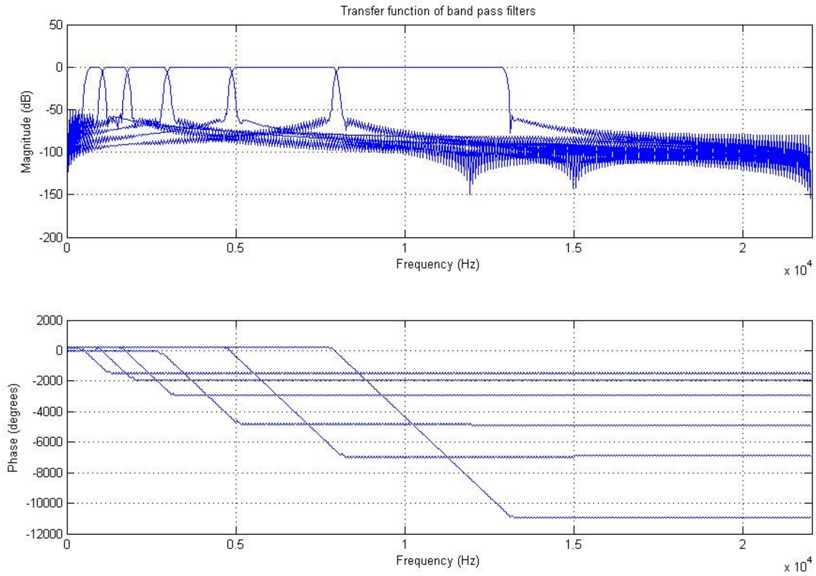


Figure 28 Transfer functions of the band pass filter bank used in the CIS and SPLS simulation using 6 channels, FIR filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz].

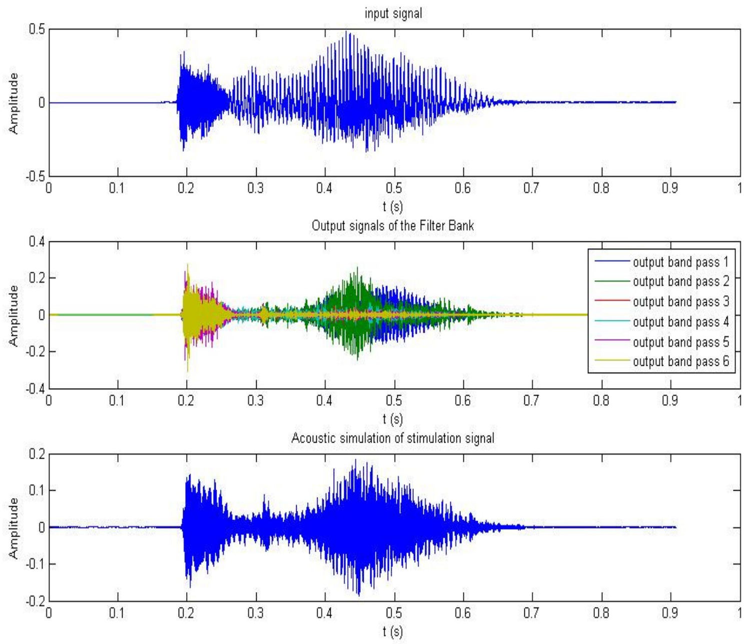


Figure 29 Results of the CIS simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz].

After showing the results of the acoustic simulation in time, both spectrograms of the simulated and the original wave sound are shown (see Figure 30, Figure 34 and Figure 38 for simulations with BW1, BW2 and BW3 respectively). The acoustic simulations can be heard in the CD attached to this master thesis, under the CIS simulations directory, with the names **sim_CIS_6_RectBW_FIR_BW1.wav**, **sim_CIS_6_RectBW_FIR_BW2.wav** and **sim_CIS_6_RectBW_FIR_BW3.wav**.

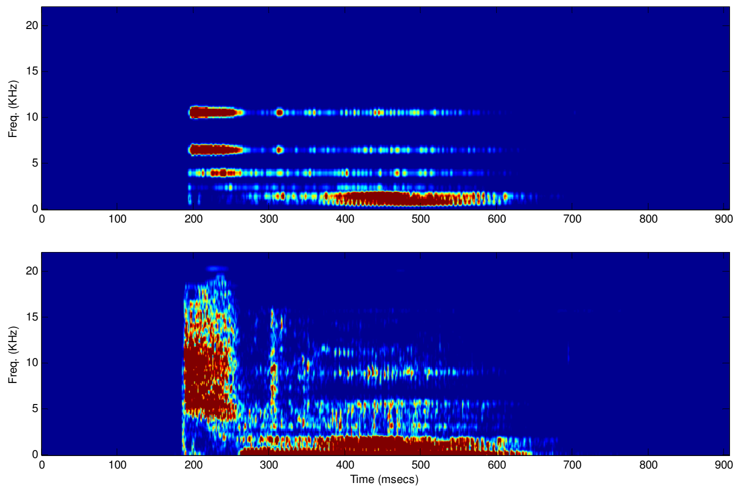
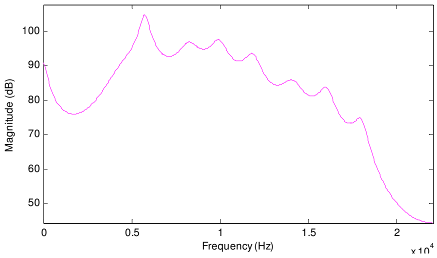
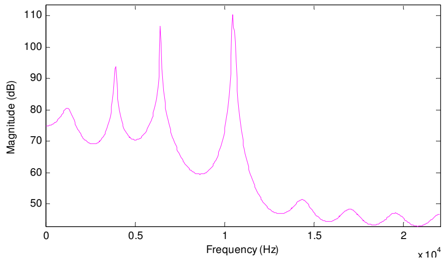


Figure 30 Top: Spectrogram of the CIS acoustic simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz]. Bottom: Spectrogram of the original word.

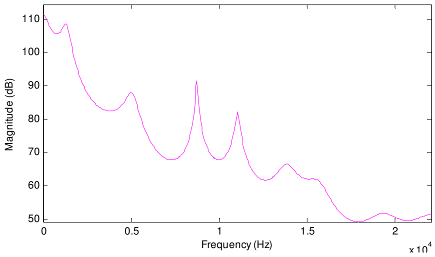
In the case of using BW1 as frequency range, it is observed in the top spectrogram of Figure 30, that the very low frequencies are not present at all. This makes the quality of the reproduction of the vowel at the end of the syllable to be poor. However, there should be enough information in the low frequencies to perceive the /a/. The consonant can be hardly due to the lack of components in the high frequency band (over 5 KHz).



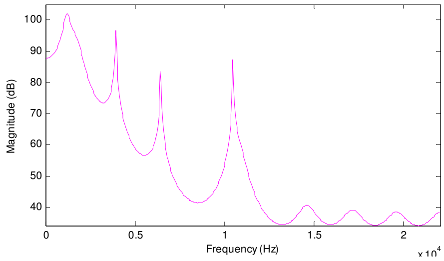
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



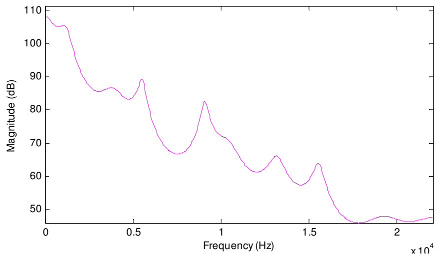
F1= 1206 Hz, F2= 3916 Hz F3= 6417 Hz



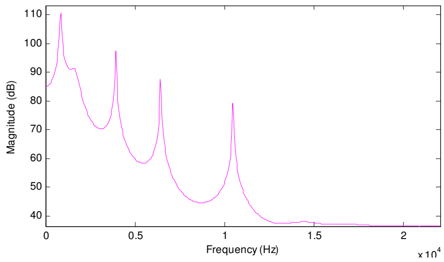
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 1206 Hz, F2= 3916 Hz F3= 6417 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 861 Hz, F2= 1550 Hz F3= 3919 Hz

Figure 31 Spectrum plots of the CIS acoustic simulation of the syllable “tra” using 6 channels, FIR filters with uniform bandwidth and using the frequency range BW1=[603 Hz, 12986 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

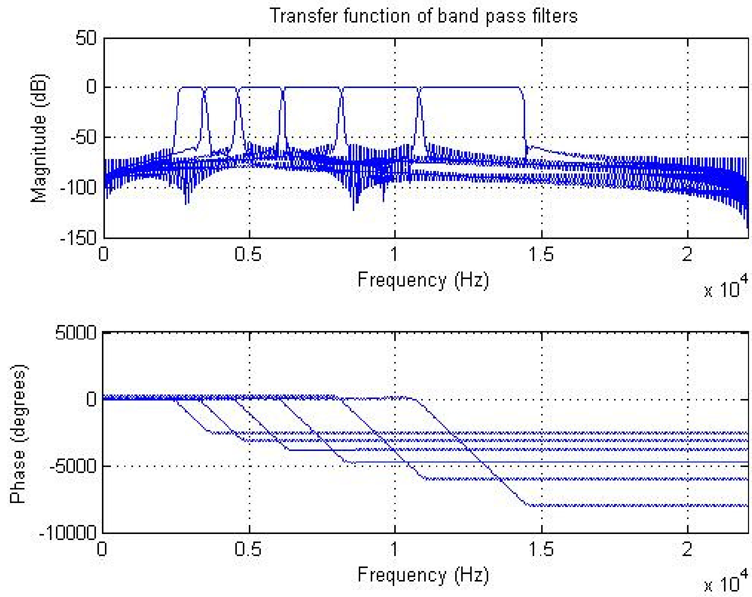


Figure 32 Transfer functions of the band pass filter bank used in the CIS and SPLS simulation using 6 channels, FIR filters with uniforms bandwidths and using the frequency range BW2=[2587 Hz, 14320 Hz].

By choosing BW2 as the frequency range, the low and middle frequencies are filtered out as can be seen in the central plot of Figure 33. Due to this, the electrical stimulation when the vowel part begins practically disappears, and it is expected that the simulated signal will be very poor at that time.

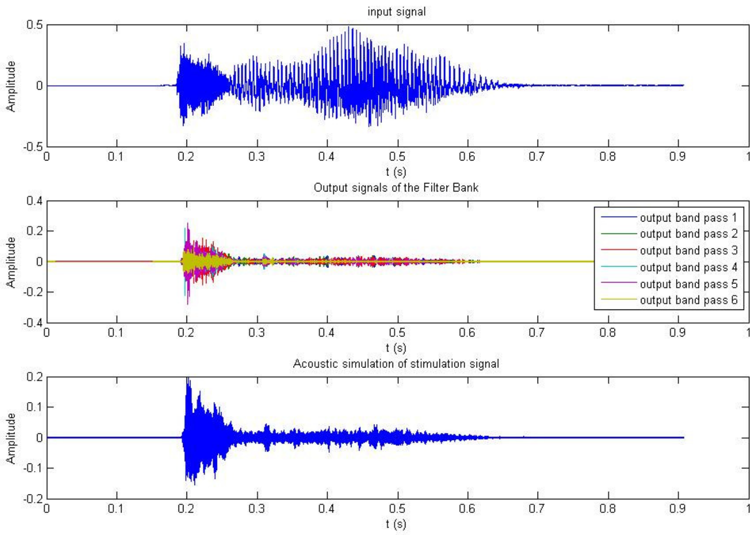


Figure 33 Results of the CIS simulation of the syllable “tra” using 6 channels, FIR filters with uniforms BW and using the frequency range BW2=[2587 Hz, 14320 Hz].

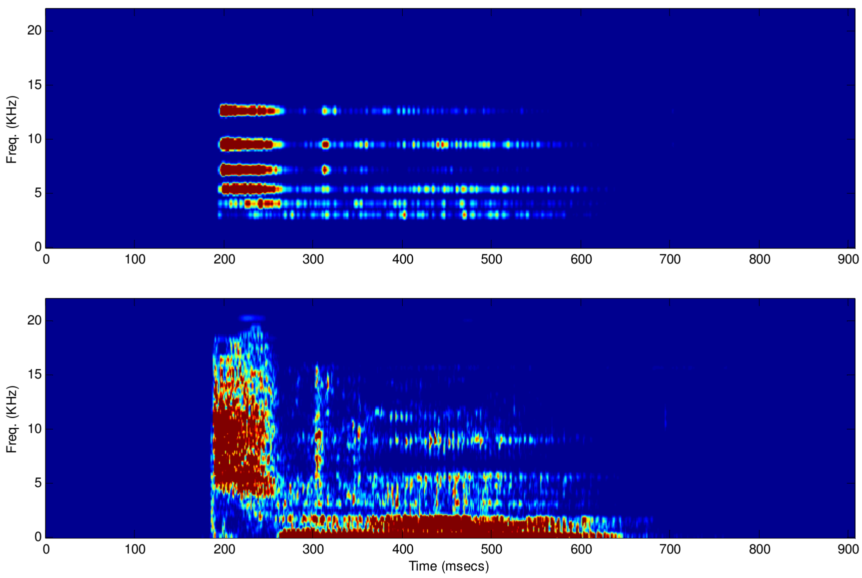
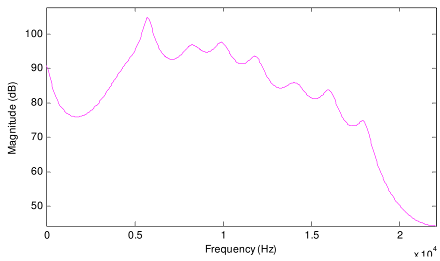
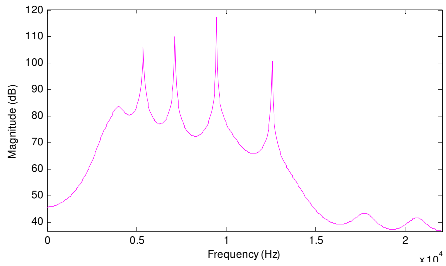


Figure 34 Top: Spectrogram of the CIS acoustic simulation of the syllable “tra” using 6 channels, FIR filters with uniforms BW and using the frequency range BW2=[2587 Hz, 14320 Hz].. Bottom: Spectrogram of the original word.

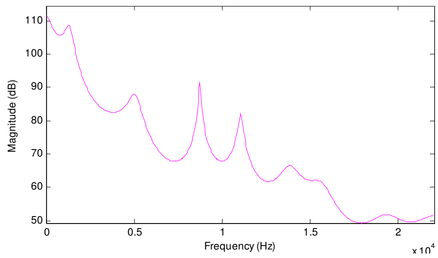
Due to the lack of low-middle frequency stimulation, the simulated sound can as expected hardly reproduce the vowel /a/ of the syllable, this part is not present in the simulation as it is shown in Figure 34, only a mix of high frequencies producing the /t/ can be heard at the beginning, obtaining an unintelligible sound.



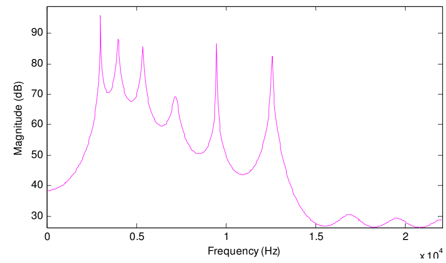
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



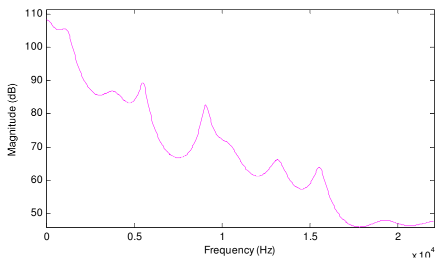
F1= 4005 Hz, F2= 5383 Hz F3= 7149 Hz



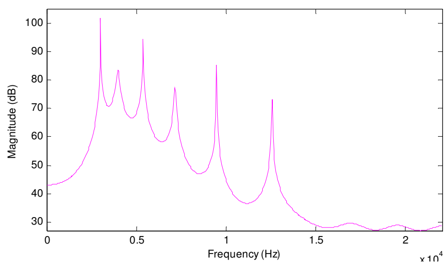
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 3015 Hz, F2= 3962 Hz F3= 5340 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 3015 Hz, F2= 3962 Hz F3= 5383 Hz

Figure 35 Spectrum plots of the CIS acoustic simulation of the syllable “tra” using 6 channels, FIR filters with uniform bandwidth and using the frequency range BW2=[2587 Hz, 14320 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

The third frequency range covers the low frequency much better than the first two choices. On the other hand, the high frequencies are poorly represented.

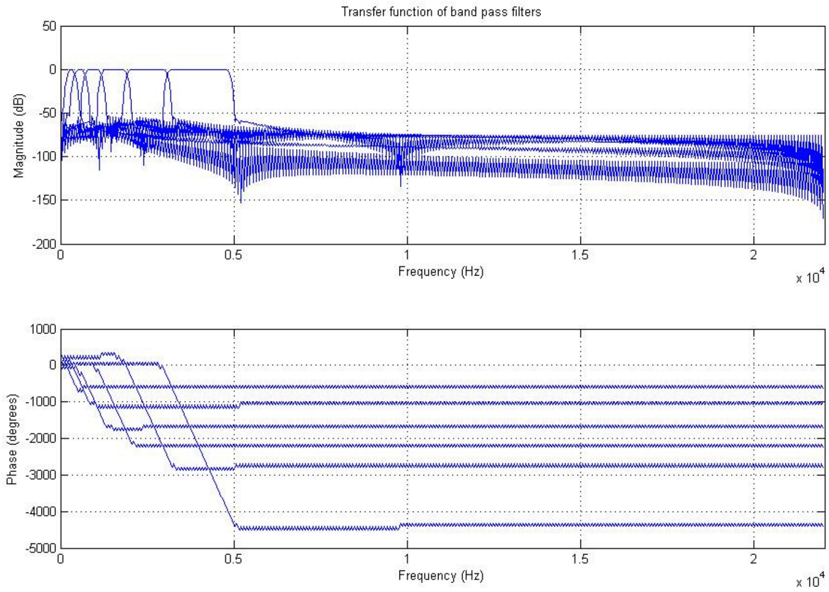


Figure 36 Transfer functions of the band pass filter bank used in the CIS and SPLS simulation using 6 channels, FIR filters with uniforms bandwidths and using the frequency range BW3=[215 Hz, 4891 Hz].

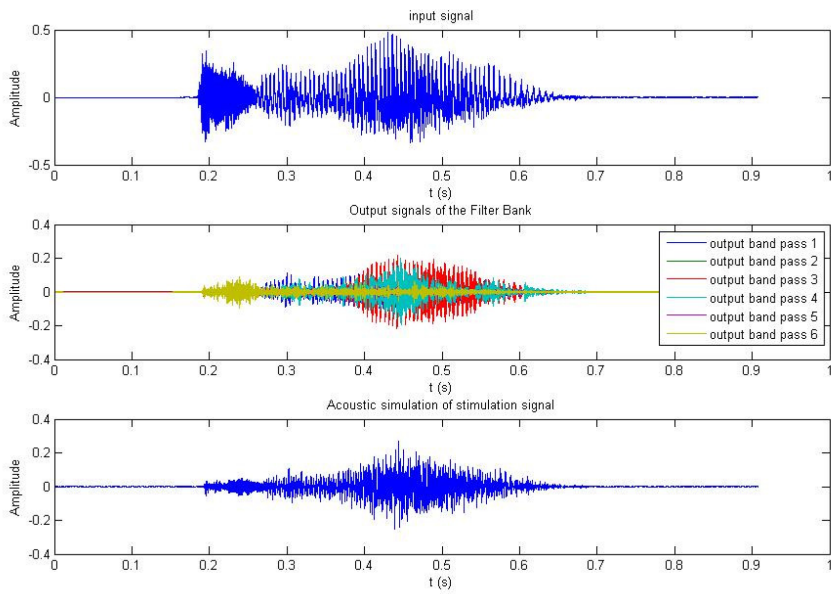


Figure 37 Results of the CIS simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].

