AN IMPROVED AUDIOMETER FOR SPECIFIC LANGUAGE IMPAIRED CHILDREN

by

ANNIE PUSHPAM PACKIARAJ (201164086)

This thesis is submitted in fulfilment of the requirements for the degree of Masters in Bioengineering

September 2012
Bioengineering Unit
University of Strathclyde
Glasgow, UK

Declaration

This thesis is the result of the author‘s original research. It has been composed by the author and has not been submitted for examination which has led to award of a degree.

Signed:

Date:

Acknowledgments

I would like to thank my supervisor Professor William Sandham for his guidance and support throughout the project. I also thank him for providing me with relevant literature regarding the project.

I would also like to thank Dr. Mark Erik Larsen for his useful suggestions and help in developing the algorithm.

Thanks to MathWorks® for their online tutorials and webinars for getting me accustomed with Matlab®.

Thanks to Miss. Supraja Suresh for her valuable information on Research.

Thanks to my parents and friends for the encouragement they have given me over the past year.

Abstract

The ability to detect only certain frequencies of sound is called a hearing impairment. Specific Language Impairment (SLI) is a very common type of impairment found in many children. Generally, children with SLI are often diagnosed with developmental language disorder and as they grew older they may find difficulty in learning new words and making a conversation. Since children with SLI have no hearing loss, an adaptive computerized program was developed to generate speech from the text entered so as to examine their language accuracy. The application was developed using the Matlab® environment. A child’s brain cannot process the rapidly changing audio information within a short period of time and hence a delayed speech will be generated. The delay speed at which a particular child can differentiate words having a similar sound was identified. The child can then be treated accordingly with the help of speech therapy software that makes the brain function better after consistent therapies.

Table of Contents

Table of Figures

1. INTRODUCTION

“A characteristic of the normal child is he doesn’t act that way often” - anonymous. Childhood is the most crucial stage of life. Children with lack of sufficient nutrition, nurturing, interaction and stimulus suffer from developmental deficits. One such developmental disorder is specific language impairment (SLI). The cause of SLI has not yet been identified, but it has been revealed that there could be a genetic basis to SLI [1]. Children with SLI have no hearing loss, and SLI is a relatively common childhood learning disability. These children may show poor performance at schools and it is important to help them in improving their language skills. The present study was to develop an audiometer that can identify the level of linguistic ability of children with SLI, with respect to perceptual acuity.

The Wernicke’s area surrounding the auditory cortex (Heschl’s gyrus) on the sylvian fissure in the frontal and temporal opercula of the brain subserves language [2]. Just as the body becomes fit when we run, the brain can learn better after consistent exercises. Generally when we are listening to a foreign language, people would slow down so that we can understand them. Similarly children with SLI cannot perceive the language appropriately and if slowed down, the perceivance will be better [3]. It has been proposed to determine the delay speed at which the children can understand so that they can undergo exercises starting at the same speed and increasing it gradually.

The audiometer to determine the delay speed as mentioned above has the following two sections:

1. A text-to-speech engine that can generate speech with different delay rates.
2. A user interface in which the text to be spoken is entered and the delay at which it is to be spoken is selected.

1.1. STRUCTURE

Section 1 is this introduction of the project. Section 2 and its subsections portray the background knowledge required for this project. The SLI condition is briefly discussed in Section 2.1, followed by a brief anatomy of the human brain in Section 2.2, and a description of brain morphology in children with SLI in Section 2.3. The anatomy of the human auditory system and the principles of audiometry are demonstrated in Sections 2.4 and 2.5, respectively. The text-to-speech synthesis which forms the basis of the speech audiometry is discussed in Section 2.6. The methodology and implementation of the project is illustrated in Section 3 and its subsections. Section 4 describes the outcome of the project. Results are discussed in Section 5, followed by the conclusions and future work in Section 6. Two appendices are included. Appendix A is a User's manual and Appendix B lists the Matlab® coding.

2. BACKGROUND

2.1. SPECIFIC LANGUAGE IMPAIRMENT

Specific language impairment (SLI) is a communication disorder that affects syntax and inflectional morphology. Inflectional morphology is the changes that happen in words to denote certain grammatical features. Approximately half a million children between the ages of 3 years and 16 years have SLI [4]. Based on different linguistic characters, children with SLI can be classified into subgroups such as children with semantic SLI, who cannot understand the human expression through language and children with pragmatic SLI, whose content of language is unusual. The former are those with phonological SLI, the children speak in long but poorly intelligible utterances and the latter are those with grammatical SLI. Additionally there is also another subgroup called familial aggregation, those with the genetic background of SLI. The degree of impairment in different subcomponents of the language differs with each SLI child.

A prominent low-level causal theory is that children with SLI have difficulties in processing brief, rapidly successive acoustic stimuli and that these difficulties lead directly to their language problems [5]. This suggests that these children are capable of processing the delayed successive acoustic stimuli. According to Corriveau et al. [6] twelve 8-to-12 year old children with impairment, and twelve normal controlled children of a similar age group were asked to listen to two different tones and respond by pressing the correct buttons associated with the tone in suitable order. Children with SLI found it difficult when the stimuli were brief with the slow interstimulus interval (75ms), but when the interstimulus interval was longer (>150ms) they did not differ from the other normal participants.

Language disorder may be due to several neurological conditions that must be ruled out before diagnosis of SLI is confirmed. There should be no evidence of cerebral palsy, traumatic brain injury, or focal brain lesions. Auditory processing in children plays an important role in word learning. The appropriate understanding of stress and syllables in a word leads to proper understanding of that particular word. The understanding levels differ from each child. Auditory processing of the speech has not been investigated in children with SLI at early periods but recent researches in infant language acquisition have shown that auditory processing of prosody plays a vital role. Prosodics of language involve variation in syllable length, loudness and pitch. There are software therapy packages available commercially which help in strengthening the skills of memory, attention, processing rate, and sequencing for children. These packages deliver the speech at same speed for all children whereas the proposed audiometer will deliver at different speed for different children.

The goal of the audiometer developed based on the above hypothesis is that, the auditory processing level of a particular child can be determined. This audiometer will also help in the development of new therapy packages. The speed of the auditory stimuli can be manually selected for each individual child in the therapy package.

2.2. HUMAN BRAIN

The adult human brain consists of almost 97 percent neural tissue and is the functional unit of the nervous system. The brain is protected by the bones of cranium, cranial meninges and the cerebrospinal fluid and it is also isolated biochemically by the blood-brain barrier from the general circulation of the body [7].

The largest part of the human brain is the cerebrum. When viewed from the anterior and superior surfaces, it can be divided into two cerebral hemispheres. The hemispheres are highly folded and covered by cortical layer. The cerebral cortex forms a series of elevated ridges called gyri that increase the surface area. The gyri are separated by deeper groves called fissures or shallow groves called sulci. The cerebrum is responsible for most of the higher mental function.

Each cerebral hemisphere is divided into four lobes: the frontal lobe, the parietal lobe, the occipital lobe and the temporal lobe – see Figure 2.1. The lobes are named after