This effect can be observed in the time representation of the simulated signal (see Figure 37). At the beginning of the syllable, where the /t/ should be heard, the stimulation values are very small, and therefore the /t/ cannot be heard clearly. Although, using this bandwidth, the acoustic simulation is the best perceived by me.

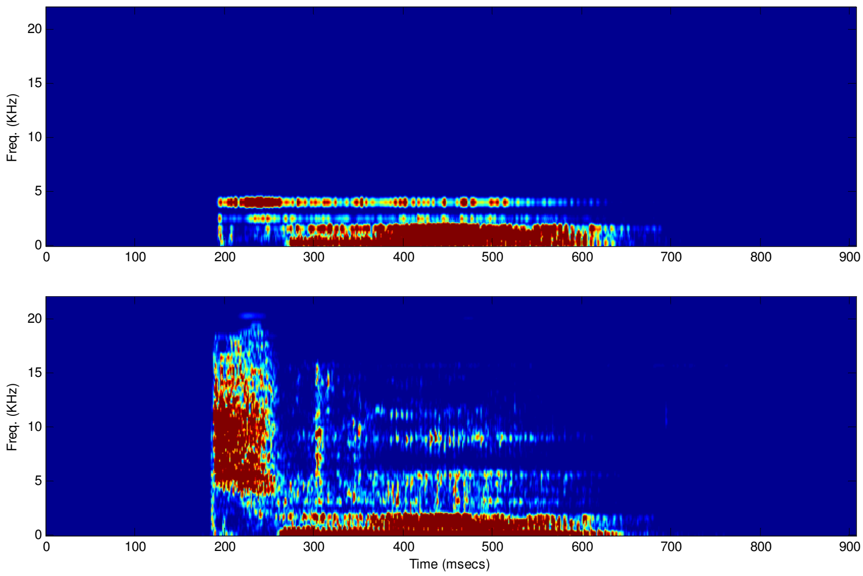
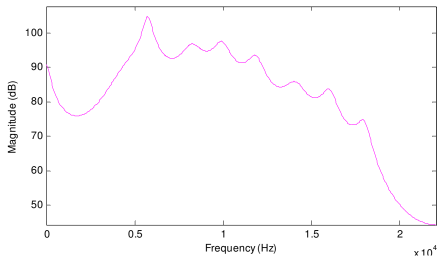


Figure 38 Top: Spectrogram of the CIS acoustic simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].. Bottom: Spectrogram of the original word.

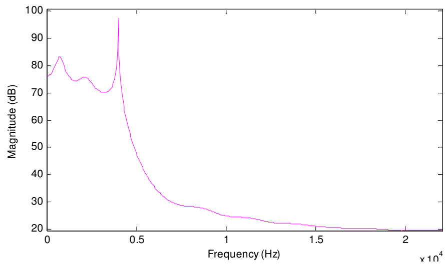
In the spectrogram shown in Figure 38, the six bands of the six corresponding channels cannot be so easily identified, as it was for the other two frequency ranges. The overlap in the low frequencies raises the quality of the voiced sound of the syllable at the end of the wave form. Using this frequency range, the voiced sounds can be perceived better.

In all plots of the spectrum for the times 200ms, 400 ms and 500 ms (see Figure 31, Figure 35 and Figure 39) it can be observed that the form is always very different than the original one. The formants have very different values; usually the center frequencies of the band pass filters. Because of this, the acoustic simulations of the

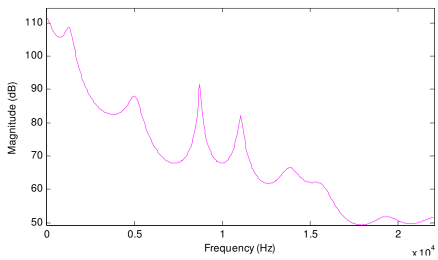
CIS strategy always sound metallic and it is not possible to distinguish if the voice is from a male or female.



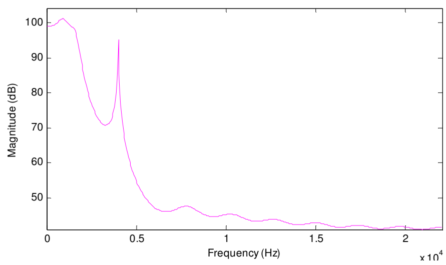
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



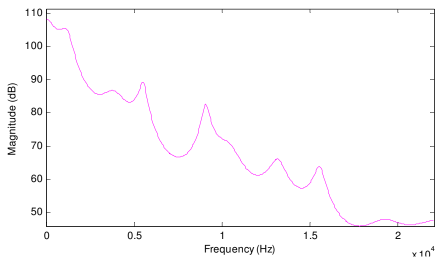
F1= 732 Hz, F2= 2110 Hz F3= 4005 Hz



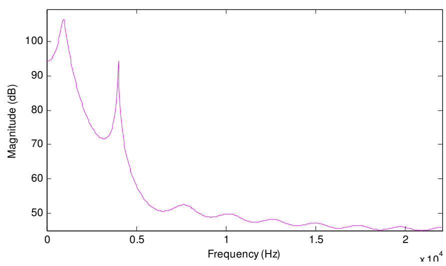
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 904 Hz, F2= 4005 Hz F3= 7795 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 947 Hz, F2= 4005 Hz F3= 7623 Hz

Figure 39 Spectrum plots of the CIS acoustic simulation of the syllable "tra" using 6 channels, FIR filters with uniform bandwidth and using the frequency range BW3=[215 Hz, 4891 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

Results by varying the number of channels

In order to be able to examine the effects of the number of channels in the acoustic simulation, the other three parameters were left to be the same. The three simulations, using 2, 4 and 8 channels, were done using the frequency range BW3=[215 Hz, 4891 Hz] and Butterworth IIR filters of 4th order with uniform bandwidth.

To expand the frequency range in the required number of channels, the MATLAB function **equal_xbm_bands** is used. This function was programmed by Bertrand Delgutte in 2000 and is as free software to download under the website http://research.meei.harvard.edu/chimera/. This function divides the frequency interval into N bands of equal width along the human basilar membrane. It is based on M.C. Liberman's cochlear frequency map for the cat scaled to match human frequency range of hearing. This function calls two other functions, which were also programmed by the same author: **cochlear_map** and **inv_cochlear_map**. The first one converts a frequency to a distance along the basilar membrane and the second one does the inverse function, conversion from distance to frequency.

The transfer functions of the band pass filters used for each one of the three simulations are depicted in Figure 40, Figure 44 and Figure 48 respectively.

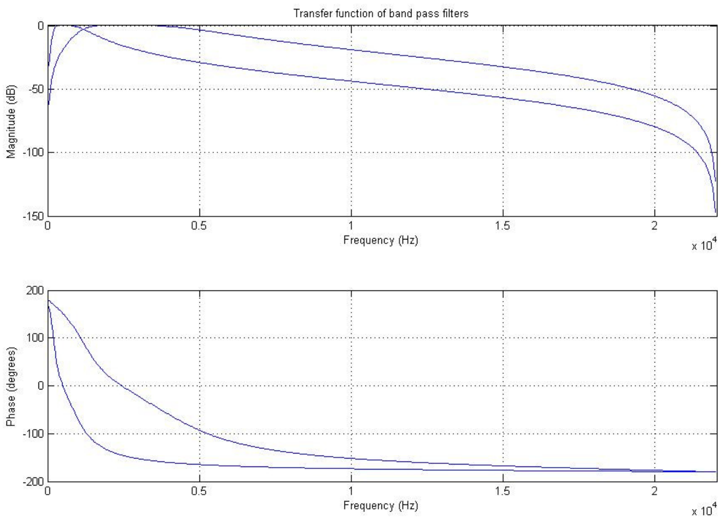


Figure 40 Transfer functions of the band pass filter bank used in the CIS and SPLS simulation using 2 channels, IIR filters (Butterworth 4th order) with uniforms bandwidths and using the frequency range BW3=[215 Hz, 4891 Hz].

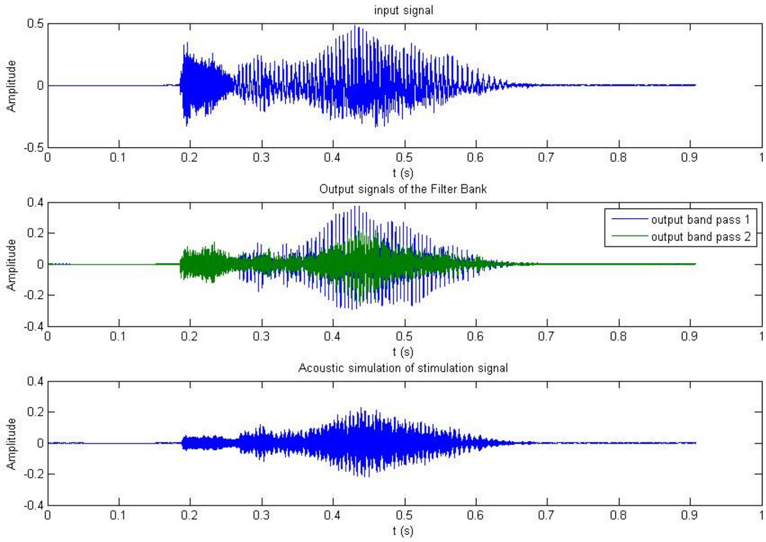


Figure 41 Results of the CIS simulation of the syllable "tra" using 2 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].

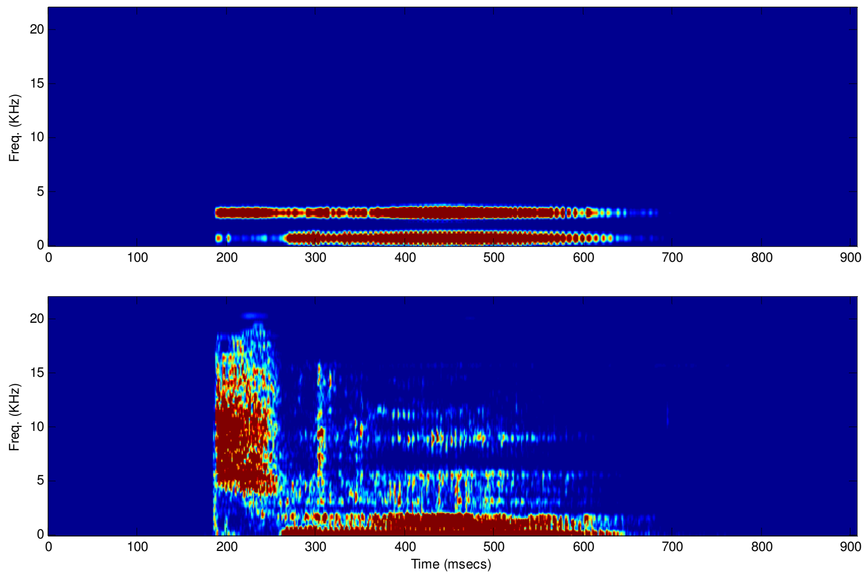
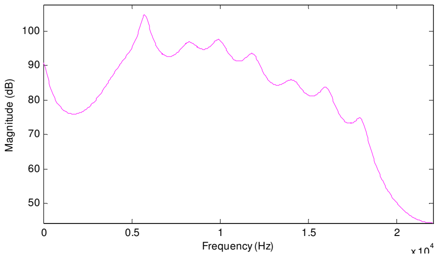
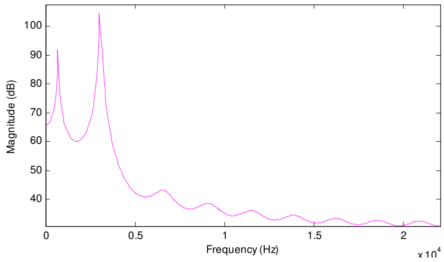


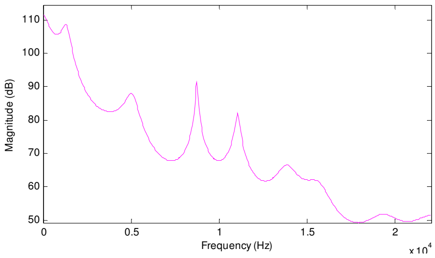
Figure 42 Top: Spectrogram of the CIS acoustic simulation of the syllable "tra" using 2 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].. Bottom: Spectrogram of the original word.



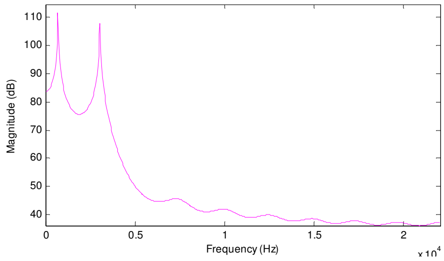
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



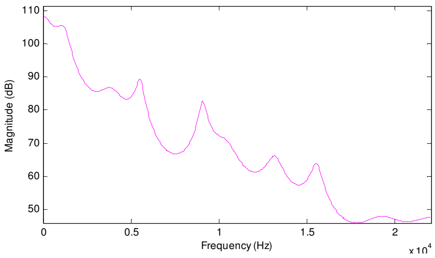
F1= 689 Hz, F2= 3015 Hz F3= 6546 Hz



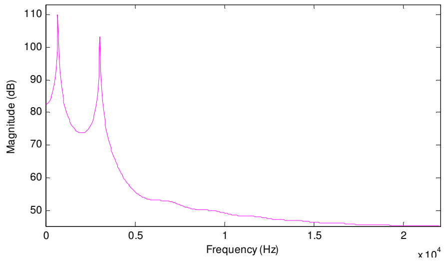
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 689 Hz, F2= 3058 Hz F3= 7278 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 689 Hz, F2= 3058 Hz F3= 15848 Hz

Figure 43 Spectrum plots of the CIS acoustic simulation of the syllable "tra" using 2 channels, IIR filters (Butterworth 4th order) with uniform bandwidth and using the frequency range BW3=[215 Hz, 4891 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

Using two channels, the sound obtained in the simulation is unintelligible. The frequency information is too low to get enough perception of the original sound. This can be observed in both the spectrogram (Figure 42) and in the spectrums (Figure 43).

For the simulation using 4 channels, an improvement in the resulting wave sound is expected. As it was discussed before, the selected frequency range BW3 makes it hard to perceive the higher frequencies. That's why the first part of the syllable is still hard to indentify. On the other hand, the low frequency information is much present in this second simulation and the vowel at the end of the signal should be better guessed. See the spectrogram comparison in Figure 46.

The plots of the spectrum for the times 200ms, 400 ms and 500 ms (see Figure 47) show as expected that the formants still have very different values. The acoustic simulations sound metallic and it is not possible to distinguish if the voice is from a male or female.

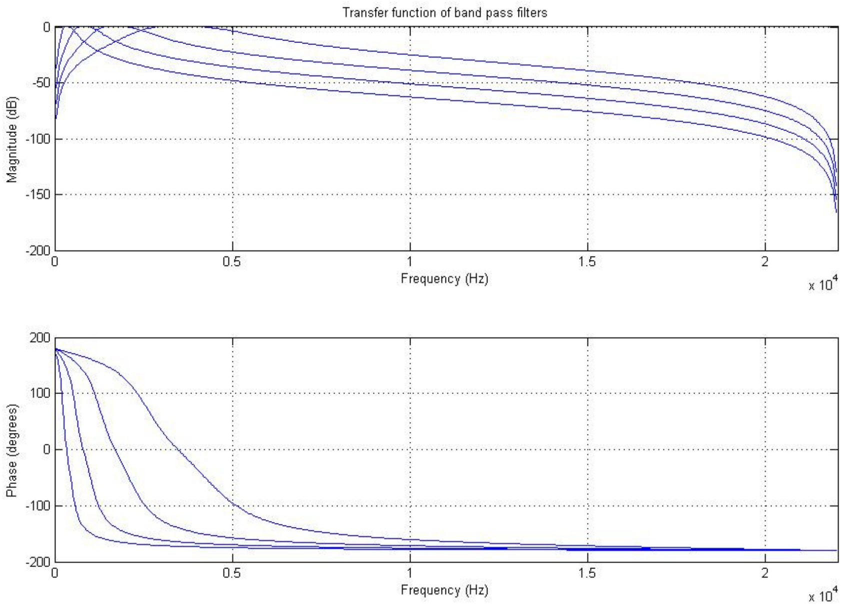


Figure 44 Transfer functions of the band pass filter bank used in the CIS and SPLS simulation using 4 channels, IIR filters (Butterworth 4th order) with uniforms bandwidths and using the frequency range BW3=[215 Hz, 4891 Hz].

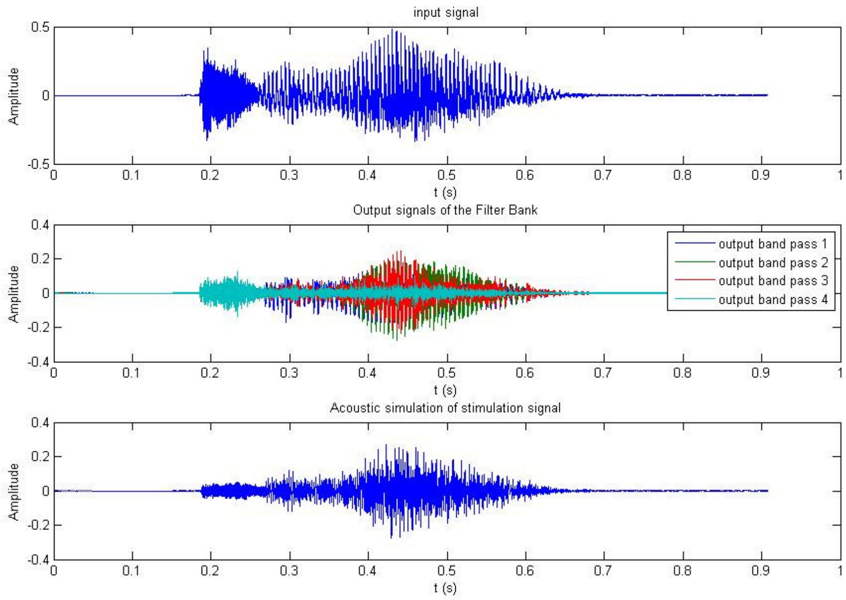


Figure 45 Results of the CIS simulation of the syllable “tra” using 4 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].

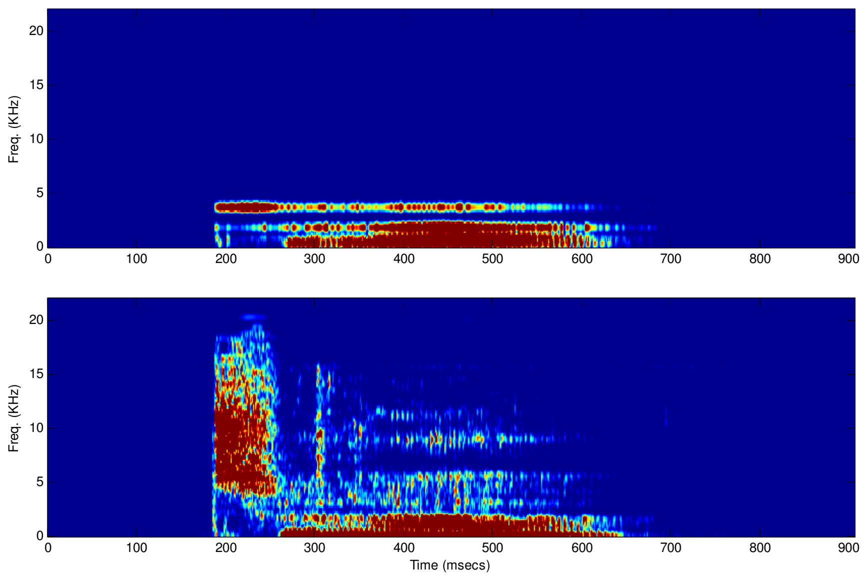
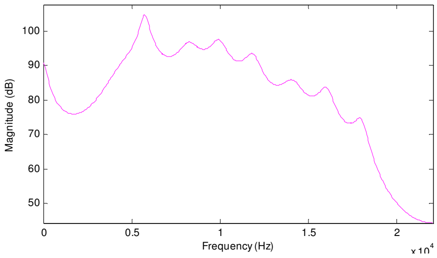
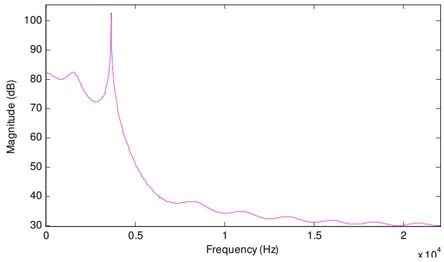


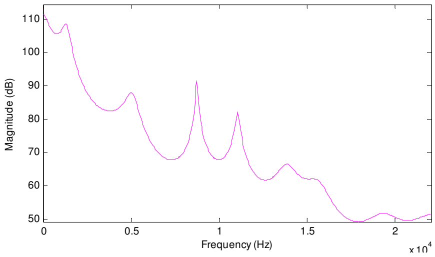
Figure 46 Top: Spectrogram of the CIS acoustic simulation of the syllable “tra” using 4 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].. Bottom: Spectrogram of the original word.



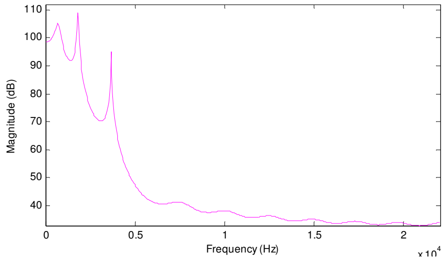
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



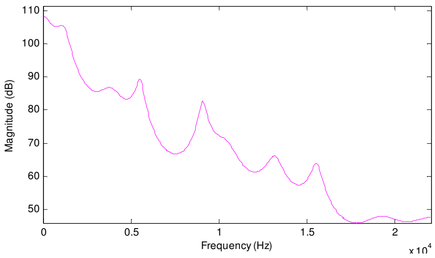
F1= 1550 Hz, F2= 3661 Hz F3= 8183 Hz



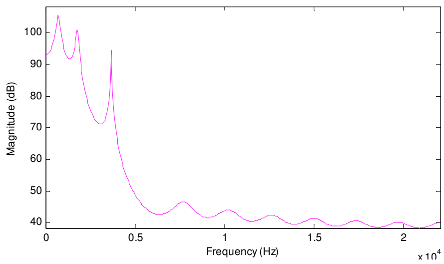
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 689 Hz, F2= 1809 Hz F3= 3661 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 689 Hz, F2= 1766 Hz F3= 3661 Hz

Figure 47 Spectrum plots of the CIS acoustic simulation of the syllable "tra" using 4 channels, IIR filters (Butterworth 4th order) with uniform bandwidth and using the frequency range BW3=[215 Hz, 4891 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

In the third simulation using 8 channels, the frequency resolution is much better than the 2 channels simulation. Characteristic information has been achieved at least in the low frequency range, with the result that the vowel /a/ can be heard clearly.

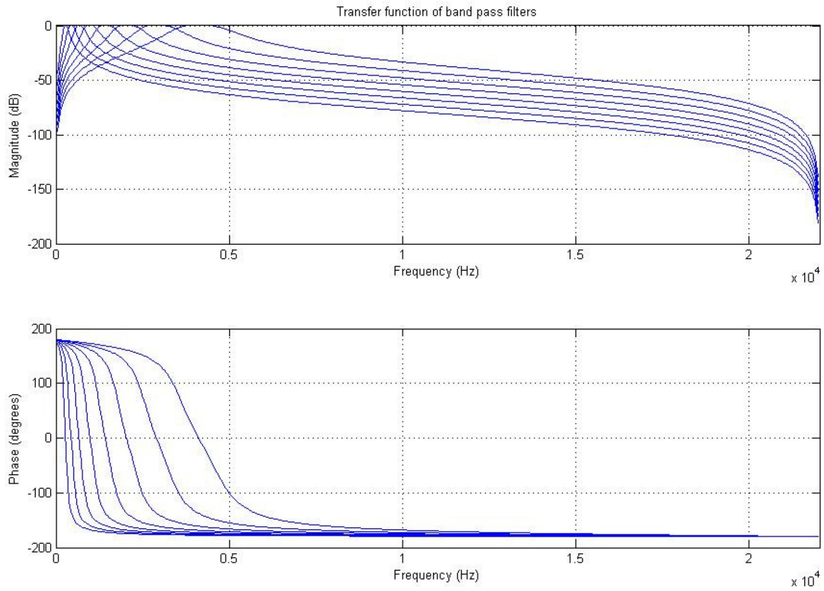


Figure 48 Transfer functions of the band pass filter bank used in the CIS and SPLS simulation using 8 channels, IIR filters (Butterworth 4th order) with uniforms bandwidths and using the frequency range BW3=[215 Hz, 4891 Hz].

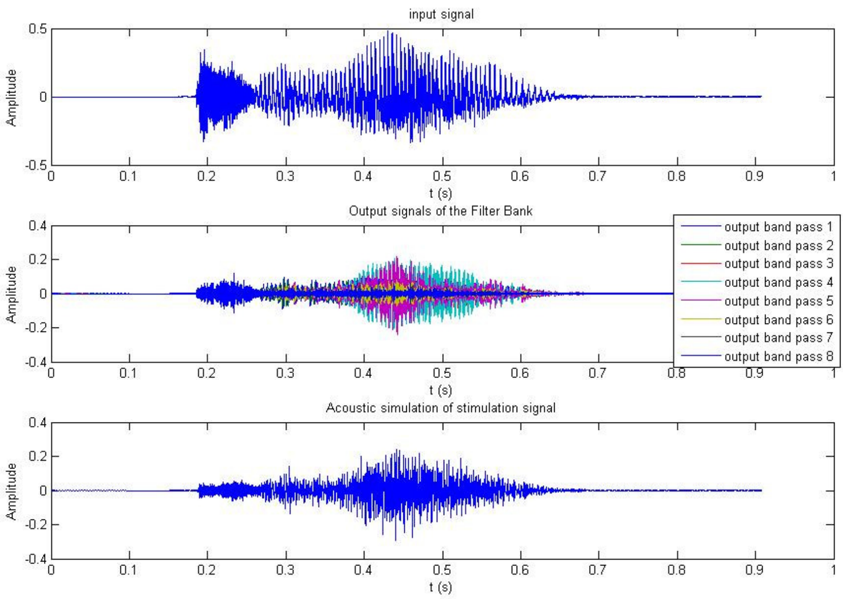


Figure 49 Results of the CIS simulation of the syllable "tra" using 8 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].

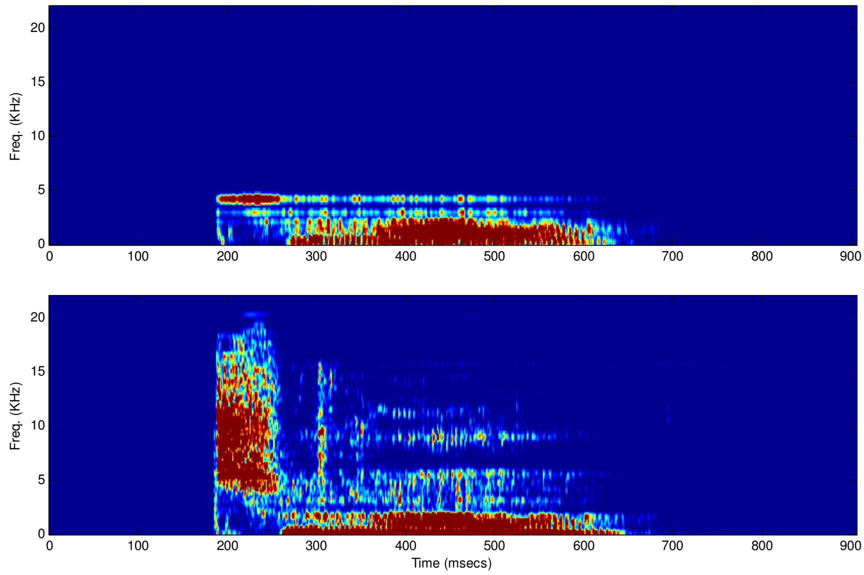


Figure 50 Top: Spectrogram of the CIS acoustic simulation of the syllable "tra" using 8 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].. Bottom: Spectrogram of the original word.

On the other hand, the consonant part of the simulation is limited to the fact that the frequency range chosen by the BW3 is for the unvoiced sounds not as optimal as for voiced sounds. That is the reason why the /t/ is hardly perceived.

In Figure 51, it can be observed that the spectrum plots for 400 ms and 500 ms of the acoustic simulation using 8 channels are tending to look like the spectrum of the original signal in such times, except for the higher frequencies. The formants for time 500 ms get closer to the original values.

The acoustic simulations can be heard in the CD attached to this master thesis, under the CIS simulations directory, with the names s**im_CIS_2_RectBW_IIR_BW3.wav**, **sim_CIS_4_RectBW_IIR_BW3.wav** and **sim_CIS_8_RectBW_IIR_BW3.wav**.

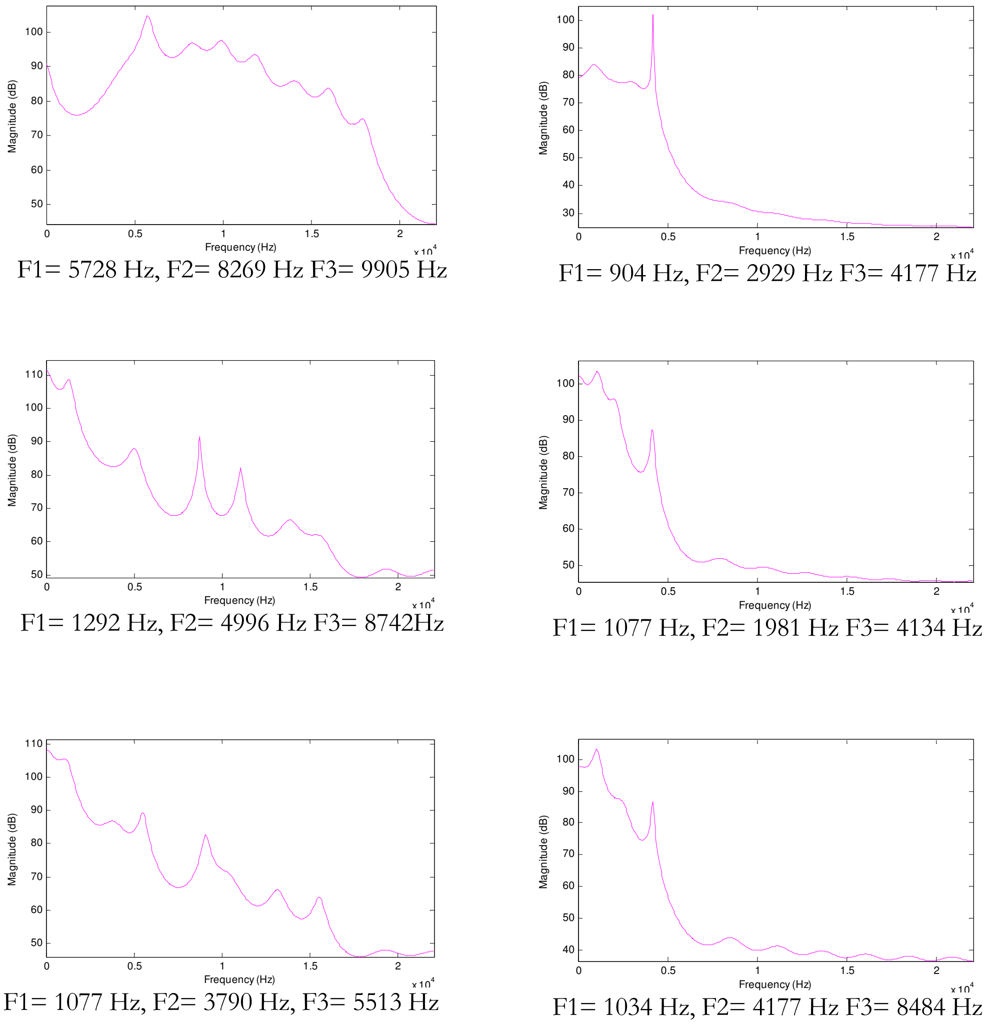


Figure 51 Spectrum plots of the CIS acoustic simulation of the syllable "tra" using 8 channels, IIR filters (Butterworth 4th order) with uniform bandwidth and using the frequency range BW3=[215 Hz, 4891 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

The results with the CIS strategy are still very hard to be perceived, when one is not familiar with such a metallic sound. It has to be remark that I got used to this metallic sounds and recognize often more than other normal hearers perceive when I play them.

Results by varying the type of filter

Two kinds of filters have been implanted for the CIS Simulation:

1. FIR filters of order 512 using the MATLAB command **fir1**, which implements the classical method of windowed linear-phase FIR digital filter design. See MATLAB Help Documentation for more information. Using such filters, a non distortion of the phase is achieved due to the constant group delay that characterizes FIR filters. The filtered signal just suffers a delay.

2. IIR Butterworth filters of 4th order using the MATLAB command **butter**, which implements the Butterworth filter design algorithm. The obtained filters are characterized by a magnitude response that is maximally flat in the pass band and monotonic overall. For this kind of filter, the group delay is not constant and there is a phase distortion of the filtered signal.

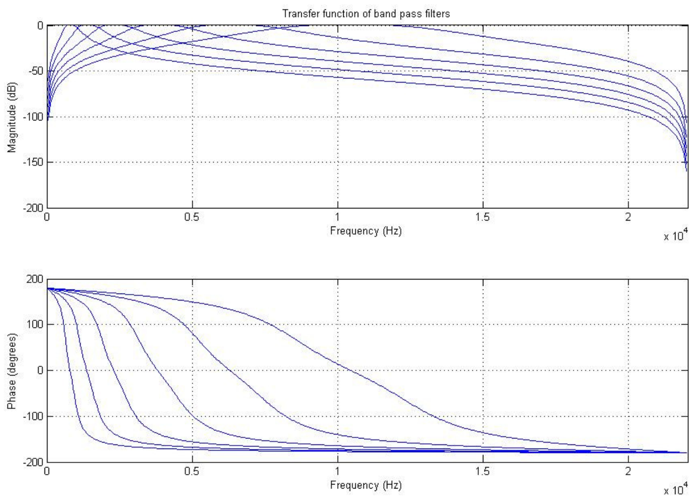


Figure 52 Transfer functions of the band pass filter bank used in the CIS and SPLS simulation using 6 channels, IIR filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz].

To evaluate the effects of the type of filter, two simulations using 6 channels, the frequency range BW1=[603 Hz, 12986 Hz] and uniform bandwidths are presented.

The results of the simulation using FIR filters are depicted in Figure 28, Figure 29, Figure 30 and Figure 31.

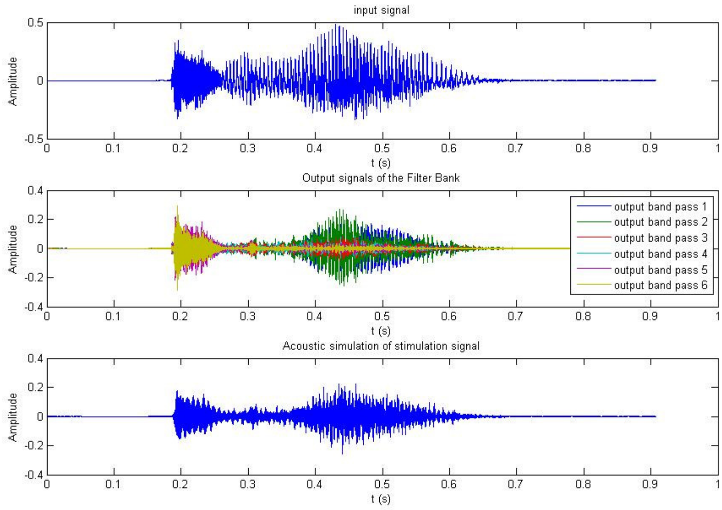


Figure 53 Results of the CIS simulation of the syllable "tra" using 6 channels, IIR 4^{th} order Butterworth filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz].

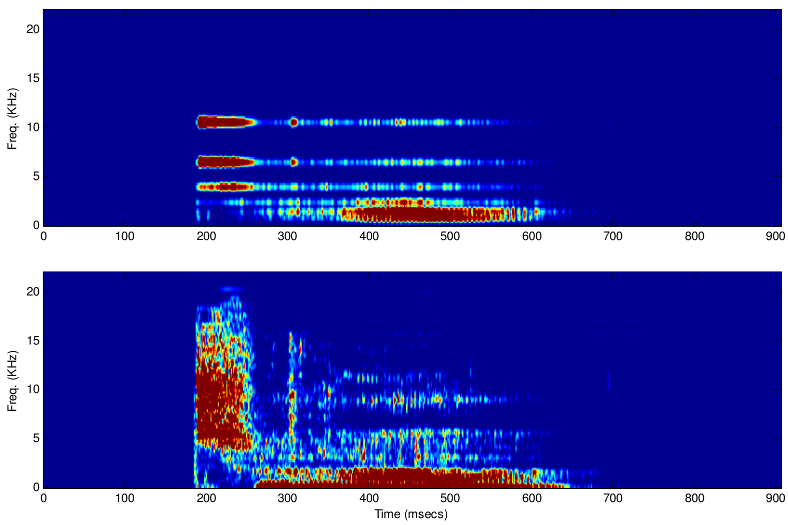
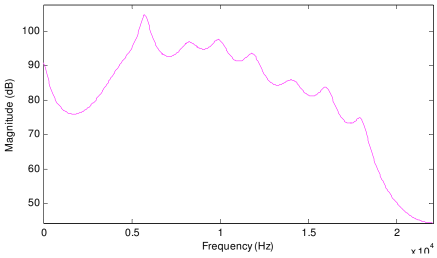
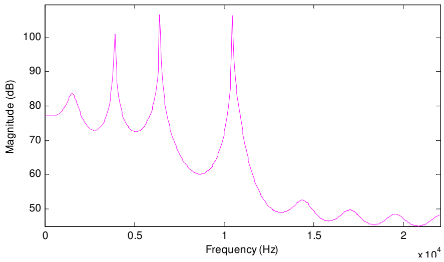


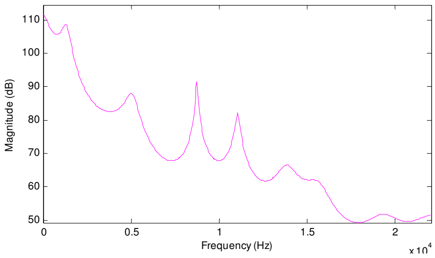
Figure 54 Top: Spectrogram of the CIS acoustic simulation of the syllable "tra" using 6 channels, IIR 4^{th} order Butterworth filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz]. Bottom: Spectrogram of the original word.



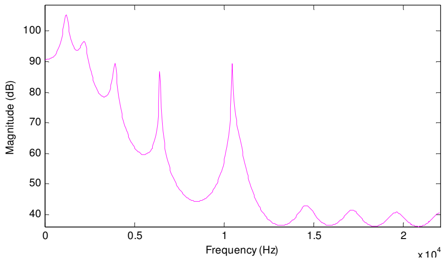
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



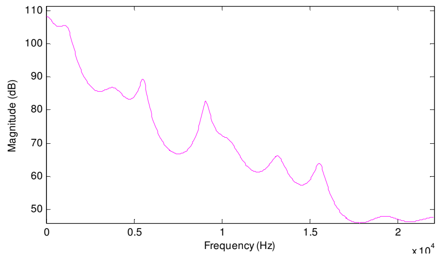
F1= 1550 Hz, F2= 3919 Hz F3= 6417 Hz



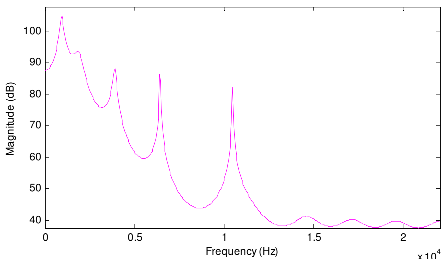
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 1206 Hz, F2= 2196 Hz F3= 3919 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 947 Hz, F2= 1852 Hz F3= 3919 Hz

Figure 55 Spectrum plots of the CIS acoustic simulation of the syllable "tra" using 6 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

The acoustic simulations obtained in both cases are very similar. Significant differences were not found. This parameter plays an important role in the implementation of the strategy in the real hardware. The advantages of FIR filters are their constant group delay, over the non-linear phase distortion of the IIR filters. On the other hand, the sparse representation of the IIR filters to get the same attenuation in the stop band, with much fewer coefficients than the FIR filter is an important advantage for the design of the memory size and computation power of the processor.

The acoustic simulations can be heard in the CD attached to this master thesis, under the CIS simulations directory, with the names **sim_CIS_6_RectBW_FIR_BW1.wav** and **sim_CIS_6_RectBW_IIR_BW1.wav.**

Results varying the type of filter's bandwidth:

Two types of filter's bandwidth were implemented:

1. Uniform bandwidth over the whole pass band: In this case, the filters will be design to have the same gain in the whole pass band.

2. Equivalent Rectangular Bandwidth: This concept was recently developed by Glasberg and Moore [Glasberg et al., 1996] to be used by the auditory filter bank models that try to simulate the human auditory periphery. To calculate the bandwidth corresponding to a center frequency f_c following formula has to be used:

$$BW_{ERB} = 24.7 \cdot (0.00437 \cdot f_c + 1)$$

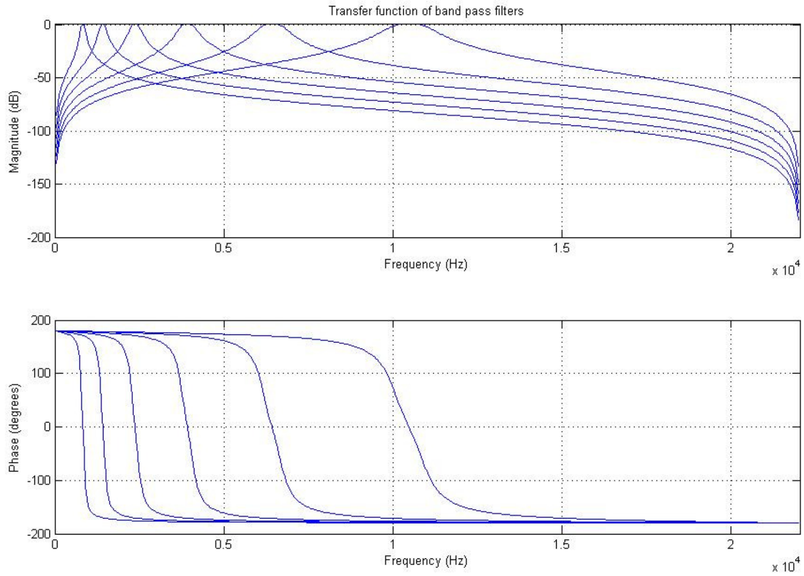


Figure 56 Transfer functions of the band pass filter bank used in the CIS and SPLS simulation using 6 channels, IIR filters with ERB and using the frequency range BW1=[603 Hz, 12986 Hz].

To evaluate the effects of the type of filter's bandwidth, two simulations using 6 channels, the frequency range BW1=[603 Hz, 12986 Hz] and Butterworth 4th order IIR filter were done. The results of the simulation using uniform bandwidths are depicted in Figure 28.

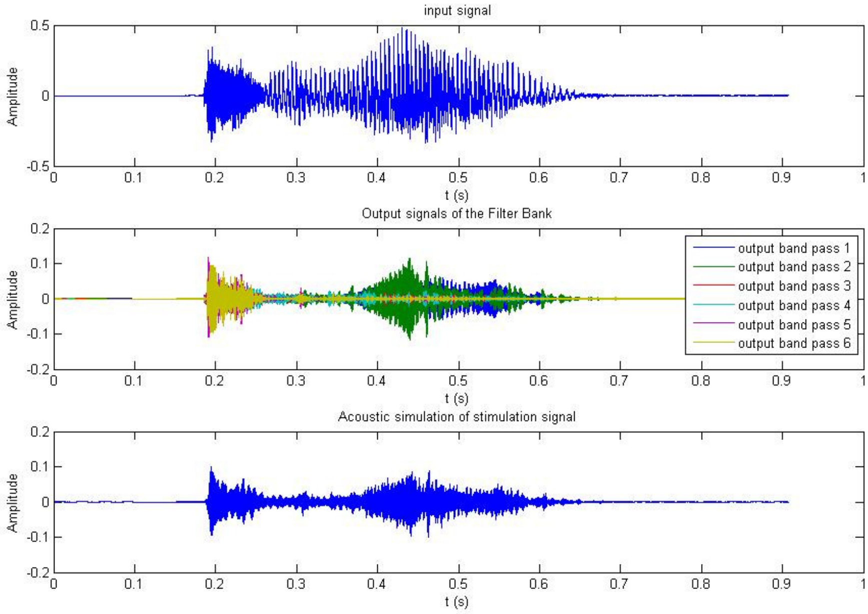


Figure 57 Results of the CIS simulation of the syllable "tra" using 6 channels, IIR 4th order Butterworth filters with ERB and using the frequency range BW1=[603 Hz, 12986 Hz].

By the simulation using ERB filters, due to the higher selectivity of the filters, the output signals of the filter bank are always lower than the outputs of the filter bank of the simulation using the uniform bandwidths (see Figure 57 compared to Figure 53). The gaps between the bands make the acoustic signal to loose important components in frequency, essential for the speech intelligibility. This effect will be less important when the number of channel grows. For low number of channels, the use of ERB filters doesn't perform a good acoustic simulation.

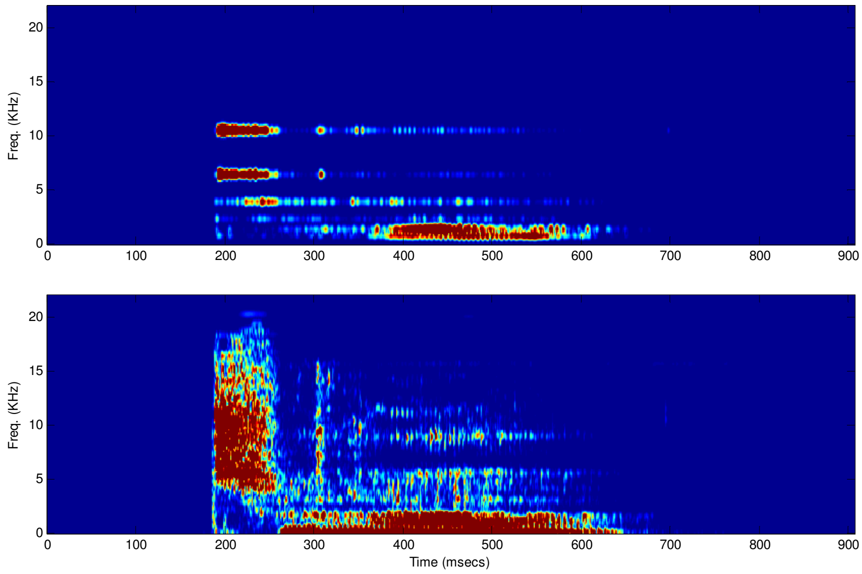
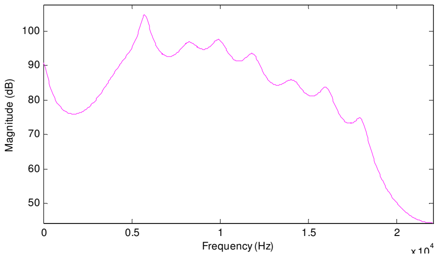


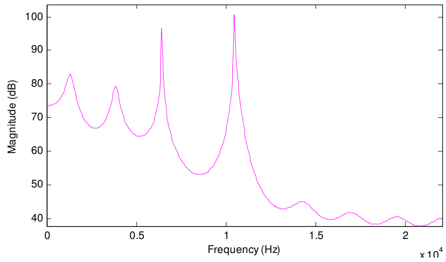
Figure 58 Top: Spectrogram of the CIS acoustic simulation of the syllable "tra" using 6 channels, IIR 4th order Butterworth filters with ERB and using the frequency range BW1=[603 Hz, 12986 Hz]. Bottom: Spectrogram of the original word.

In Figure 59, the spectrum of the simulation for the three selected times show that the selectivity of the band pass filters using the ERB model does not improve the acoustic simulations. While in the results for uniform bandwidths (see Figure 55), the spectrum below 5000 Hz, at least for 400 ms and 500 ms, tends to the original one, for the ERB filters, the lack of information in the frequency regions between the band pass filters leads to very different spectrums.

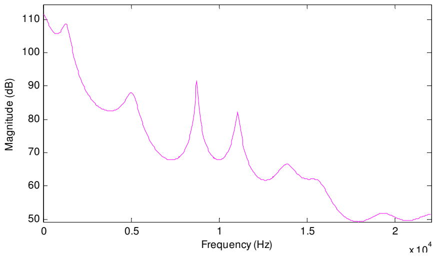
The acoustic simulations can be heard in the CD attached to this master thesis, under the CIS simulations directory, with the names **sim_CIS_6_RectBW_IIR_BW1.wav** and **sim_CIS_6_ERB_IIR_BW1.wav.**



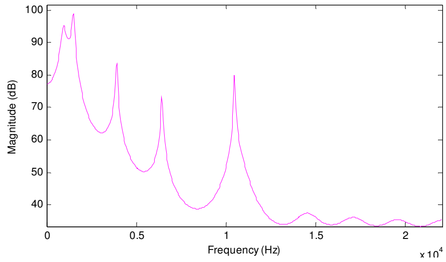
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



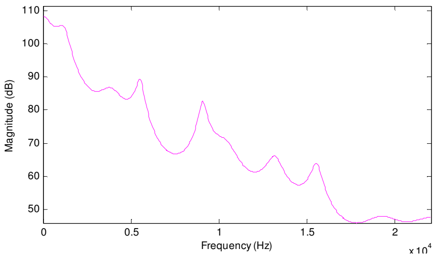
F1= 1292 Hz, F2= 3033 Hz F3= 6417 Hz



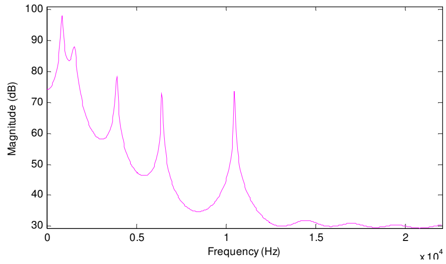
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 947 Hz, F2= 1507 Hz F3= 3919 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 861 Hz, F2= 1550 Hz F3= 3919 Hz

Figure 59 Spectrum plots of the CIS acoustic simulation of the syllable "tra" using 6 channels, IIR filters (Butterworth 4th order) with ERB and using the frequency range BW1=[603 Hz, 12986 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

Simulation_SPLS

Results changing the frequency range

The same frequency ranges as for the CIS strategy were here selected. Like before, the three simulations were done choosing six channels and FIR filters with uniform bandwidth. The results for BW1, BW2 and BW3 are depicted in Figure 60, Figure 63 and Figure 66 respectively.

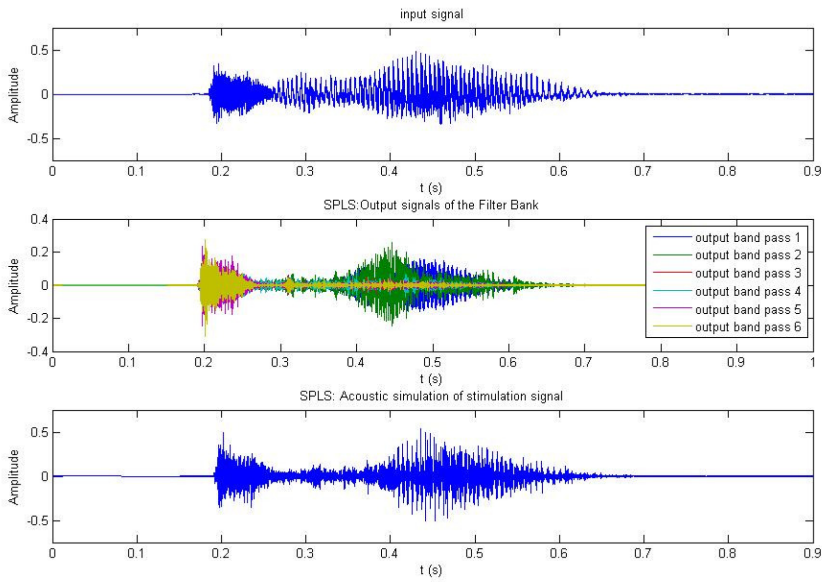


Figure 60 Results of the SPLS simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz].

Both spectrograms of the simulated and the original wave sound for each frequency range are shown in Figure 61, Figure 64 and Figure 67. The acoustic simulations can be heard in the CD attached to this master thesis, under the SPLS simulations directory, with the names **sim_SPLS_6_RectBW_FIR_BW1.wav**, **sim_SPLS_6_RectBW_FIR_BW2.wav** and **sim_SPLS_6_RectBW_FIR_BW3.wav**.

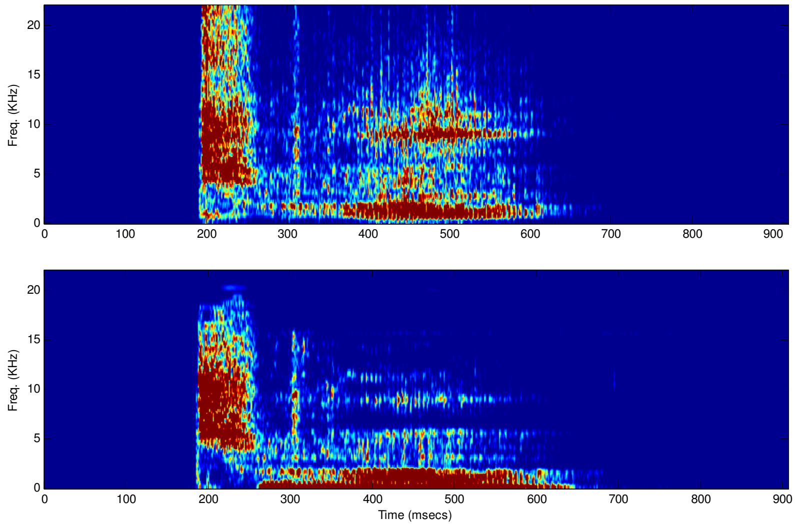
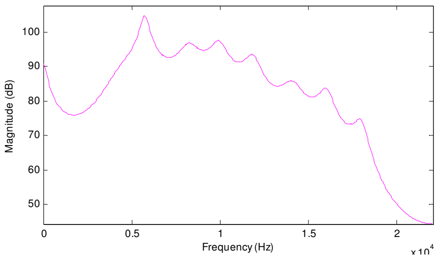


Figure 61 Top: Spectrogram of the SPLS acoustic simulation of the syllable “tra” using 6 channels, FIR filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz]. Bottom: Spectrogram of the original word.

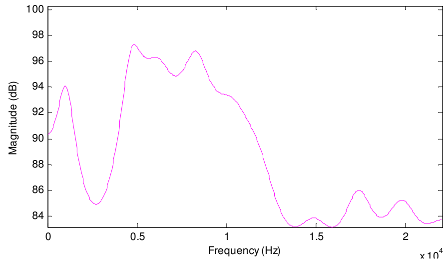
In the simulation using the frequency range BW1, the syllable “tra” can be heard clearly and much better than the simulation using the CIS strategy. The part where the “r” should be heard (between 270 ms and 370 ms) is the part where the simulation has a worse performance, but the syllable can still be good perceived.

In the spectrum plots for this simulation (Figure 62), we can see that in the voiced part of the signal, from 400 ms on where it is plotted in the middle and bottom right figures, the spectrum gets very similar to the original one, and the three first formants get close to the original ones. That is why we are able to recognize the female voice.

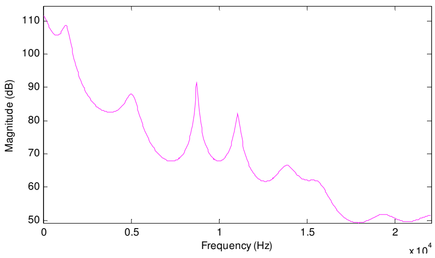
Using the SPLS strategy, this acoustic simulation could be recognized by most of the normal listeners.



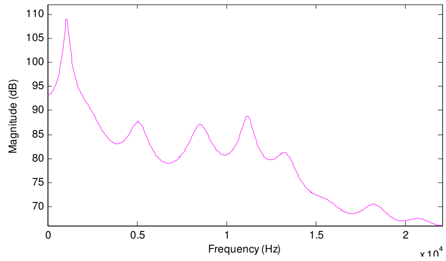
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



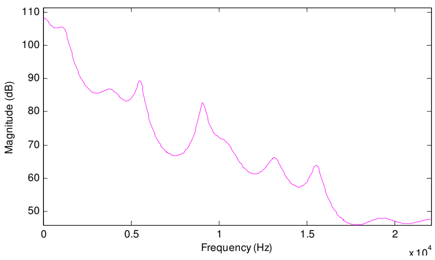
F1= 991 Hz, F2= 4823 Hz F3= 6029 Hz



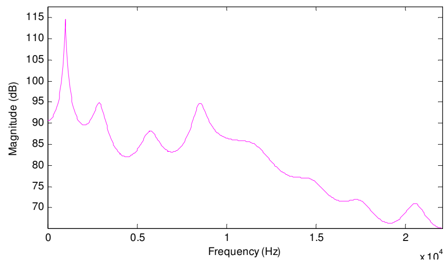
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 1077 Hz, F2= 5082 Hz F3= 8527 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 991 Hz, F2= 2885 Hz F3= 5728 Hz

Figure 62 Spectrum plots of the SPLS acoustic simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

For the simulation using the frequency range BW2, as it is seen in Figure 64, due to the absence of the low frequencies, we can expect that the /a/ won't be heard at all. This result is similar to the ones obtained with the CIS strategy. The syllable won't be right perceived by the listeners either. So, this frequency range makes problems for the recognition of speech and should be avoided. There is a need of low frequency stimulation in order to get an acceptable performance, independent of the speech strategy.

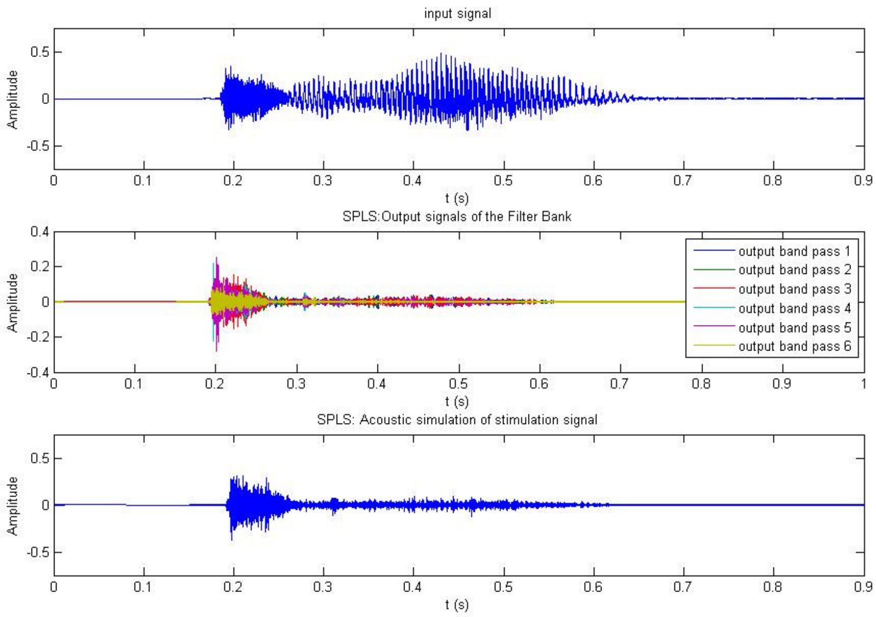


Figure 63 Results of the SPLS simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW2=[2587 Hz, 14320 Hz].

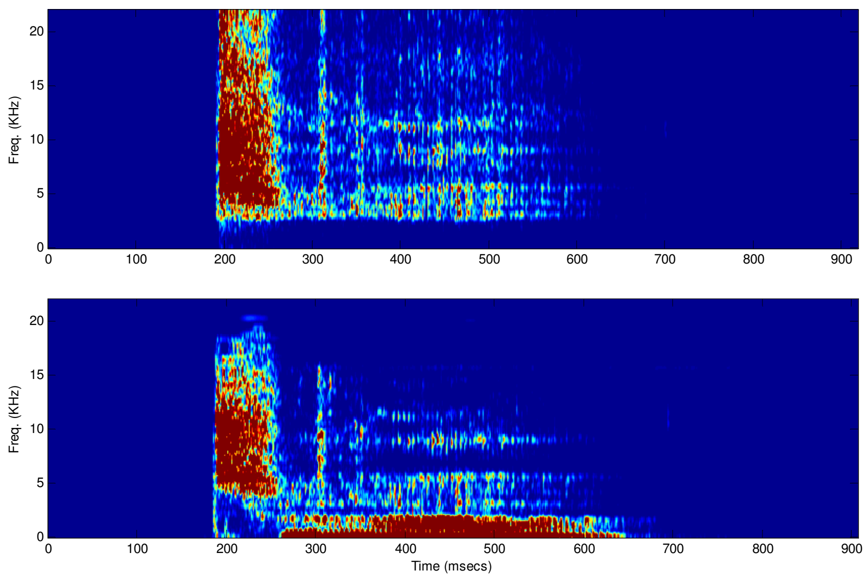
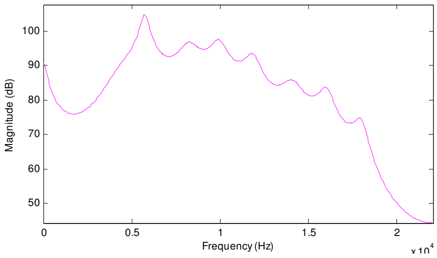
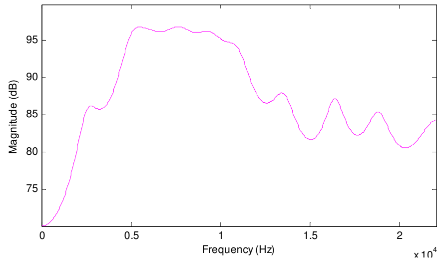


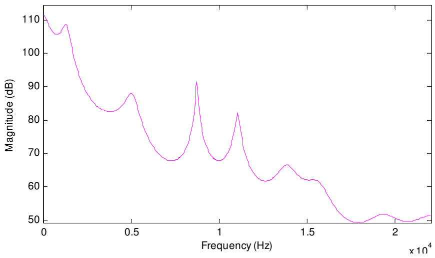
Figure 64 Top: Spectrogram of the SPLS acoustic simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW2=[2587 Hz, 14320 Hz]. Bottom: Spectrogram of the original word.



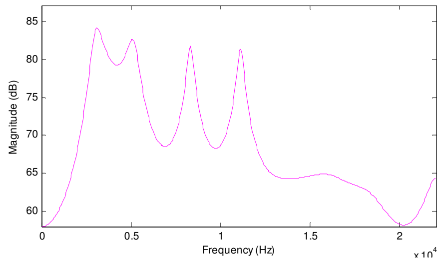
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



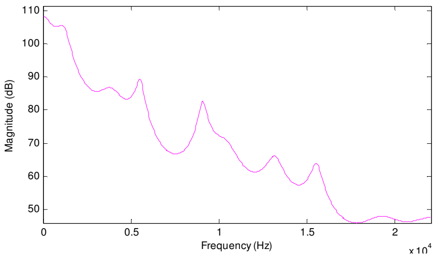
F1= 2756 Hz, F2= 5469 Hz F3= 7623 Hz



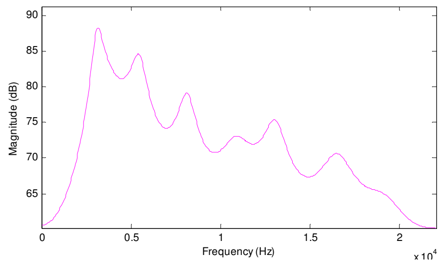
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 3101 Hz, F2= 5082 Hz F3= 8312 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 3144 Hz, F2= 5426 Hz F3= 8096 Hz

Figure 65 Spectrum plots of the SPLS acoustic simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW2=[2587 Hz, 14320 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

When we heard the simulation, the /t/ at the beginning can be heard and then we perceived an /i/ instead of a "ra". This can be explained by looking at the spectrum of the signal for 400 ms and 500 ms (Figure 65). The typical formant frequencies F1, F2, and F3 for the vowel /a/ are approximately 600 Hz, 1000 Hz, and 2500 Hz respectively, while for the vowel /i/ the first three formants are 200 Hz, 2300 Hz, and 3000 Hz. The formants found in these two plots are of around 3000 Hz, typical values for the third formant of /i/.

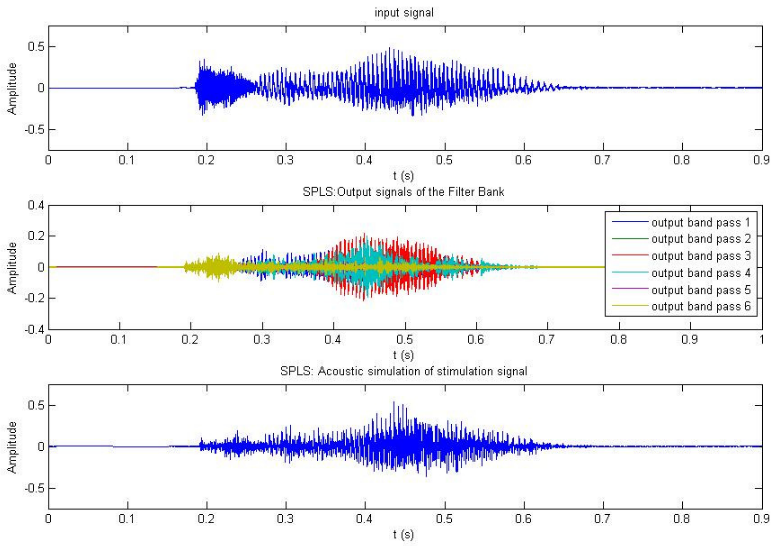


Figure 66 Results of the SPLS simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].

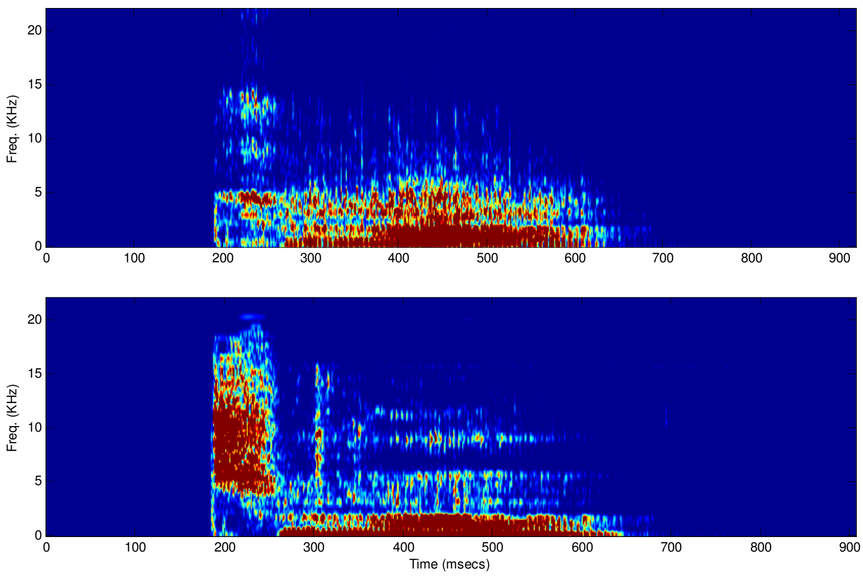
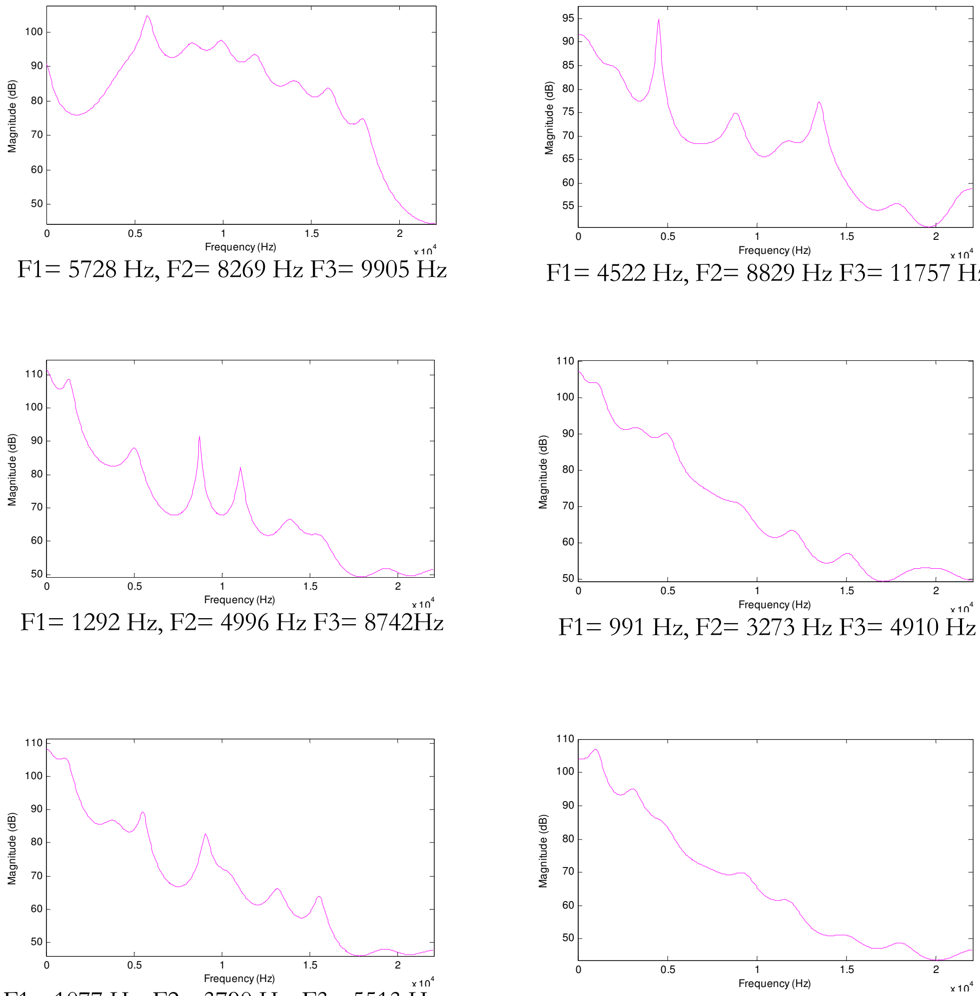


Figure 67 Top: Spectrogram of the SPLS acoustic simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz]. Bottom: Spectrogram of the original word.



F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz

F1= 4522 Hz, F2= 8829 Hz F3= 11757 Hz

F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz

F1= 991 Hz, F2= 3273 Hz F3= 4910 Hz

F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz

F1= 991 Hz, F2= 3058 Hz F3= 9130 Hz

Figure 68 Spectrum plots of the SPLS acoustic simulation of the syllable "tra" using 6 channels, FIR filters with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

For the simulation using the third frequency range, the syllable can be good recognized. As it is shown in Figure 67, the obtained wave sound contains minimal high frequency information, but it is enough to produce a good sensation to understand the /t/. The plot of Figure 68 at the top right, also shows that the spectrum at the beginning of the syllable is very different than the original one. For the voiced part of the syllable, the spectrum tends more and more to the original one. In this simulation, the voice of a female speaker can also be recognized.

Results by varying the number of channels

The three simulations, using 2, 4 and 8 channels, were done like for the CIS strategy, using the frequency range BW3=[215 Hz, 4891 Hz] and Butterworth IIR filters of 4^{th} order with uniform bandwidth.

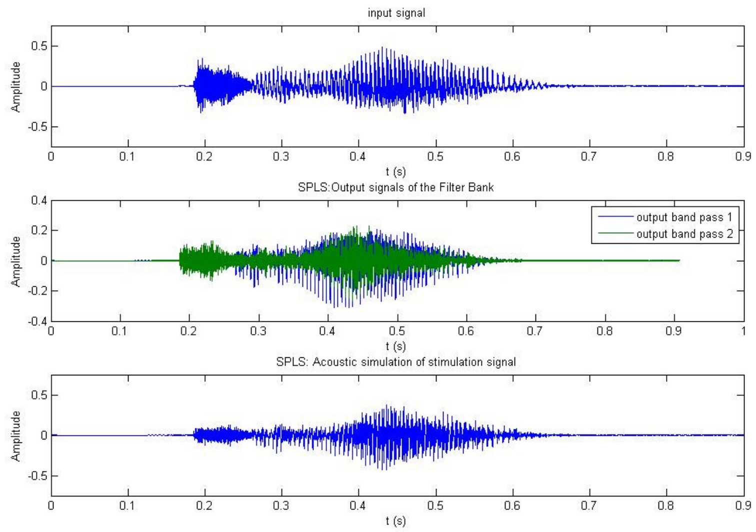


Figure 69 Results of the SPLS simulation of the syllable "tra" using 2 channels, IIR (Butterworth 4^{th} order) filters with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].

The acoustic simulations can be heard in the CD attached to this master thesis, under the SPLS simulations directory, with the names **sim_SPLS_2_RectBW_IIR_BW3.wav**, **sim_SPLS_4_RectBW_IIR_BW3.wav** and **sim_SPLS_8_RectBW_IIR_BW3.wav**.

For the simulation using only 2 channel, the SPLS performs strongly compared to the CIS strategy. It is possible without problems to understand and recognized the syllable. The spectrogram in Figure 70 shows the good reconstruction of the low frequency information, which as we have seen, is the most important for speech understanding.

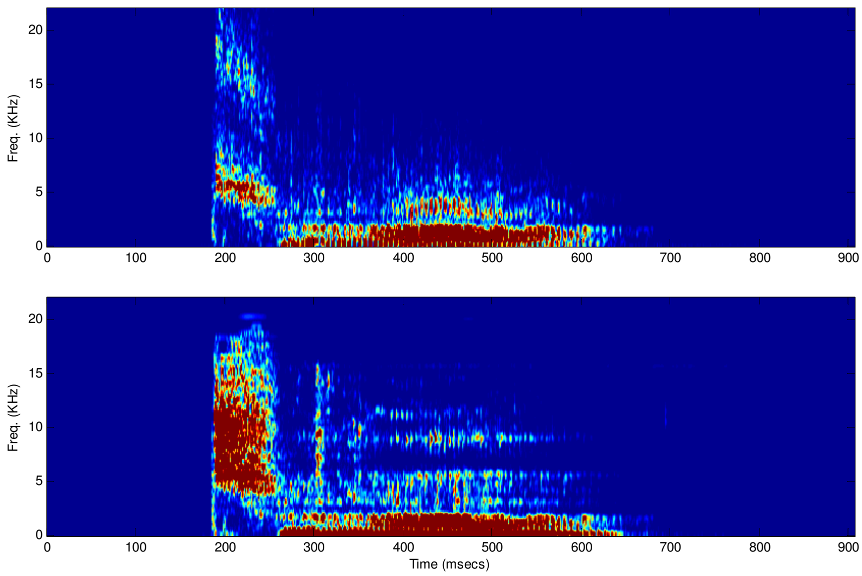
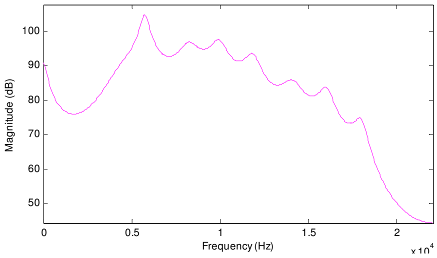
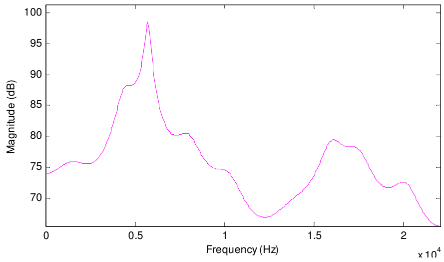


Figure 70 Top: Spectrogram of the SPLS acoustic simulation of the syllable "tra" using 2 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].. Bottom: Spectrogram of the original word.

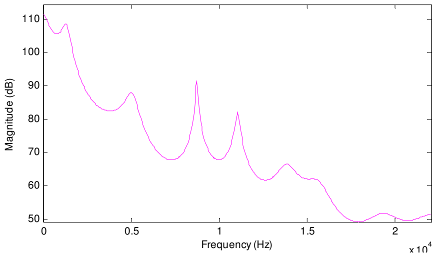
Although the low portion of high frequency information, due the limitation of the frequency range to 5000 Hz, the unvoiced sound can be perceived without problems. From the plots of the spectrum depicted in Figure 71, it can be seen how good the simulation performs in the voiced part of the sound. Compared to the CIS strategy, where nothing could be perceived at all, the spectrum in 400 ms and 500 ms have the same form as the original ones. Even the first formants are in the near of the original ones. Compare these two plots with the ones obtained for the same simulation using the CIS strategy in Figure 43.



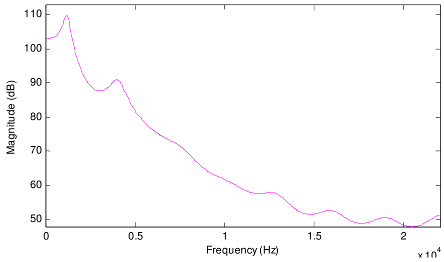
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



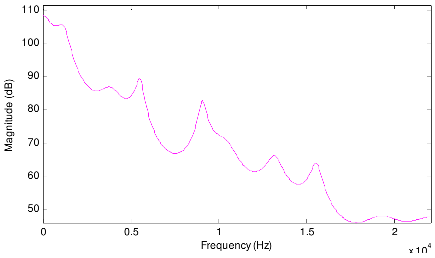
F1= 1550 Hz, F2= 4651 Hz F3= 5728 Hz



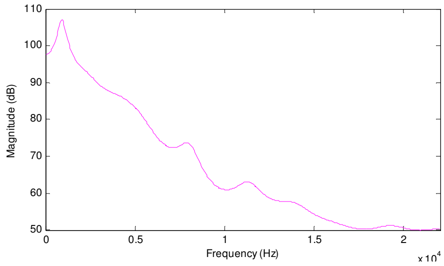
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 1163 Hz, F2= 4005 Hz F3= 12618 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 904 Hz, F2= 7881 Hz F3= 11283 Hz

Figure 71 Spectrum plots of the SPLS acoustic simulation of the syllable “tra” using 2 channels, IIR filters (Butterworth 4th order) with uniform bandwidth and using the frequency range BW3=[215 Hz, 4891 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

For 4 and 8 channels, the results obtained are qualitative very similar to the obtained using 2 channels as it can be seen when the spectrums of both simulations are considered, Figure 73 for 4 channels and Figure 76 for 8 channels. If we observed the plots of the spectrum for 200 ms, it can be seen that the peaks get broader when the number of channels increase (see Figure 74 Figure 77 top right) trying to expand the spectrum to get the original one, similar to a high band signal.

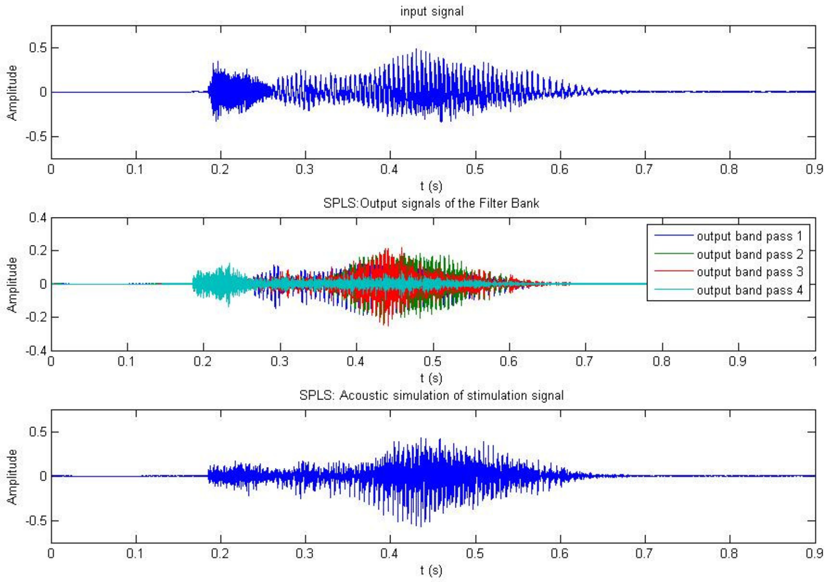


Figure 72 Results of the SPLS simulation of the syllable "tra" using 4 channels, IIR (Butterworth 4th order) filters with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].

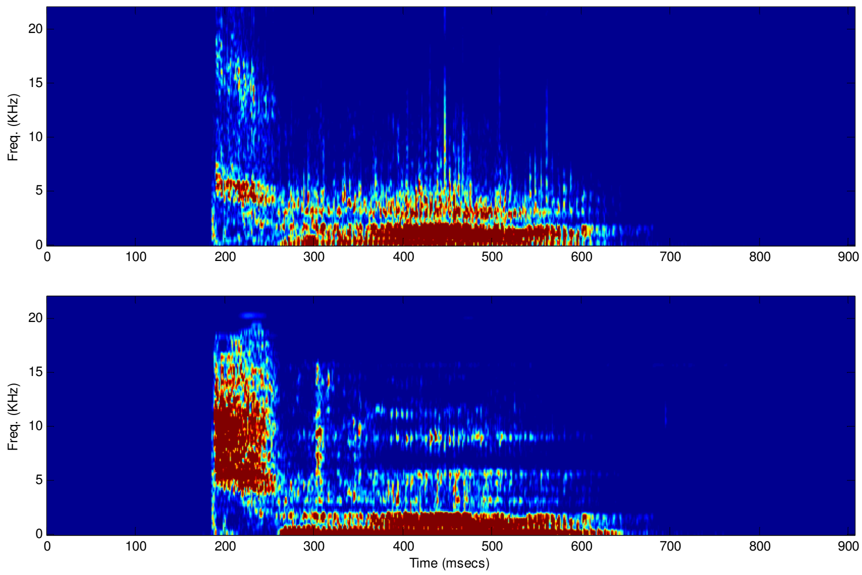
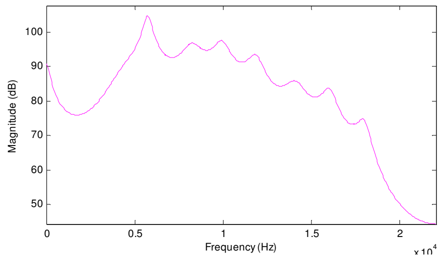
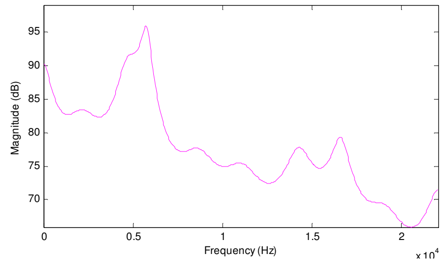


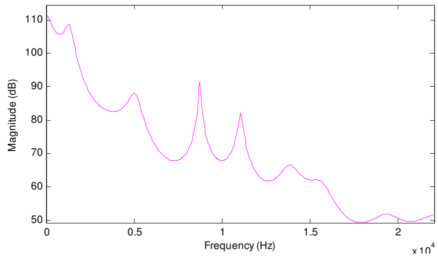
Figure 73 Top: Spectrogram of the SPLS acoustic simulation of the syllable "tra" using 4 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].. Bottom: Spectrogram of the original word.



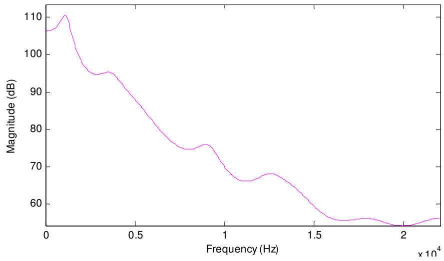
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



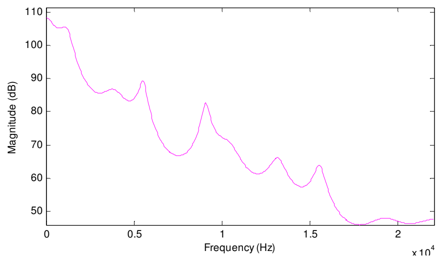
F1= 2153 Hz, F2= 5728 Hz F3= 8527 Hz



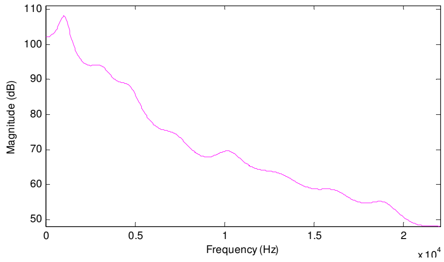
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 1077 Hz, F2= 3488 Hz F3= 8958 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 1034 Hz, F2= 2929 Hz F3= 10164 Hz

Figure 74 Spectrum plots of the SPLS acoustic simulation of the syllable “tra” using 4 channels, IIR filters (Butterworth 4th order) with uniform bandwidth and using the frequency range BW3=[215 Hz, 4891 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

For 400 ms and 500 ms, the spectrums get more exact information of the formants, when the number of channels increases (see Figure 74 and Figure 77 middle and bottom right plots). It is expected that the number of channels has its importance when speech has to be understand in a noisy environment or for music perception. In such cases, the more fine information added to the signal will play an important role. .

Test adding noise or using music sounds were for this master thesis out of our focus, because of the time limitation, but sure a very interesting point.

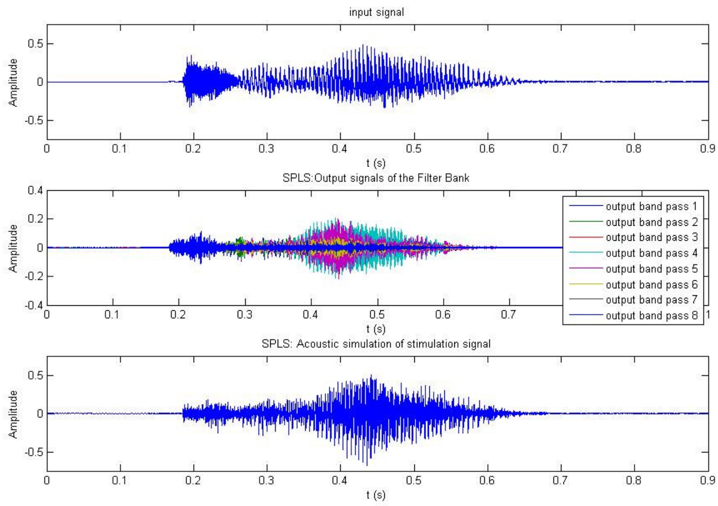


Figure 75 Results of the SPLS simulation of the syllable "tra" using 8 channels, IIR (Butterworth 4th order) filters with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].

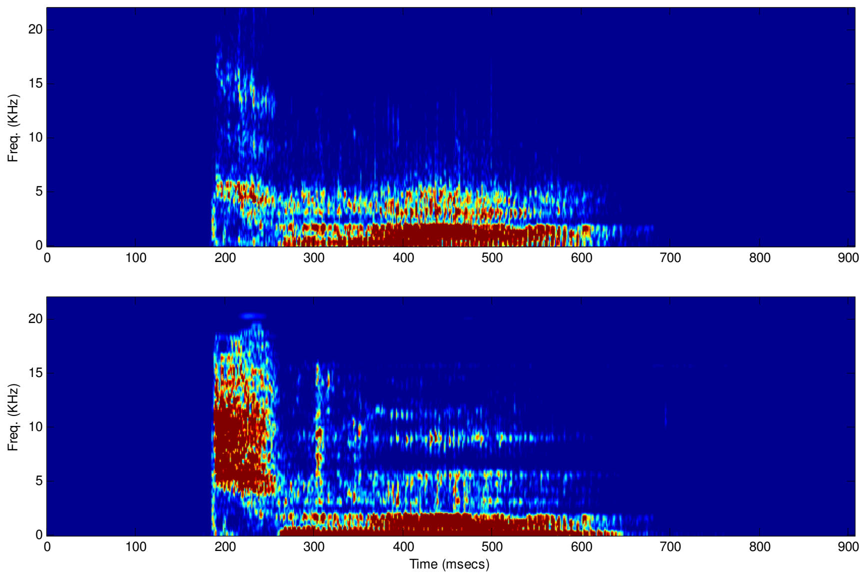
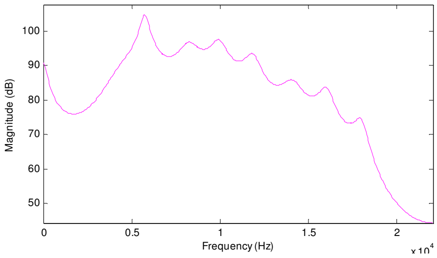
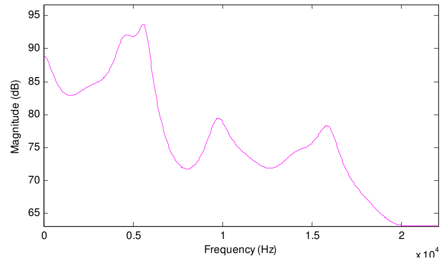


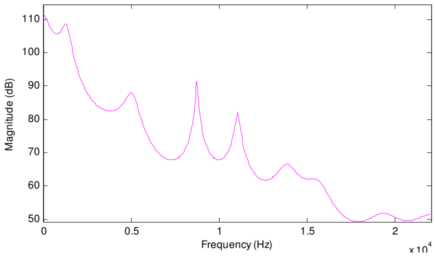
Figure 76 Top: Spectrogram of the SPLS acoustic simulation of the syllable "tra" using 8 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW3=[215 Hz, 4891 Hz].. Bottom: Spectrogram of the original word.



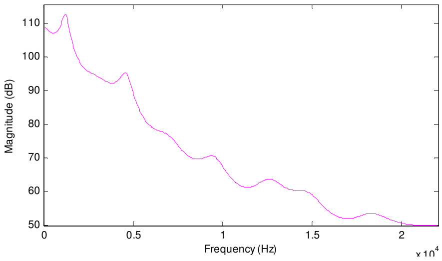
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



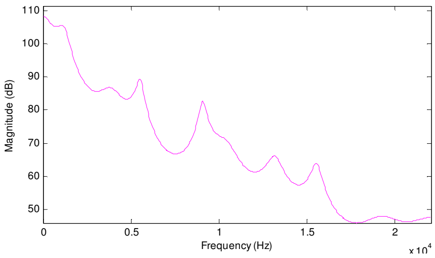
F1= 4651 Hz, F2= 5599 Hz F3= 9776 Hz



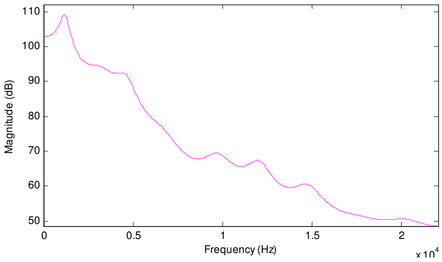
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 1206 Hz, F2= 4564 Hz F3= 9388 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 1163 Hz, F2= 4393 Hz F3= 9647 Hz

Figure 77 Spectrum plots of the SPLS acoustic simulation of the syllable "tra" using 8 channels, IIR filters (Butterworth 4th order) with uniform bandwidth and using the frequency range BW3=[215 Hz, 4891 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

Results by varying the type of filter

Like for the CIS strategy, to evaluate the effects of the type of filter, two simulations using 6 channels over the frequency range BW1=[603 Hz, 12986 Hz] and uniform bandwidths were done. The results using FIR filter were already shown in Figure 60, Figure 61 and Figure 62. For the IIR filter simulation, in order to conserve the phase information, which is the one that decides the pulse firing, zero-phase filtering was achieved by forward-backward filtering. This method is a very usual one, when the signal can be filtered "offline" (the signal is stored and can be filtered at once without time constraints). The input signal will be first filtered with the iir recursive filter. The output of this first stage will be flipped in time. Then the flipped signal will be filtered again with the recursive iir filter. The final output is then this result flipped in time. An scheme of this method is depicted in Figure 78.

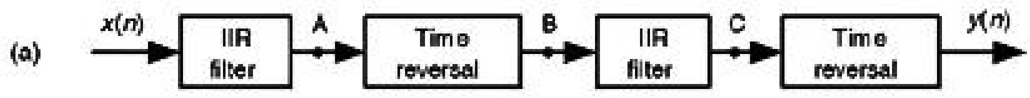


Figure 78 Block diagram of the forward-backward zero-phase filtering method.

To show this analytically, let denote $x(n)$ to be our input signal, $v(n)$ to be the output of the first filter stage (forward stage) at point A. Let denote $h(n)$ to be the impulse response of the iir filter. Then we know that

$$v(n) = h(n) * x(n)$$

At point C, let denote $w(n)$ the result signal of the backward stage. This signal is

$$w(n) = h(n) * v(-n)$$

The final output signal $y(n)$ is defined as

$$y(n) = w(-n) = h(-n) * v(n)$$

Using the z-transform and its properties, we can write

$$Y(z) = H(1/z) \cdot H(z) \cdot X(z)$$

On the unit circle where $z = e^{jw}$,

$$Y(e^{jw}) = H(e^{-jw}) \cdot H(e^{jw}) \cdot X(e^{jw})$$

It is important to remark that

$$H(e^{jw}) = | H(e^{jw}) | \cdot e^{j\phi(H(e^{jw}))} \quad \text{and} \quad H(e^{-jw}) = | H(e^{jw}) | \cdot e^{-j\phi(H(e^{jw}))}$$

Using these two relations, we obtain that

$$Y(e^{jw}) = | H(e^{jw}) |^2 \cdot X(e^{jw})$$

That means, the output signal conserves the phase of the input signal, achieving a zero-phase filtering.

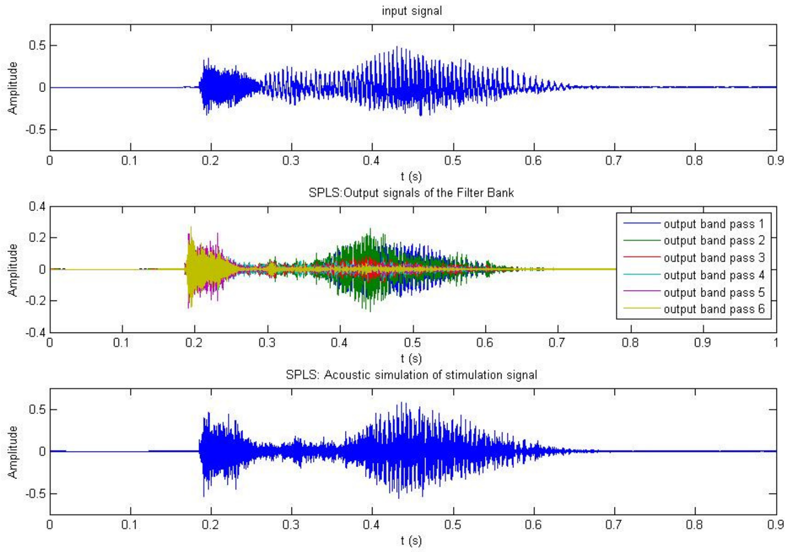


Figure 79 Results of the SPLS simulation of the syllable "tra" using 6 channels, IIR (Butterworth 4th order) filters with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz].

Both simulations have a good sound quality and the syllable can be perceived easily in both cases. From the spectrograms shown in Figure 61 for the FIR filters and Figure 80 for the IIR filter, no significant differences can be found.

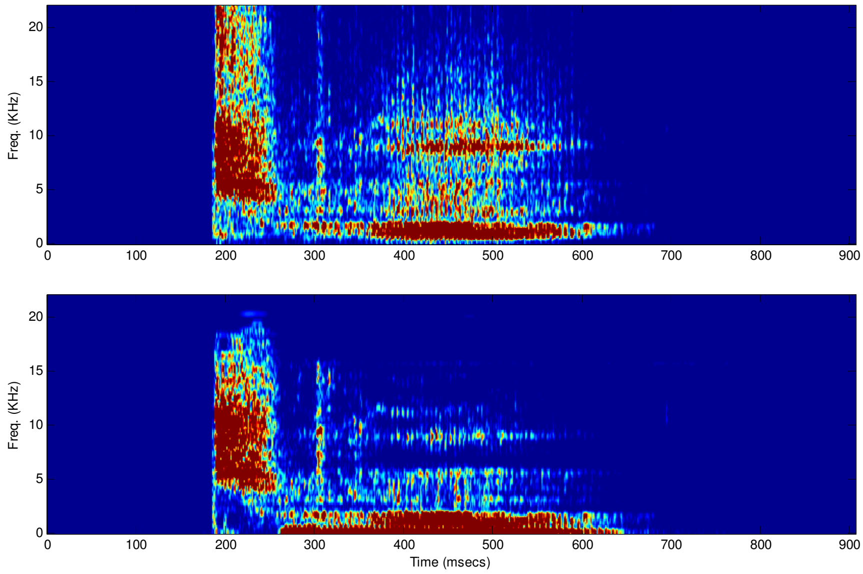
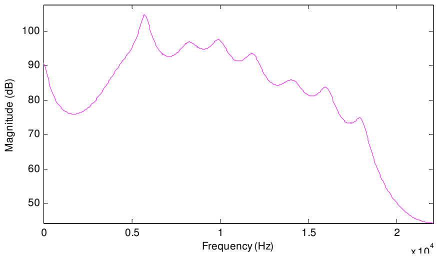


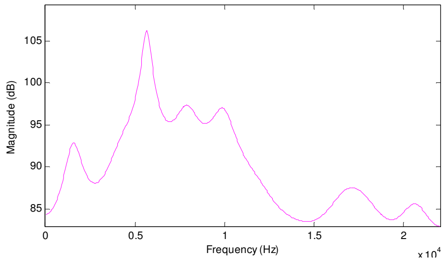
Figure 80 Top: Spectrogram of the SPLS acoustic simulation of the syllable "tra" using 6 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz]. Bottom: Spectrogram of the original word.

However, observing the spectrums of the simulations, in special for the time 400 ms and 500 ms, the IIR simulations provide more similar spectrums than the simulations using FIR filters. The formants are also more similar to the original ones in the IIR simulation.

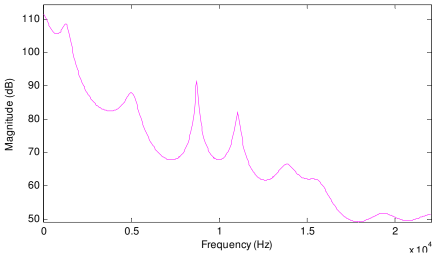
Both simulations, compared to the CIS wave sounds, are much better. The wave sounds can be found in the CD attached to the master thesis, under the SPLS simulations directory, with the names s**im_SPLS_6_RectBW_FIR_BW1.wav** and **sim_SPLS_6_RectBW_IIR_BW1.wav**.



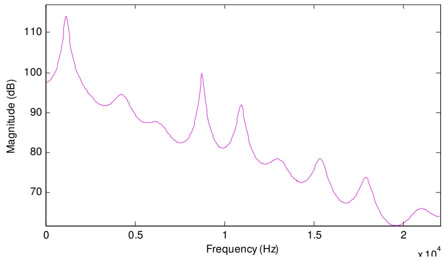
F1= 5728 Hz, F2= 8269 Hz F3= 9905 Hz



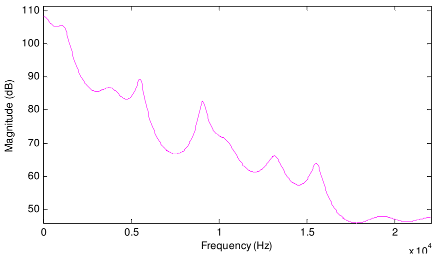
F1= 1593 Hz, F2= 5642 Hz F3= 7881 Hz



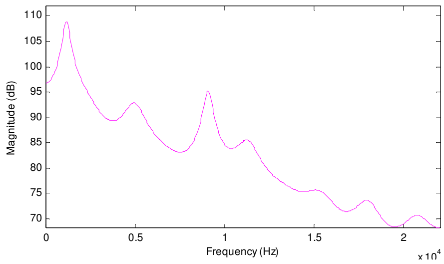
F1= 1292 Hz, F2= 4996 Hz F3= 8742Hz



F1= 1163 Hz, F2= 4264 Hz F3= 6115 Hz



F1= 1077 Hz, F2= 3790 Hz, F3= 5513 Hz



F1= 1163 Hz, F2= 4953 Hz F3= 9087 Hz

Figure 81 Spectrum plots of the SPLS acoustic simulation of the syllable "tra" using 6 channels, IIR filters (Butterworth 4th order) with uniforms BW and using the frequency range BW1=[603 Hz, 12986 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

<u>Results varying the type of filter's bandwidth</u>

To evaluate the effects of the type of filter's bandwidth, again two simulations using 6 channels, the frequency range BW1=[603 Hz, 12986 Hz] and Butterworth 4th order IIR filter were done. The results of the simulation using uniform bandwidths are depicted in Figure 79, Figure 80 and Figure 81.

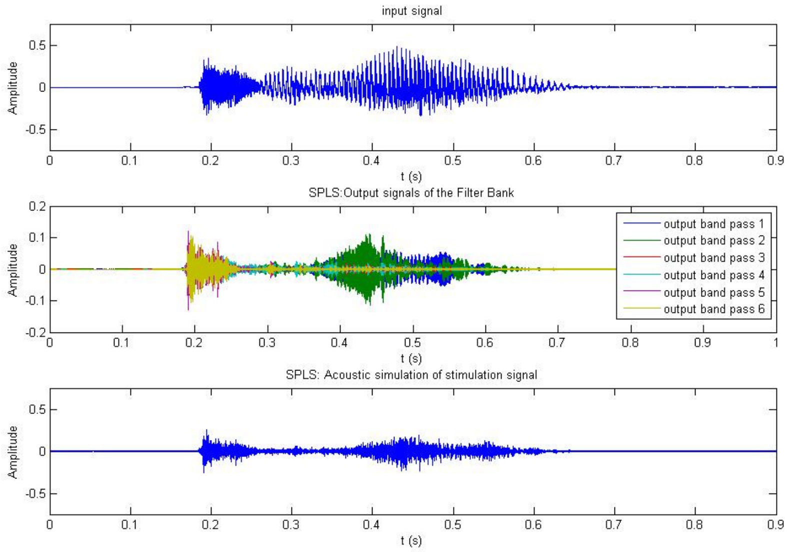


Figure 82 Results of the SPLS simulation of the syllable “tra” using 6 channels, IIR (Butterworth 4th order) filters with ERB and using the frequency range BW1=[603 Hz, 12986 Hz].

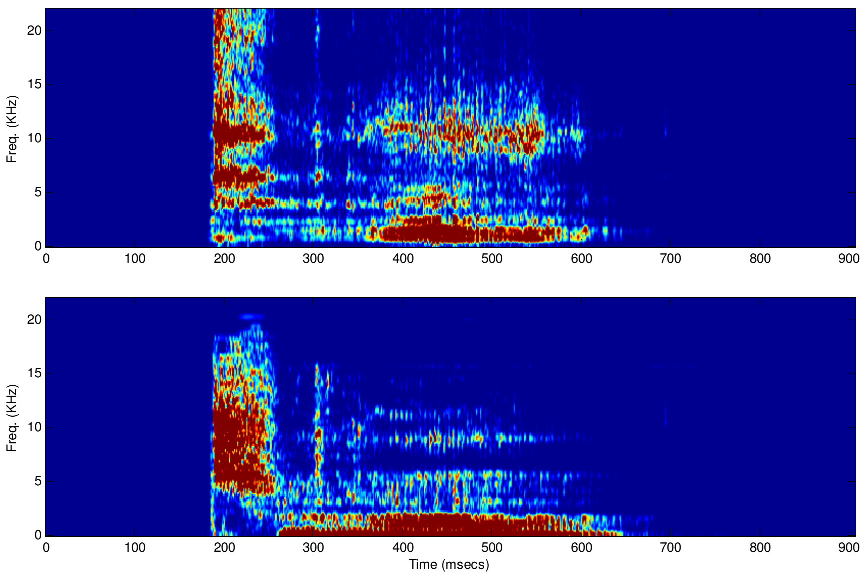


Figure 83 Spectrogram of the SPLS acoustic simulation of the syllable “tra” using 6 channels, IIR filters (Butterworth 4th order) with ERB and using the frequency range BW1=[603 Hz, 12986 Hz]. Bottom: Spectrogram of the original word.

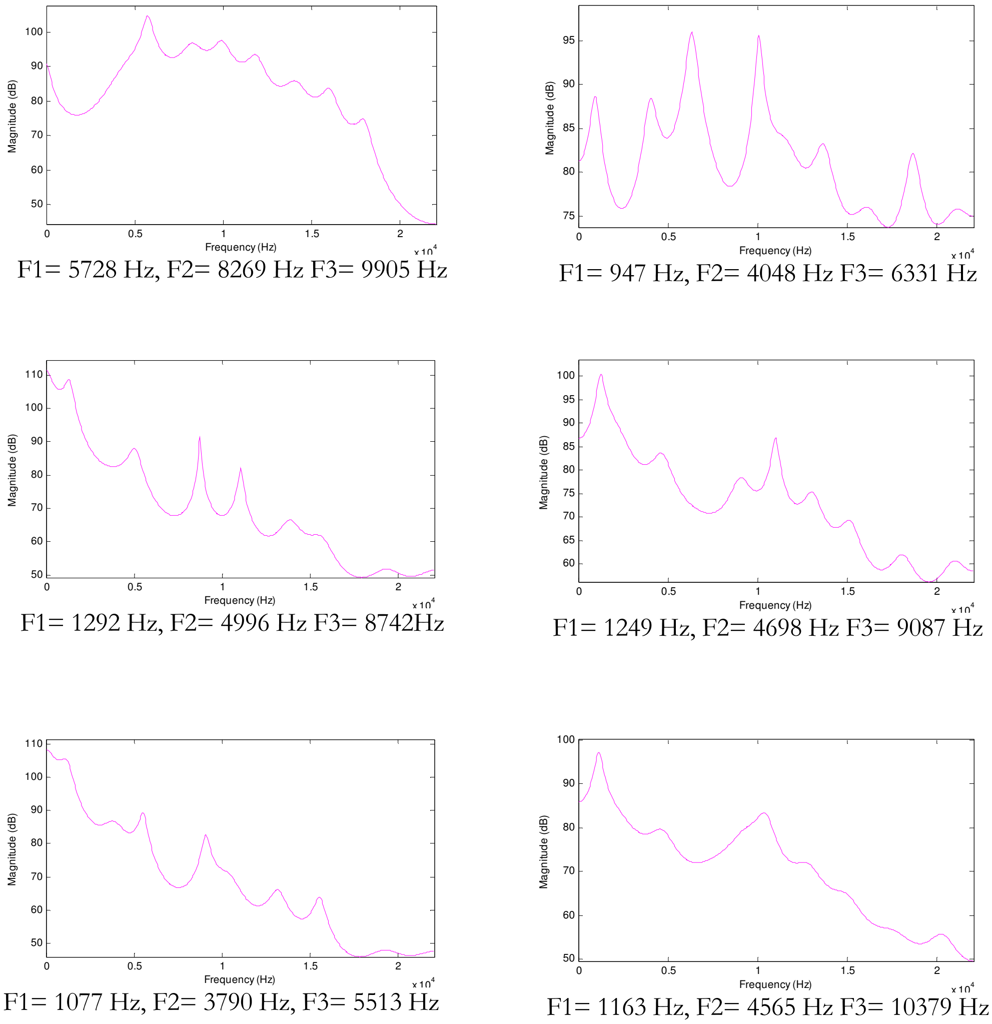


Figure 84 Spectrum plots of the SPLS acoustic simulation of the syllable "tra" using 6 channels, IIR filters (Butterworth 4th order) with ERB and using the frequency range BW1=[603 Hz, 12986 Hz] (left column) versus original signal (right column). Top row: time 200ms. Middle row: time 400ms. Bottom row: time 500ms.

The results obtained by the SPLS simulation using ERB filters, due to the high selectivity of the filters, don't sound better than the acoustic simulation using uniform bandwidths. The gaps produced between the bands make the acoustic signal to loose important components in frequency, which do the signal to be less intelligible. The syllable can still be understood, but it is not possible to recognize if the signal is from a male or female speaker. This effect can be seen in the plots shown in Figure 84. The spectrum is for example, for the time 200 ms (top right plot) totally distorted. For the other two times, 400 ms and 500 ms, the spectrums are also more different than the original ones compared with the simulation using the uniform bandwidths (Figure 81). As we already mention in the CIS simulations, for low number of channels, the use of ERB filters doesn't perform a better acoustic simulation than the use of uniform bandwidth filters.

The acoustic simulations can be heard in the CD attached to this master thesis, under the CIS simulations directory, with the names **sim_SPLS_6_RectBW_IIR_BW1.wav** and **sim_SPLS_6_ERB_IIR_BW1.wav.**

Chapter 4: Test results and discussion

In this chapter, the results of the "Heidelberger Laudifferenzierungstest" (H-LAD) using different CIS and SPLS simulations will be presented. An explanation of this test and what it is for will be done in the corresponding subsection. The results will be depicted in simple graphs to be able to compare the results easily. In the second part of this chapter, the percentage of the simultaneous firing between neighbor electrodes, calculated during different simulations of the syllable "tra" , simulations of a piece of music and simulations of speech in noise using the SPLS strategy are outlined.

Results of the HLAD Test

The H-LAD Test is used to check if children speaking the German language have a sound analysis or differentiation impairment and the severity of such debility. The test is the result of a common project developed by the "dyslexia workgroup" and the department of vocal, speech and hearing disorders of the Clinic University of Otolaryngology Heidelberg and the department of Child and Adolescent Psychiatry of the Ruprecht-Karks-University Heidelberg. In search of dyslexia reasons, it is sure that there is not an exclusive factor, but more causes are responsible for the children's difficulties in reading and writing. It is thought that one of these factors is the impairment of the phonological processing. This debility can appear in one of the three levels of such processing mechanism; at the auditory analysis and discrimination stage, at the phonological temporary storage or/and in the conversion of the phonological information to articulations mechanisms.

In the practice, lot of normal hearing children having problems in reading and writing correctly in the German language, have big problems trying to differentiate similar sounds and to articulate them. This is not a specific problem of children with dyslexia; it is also present in children with speech development disorders. In order to get the right therapy for such children, a test that is able to diagnose such problems was necessary.

The H-LAD Test consists of two subtests. The first subtest checks the ability to differentiate consonants and the second subtest checks the ability of analysis and differentiation of consonants accumulation in a word. The first subtest consists itself in three subsets. The first two subsets check the differentiation ability of the children in words and syllables which have consonants that are produced with a different articulation mode but in the same place of articulation. The third subset consists of differentiating pairs of words with consonants that are produced in a different place of articulation but with the same articulation mode.

We have worked with subsets 1a and 1b of the first subtest. The words pairs and syllable pairs used for the test are listed in Table 1. All these word and syllable pairs were simulated with the CIS and SPLS strategies using different parameters for the number of channels and type of filter's bandwidth. The frequency range and the type of filter remain for all the same. These two parameters were selected to meet the same conditions used in the tests done by Chen et al. and presented in his paper (Chen et al., 2009), BW3=[215 Hz,4891Hz] and IIR Butterworth 4th order filters. The CIS simulations using 4 channels and uniform bandwidth filters and using 8 channels and ERB filters were chosen to test the performance of the CIS strategy. The SPLS simulations using 2 channels with uniform bandwidth filters and using 4 channels with ERB filters, were the selected ones to test the performance of the SPLS strategy. All these simulations can be heard and used for test purposes in the CD attached to the master thesis, under the directory H-LAD Test Data. Name and test content are described in Table 2.

The test procedure is as follows: the subject hears the pair of words (in test 1a) or pair of syllables (in test 1b) and has to answer if what he/she heard was the same word or syllable or different. Both cases are present in the test, as it can be seen in the Table 1.

1A. Words for differentiation of consonants with a different articulation mode, the same place of articulation		**1B Syllables for differentiation of consonants with a different articulation mode, the same place of articulation**	
Kuß	Guß	tra	tra
Reisen	Reißen	kra	gra
Dreck	Dreck	gla	kla
Kriechen	Griechen	dra	tra
Gasse	Kasse	bra	pra
Paß	Paß	da	ta
Kern	gern	ba	ba
Seide	Seite	ka	ga
Dreck	Treck	ba	pa
Blatt	Platt		
Kord	Kocht		
Gasse	Gasse		
Baß	Paß		
Scharrt	Schacht		
Klette	Glätte		

Table 1.List of words and syllables used in the subtest 1a and 1b.

Wav file name	Contents of the wav file
A_1_full_4_RectBW_CIS.wav	CIS acoustical simulation using 4 channels and uniform filter's bandwidths of subtest 1a, in the order listed above.
B_1_full_8_ERB_CIS.wav	CIS acoustical simulation using 8 channels and ERB filters of subtest 1a, in the inverse order listed above.
A_1_full_4_ERB_SPLS.wav	SPLS acoustical simulation using 4 channels and ERB filters of subtest 1a, in the order listed above.
B_1_full_2_RectBW_SPLS.wav	SPLS acoustical simulation using 2 channels and uniform filter's bandwidths of subtest 1a, in the inverse order listed above.
A_2_full_4_RectBW_CIS.wav	CIS acoustical simulation using 4 channels and uniform filter's bandwidths of subtest 1b, in the order listed above.
B_2_full_8_ERB_CIS.wav	CIS acoustical simulation using 8 channels and ERB filters of subtest 1b, in the inverse order listed above.
A_2_full_4_ERB_SPLS.wav	SPLS acoustical simulation using 4 channels and ERB filters of subtest 1b, in the order listed above.
B_2_full_2_RectBW_SPLS.wav	SPLS acoustical simulation using 2 channels and uniform filter's bandwidths of subtest 1b, in the inverse order listed above.

Table 2.Wav file names of the simulations and their contents

Then, they are asked to say what they have heard again. The first part, the answer if it is the same or different words or syllables, is the auditory part of the test. The second part, when they are asked to repeat what they have heard, is the kinesthetic part of the test. For the H-LAD test is important to get both values in order to find in which of the mentioned three stages of the phonological processing may the disorder be. For our purpose, the kinesthetic part tells us the degree of words/syllables recognition.

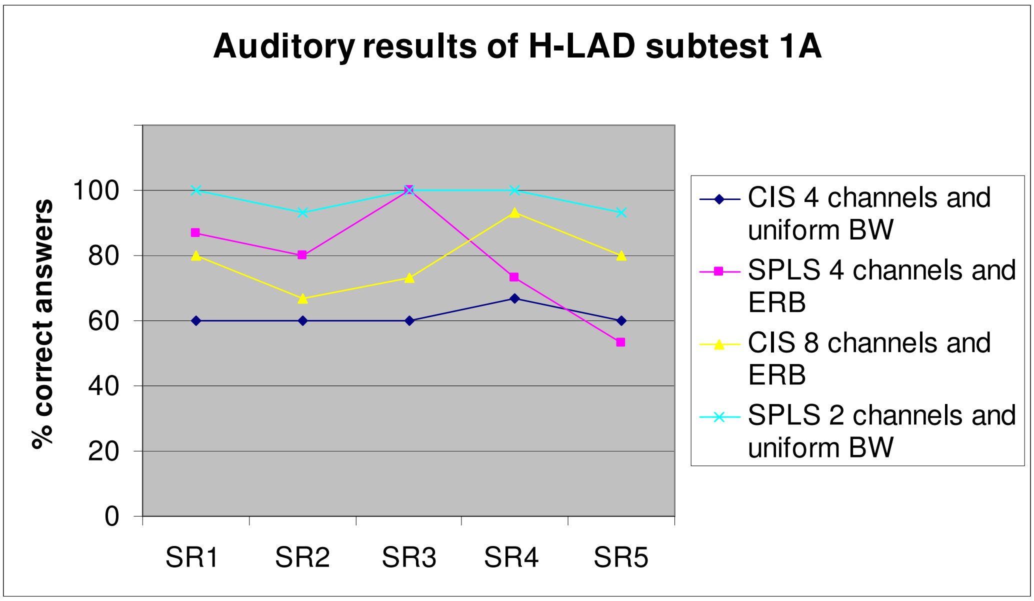


Figure 85 Results of the auditory part of the H-LAD subtest 1a.

In Figure 85 and Figure 86 are the results of the subtest 1a and in Figure 87 and Figure 88 are the results of the subtest 1b. As it can be observed in all graphs, the best results are obtained using the acoustic simulation of the SPLS strategy with 2 channels and filters with uniform bandwidth. In most of the subjects, the auditory highest score is reached, and in the kinesthetic part, the results using this simulation are much better than using the other three simulations. It has to be remarked, that the results of the acoustic simulation of the SPLS strategy using 4 channels, but ERB filters, do not reach the great results of the simulation of the same strategy using lower number of channels but using the uniform bandwidth. As it was mentioned in chapter 3, the use of ERB filters is not appropriate for the strategy. They filter out too much important information of the frequencies between the bands. Even the results of the simulation of the CIS strategy, using 8 channels and ERB filters produces better results than the SPLS 4 channels and ERB in the kinesthetic part of the test.

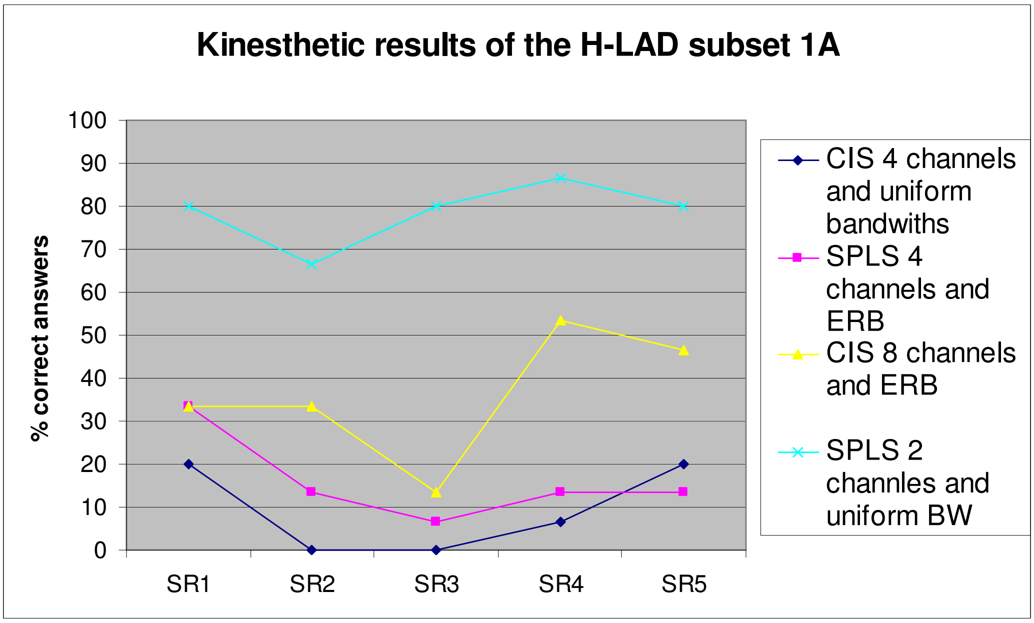


Figure 86 Results of the kinesthetic part of the H-LAD subtest 1a.

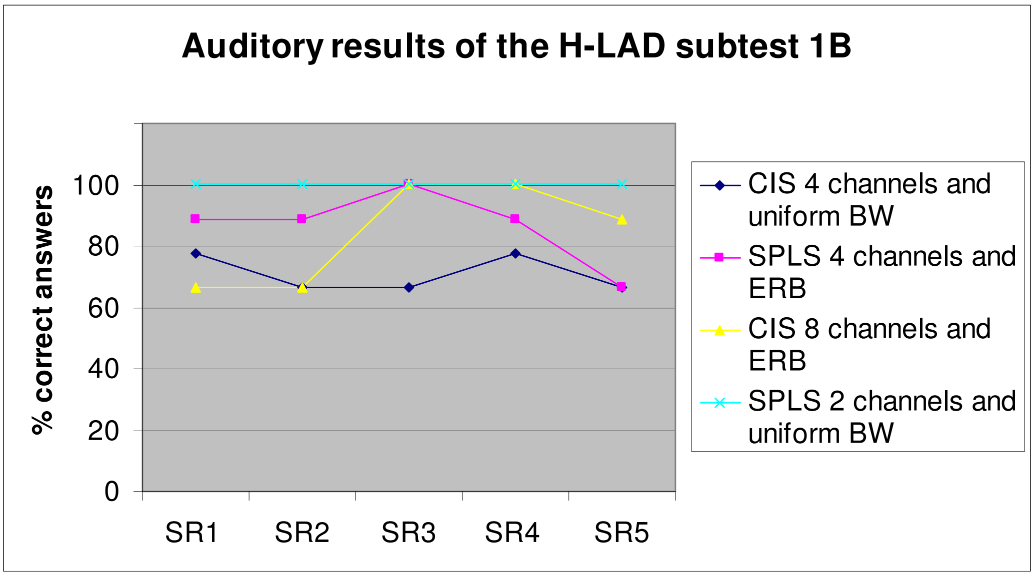


Figure 87 Results of the auditory part of the H-LAD subtest 1b.

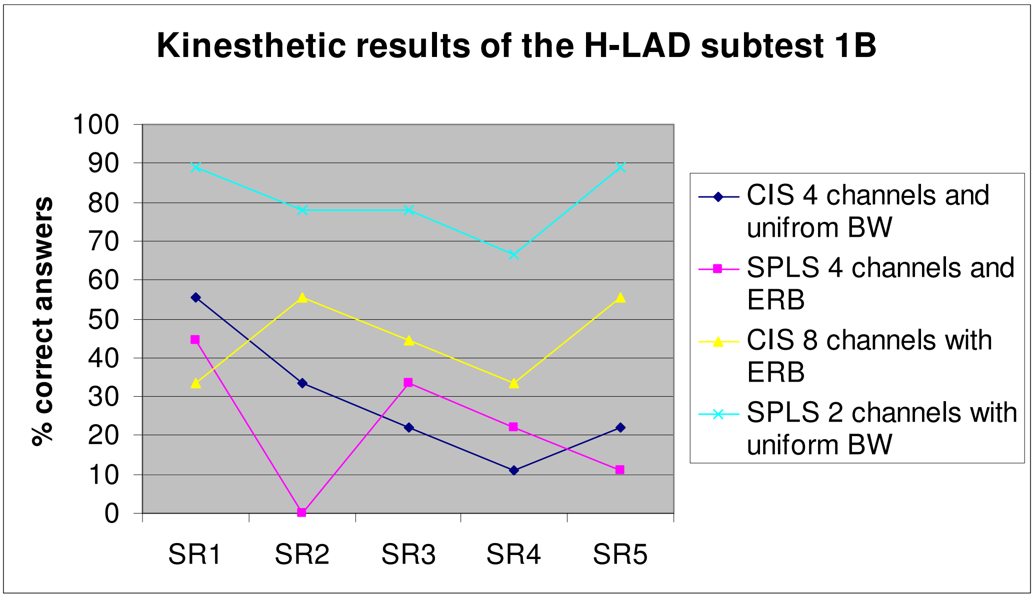


Figure 88 Results of the kinesthetic part of the H-LAD subtest 1b.

An observation should be made in how the test was done. Most of the subjects had problems in recognizing words or syllables, specially using the CIS strategy. The metallic sound was for most of them very uncomfortable and the pair of words or syllables had to be played twice or more times. However, the SPLS strategy was more comfortable for their ears and they could start to recognize the words or syllables. In special with the simulation of the SPLS with 2 channels and filters with uniform bandwidths, the subjects could respond at once, fast and easily to the test.

Results of the simultaneous stimulation in the SPLS strategy

As it was remark in chapter 2, some measurements were done by Chen et al. and presented in Chen et al, 2009 about simultaneous firing of electrodes in the SPLS strategy. They processed a 49-second piece of sound (including male or female English speech, Chinese speech, and a piece of music) with an 8 channels SPLS strategy. A final percentage of 1.9% of simultaneous firing between 2 adjacent channels was calculated. In order to prove this statement in the case of German language, this measurement was done for every simulation done in chapter 3, and also for a piece of music of 14 s and for a piece of the H-LAD test with noise of 25 s. In all cases only the simultaneous firing of neighbor electrodes was calculated. The highest results obtained in each simulation are listed in Table 3.

Simulation	Highest percentage of interactions
SPLS_6_RECTBW_FIR_BW1	Channel 2 and 3 = 2.539988%
SPLS_6_RECTBW_FIR_BW2	Channel 2 and 3 = 1.495853%
SPLS_6_RECTBW_FIR_BW3	Channel 4 and 5 = 1.448953%
SPLS_2_RectBW_IIR_BW3	Channel 1 and 2 = 1.227500%
SPLS_4_RectBW_IIR_BW3	Channel 2 and 3 = 2.125000%
SPLS_8_RectBW_IIR_BW3	Channel 5 and 6 = 4.072500%
SPLS_6_REctBW_IIR_BW1	Channel 2 and 3 = 4.002500%
SPLS_6_ERB_IIR_BW1	Channel 2 and 3 = 3.140000%
longsong iir 8 channels RectBW (15 s)	Channel 5 and 6 = 4.323928%
longsong fir 8 channels RectBW (15s)	Channel 5 and 6 = 2.981215%
track5 fir 8 channel uniform (25 s)	Channel 5 and 6 = 4.160727%
track5 iir 8 channel uniform (25 s)	Channel 5 and 6 = 2.643769%

Table 3. Results of the measurements of simultaneous firing in the SPLS strategy

From the results listed above, it can be observed that in the worst case, a simultaneous stimulation between adjacent electrodes happened 4,3% of the time.

This value is close to the calculated by Chen et al. and it is too small to develop extra inhibitory procedures mechanism for the strategy to avoid this effect.

Conclusion

In chapter 2 the development of the speech strategies in the last decades was exposed. It could be read that for the multichannel strategies, it was thought that increasing the number of channels would increase the performance of the strategy. Even with the introduction of virtual channels, the number of possible elicited positions could be raised up to 22 sites. However, it looks like that for the new strategies, which extracts more fine structure from the signal and uses this information to modulate the firing muster; the number of channels doesn't play such an important role as it was expected. It seems to be more important for the SPLS strategy in order to achieve good speech recognition to get broader filters to obtain more phase information per channel, than to get more channels with narrower filters and therefore lower phase information for each band. That is the explanation why we got better test results for the simulation of the SPLS using 2 channels and filters with uniform bandwidths than for the simulation of the same strategy using more channels but with the ERB filters.

For me, it is still open how reliable is the acoustical simulation proposed by Chen et al. in their paper. Their proposal, exposed in page 49, is based on the signal obtained in the zero-crossing stage. For the electrical stimulation, they perform a kind of decimation of the zero-crossings events, in order to decrease the firing rates for the high frequency channels and avoid the not desirable simultaneous electrode stimulation. But for the acoustic simulation, all zero-crossings events are being taken in account. This inconsistency makes me have some doubts of the great expected performance of the SPLS strategy using the schema proposed for the electrical stimulation. It would be very interesting to generate new simulations of the SPLS strategy using the decimated zero-crossings event signal to evaluate its acoustical performance.

BIBLIOGRAPHY

Boex C, Pelizzone M, and Montandon P, "*Improvements in speech recognition with the CIS strategy for the Ineraid multichannel intracochlear implant*," in Advances in cochlear implants (I. Hochmair-Desoyer and E. Hochmair, eds.), pp. 136-140, Vienna: Manz, 1994

Borden GJ, Harris KS, and Raphael LJ, *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Baltimore, Maryland: Williams and Wilkins, 1994

Chen J; Wu X; Li L; Chi H, 2009.*Simulated Phase-Locking Stimulation: An Improved Speech Processing Strategy for Cochlear Implants.* Karger, ORL :71: Pages 221-227.

Clark GM. The University of Melbourne-Nucleus multielectrode cochelar implant. *Adv. Otorhinolaryngol.*;38:1-189. Basel:Karger. 1987.

Cooper FS, Delattre PC, Liberman AM, Borst JM, and Gerstman LJ*, Some experiments on the perception of synthetic speech sounds*, Journal of the Acoustical Society of America, vol. 24, no. 6, pp. 597-606, November 1952

Donaldson GS, Kreft HA, Litvak L. *Place-pitch discrimination of single-versus dual-electrode stimuli by cochlear implant usere (L).* J Acoust Soc Am. 2005; 118(2):623-26.

Dorman M, Hannley M, Dankowski K, Smith L, and McCandless G, *Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant.* Ear and Hearing, vol. 10, pp. 44-49, 1989.

Dorman MF, Loizou PC, Spahr AJ et al. *A comparison of the speech understanding provided by acoustic models of fixed-channel and channel picking signal processors for cochlear implants.* J Speech Lang Hear Res 45: 783-8, 2002

Fant G*, Acoustic Theory of Speech Production*. The Hague, Netherlands: Mouton, 1970.

Firszt JB, Koch DB, Downing M, Litvak L. *Current Steering creates additional pitch percepts in adult cochlear implant recipients.* Otol Neurotol. 2007; 28(5):629-36

Glasberg BR, Moore BCJ. *Derivation of auditory filter shapes from notched noise data.* Hear Res 47: 103-38, 1990

Glasberg BR, Moore BCJ. *A revision of Zwicker's loudness model.* Acta Acustica*,* vol. 82, pp. 335-345, 1996.

Goldwyn JH, Bierer SM, Arenberg Bierer J. *Modeling the electrode-neuron interface of cochlear implants: Effects of neural survival, electrode placement, ans the partial tripolar configuration.* Hearing Research 268 (2010): 93-104

Graeme C. *Cochlear Implants: Fundamentals and Applications*. Springer, 2003.

Graeme MC. 2006. *The multiple-channel cochlear implant: the interface between sound and the central nervous system for hearing, speech and language in deaf people- a personal perspective.* Royal Society, Vol. 361 no. 1469: 791-810

Greenwood DD, *A cochlear frequency position function for different species: 29 years later.* J Acoustic Soc. Am 87: 2592-606. 1990

Helmholtz H. 1863. *Die Lehre von der Tonempfindungen als physiologische Grundlage für die Theorie der Musik..* Vieweg. Braunschweig (Eng. Translation: A. J. Ellis, On the sensation of tones, 1885).

Kiefer J, Hohl S, Sturzebecher E, Pfennigdorff T, Gstoettner W;*Comparison of speech recognition with different speech coding strategies (SPEAK, CIS and ACE) and their relationship to telemetric measure of compound action potential in the nucleus CI 24M cochlear implant system.* Audiology, 2001, 40: 32-42 109(11):1009-20

Koch DB, Downing M, Osberger MJ, Litvak L. *Using current steering to increase spectral resolution in CII and HiRes 90K users.* Ear Hear. 2007; 28(2 Suppl): 39S-41S

Krishnan A, Xu Y, Gandour JT, Cariani PA. *Human frequency-following response: representation of pitch contours in Chinese tones.* Hearing Research 189, 2004,1-12

Lawson DT, Wilson BS, Zerbi M, Finley CC. *Speech Processings for auditory prostheses.* Third Quarterly Progress Report, NIH Project. N01-DC-5-2103. Bethesda (MD). Neural Prosthesis Program. National institute of Health. 1996

Litvak L, Overstreet E, Mishra L. *Steering current through simultaneous activation of intracochlear electrodes in the Clarion CII cochlear implant: Frequency resolution.* In : *Proceedingsof the 2003 Conference on Implantable Auditory Prosthese. 2003*

Loizou PC. *Mimicking the human Ear: An Overview of signal-Processing Strategies for Converting Sound into Electrical Signals in Cochlear Implants*. IEEE Signal Processing Magazine, September 1998. Pages 101-130

Nogueira W, Litvak L, Edler B, Ostermann J, Büchner A. *Signal Processing Strategies for Cochlear Implants using Current Steering.* EURASIP Journal on Advances in Signal Processing. Hindawi Publishing Corperation. Volume 2009

Oppenheim AV, Schafer RW. *Discrete-time Signal Processing.* Prentice Hall Signal Processing Series. Chapter 10. 1989

Poroy O, Loizou PC. *Pitch perception using virtual channels.* In: *Proceedings of the 2001 Conference on Implantable Auditory Prostheses.* 2001

Rattay F; Lutter P, 1997. *Speech Sound Representation in the Auditory Nerve: Computer Simualtion Studies on Inner Ear Mechanism.* ZAMM, Z. angew. Math. Mech. 77 12, 935-943

Roggero MA, Robles L, Rich NC, Costalupes JA. *Basilar Membrane motion and spike initiation in the cochlear nerve*: in Moore, Patterson (eds): *Ausitory Frequency Sekectivity.* New York, Plenum, 1986

Rutherford W. 1886. *A new theory of hearing.* J. Anat. Physiol, 21, 166-168.

Smith III JO; Abel JS. *Bark and ERB Bilinear Transforms.* IEEE Transactions on Speech and Audio Processing, December 1999.

Smith ZM; Delgutte B; Oxenham AJ. *Chimearic sounds reveal dichotomies in auditory perception.* Nature, 2002.

Stone MA, Moore BCJ, Alcántara JI, Glasberg BR. *Comparison of different forms of compression using wearable digital hearing aids.* Journal of the Acoustical Society of America, 1999. Vol. 106, no. 6; pp: 3603-3619.

Wever EG, 1954. *Theory of hearing.* Dover, New York

White M., Merzenich M, and Gardi J, *Multichannel cochlear implants: Channel interactions and processor design* Archives of Otolaryngology, vol. 110, pp. 493-501, 1984

Wilson BS; Finley CC; Lawson DT; Wolford RD; Eddington DK; Rabinowitz WM . *Better Speech Recognition with Cochlear Implants.* Letter to Nature, Vol. 352, July 1991

Wilson BS; Lawson DT; Finley CC; Wolford RD. *Coding Strategies for multichannel cochlear prostheses.* Am J Otol 12(1 Suppl):56-61 1991

Wilson BS, Lawson DT, Zerbi M, Finley CC. *Speech processors for auditory prostheses: Virtual channel interleaved sampling (VCIS) processors- Initial studies with subject SR2.* First Quarterly Progress Report, NIH Project. N01-DC-2-2401. Bethesda (MD). Neural Prosthesis Program. National institute of Health. 1992

Wilson BS; Finley CC; Lawson DT; Wolford RD; Mariangeli. *Design and Evaluation of a Continuous Interleaved Sampling (CIS) Startegy for multichannel cochlear implants.* Journal of Rehabilitation Research and Development, Vol 30. No 1, 1993. Pages 110-116.

Wilson BS. *Signal Processing.* In Tyler RS, editors, *Cochlear implants: audiological foundations.* Chap. 2, San Diego (CA). Singular Publishing Group. In Pres 1993

Wilson BS, Lawson DT, Zerbi M. *Speech processors for auditory prostheses: Identification of virtual channels on the basis pitch.* Third Quarterly Progress Report, NIH Project. N01-DC-2-2401. Bethesda (MD). Neural Prosthesis Program. National institute of Health. 1993

Wilson BS, Lawson DT, Zerbi M. *Speech processors for auditory prostheses: Evaluation of VCIS processors.* Sixth Quarterly Progress Report, NIH Project. N01-DC-2-2401. Bethesda (MD). Neural Prosthesis Program. National institute of Health. 1994

Wilson BS, Lawson DT, Zerbi M, Finley CC. *Recent developments with the CIS Strategies.* In : Hochmeier Desoyer, Hochmaier Es, editor. *Advances in cochlear implants: Proceedings of the 3rd International Cochlea Implants Conference.* Vienna (Austria). 1993. Manz 1994; p: 103-1

Wilson BS, Lawson DT, and Zerbi M, "Advances in coding strategies for cochlear implants," *Advances in Otolaryngology - Head and Neck Surgery*, vol. 9, pp. 105-129, 1995.

Wilson BS. *Speech Processing Strategies.* In Cooper, Craddock, editor, *Cochlear implants: a practical Guide.* 2nd Edition London (England). Whurr, 2006, p:21-69.

Wilson BS; Dorman MF. *Cochlear Implants: Current Designs and future possibilities.* Journal of Rehabilitation Research and Development, Volume 45, Number 5, 2008

Xu L, Tsai Y, Pfingst BE. *BE: Features of stimulaton affecting tonal-speech perception: implication for cochlear prostheses.* J Acoust Soc Am. 2002; 112:247-258

Zeng FG and Shannon R, "*Loudness balance between acoustic and electric stimulation,*" Hearing Research, vol. 60, pp. 231-235, 1992

Zeng FG, *Cochlear Implants in China.* Audiology 1995; 34:61-7

Zeng FG, Cao KL, Wang ZZ. *Progress in cochlear implants.* Chin J Otolaryngol 1998; 33: 123-125

Zeng FG. *Temporal pitch in electric hearing.* Hear Res. 2002; 174 (1-2): 101-6.

Ziese M, Stutzel A, von Specht et al. *Speech understanding with the CIS and the n-of-m strategy in the Med-El COMBI 40+ system.* ORL J Otorhinolaryngol Relat Spec 2000,62: 321-9

Zwicker E, Fastl H. *Psychoacoustics: Facts and Models.* . Berlin, Heidelberg. New York: Springer-Verlag. Pp:103-5, 1990

Websites

http://www.britannica.com/EBchecked/media/18100/Model-showing-the-distribution-of-frequencies-along-the-basilar-membrane

http://openlearn.open.ac.uk/mod/oucontent/view.php?id=398672

http://www.haskins.yale.edu/featured/heads/mmsp/acoustic.html

http://research.meei.harvard.edu/chimera/index.html

http://en.wikipedia.org/wiki/Cochlear_implant#Parts_of_the_cochlear_implant

https://ccrma.stanford.edu/~jos/fp/Zero_Phase_Filters_Even_Impulse.html

http://www.yahyaozturk.com/pages/images/stories/image8.jpg

http://www.utdallas.edu/~loizou/speech/colea.htm

http://research.meei.harvard.edu/chimera/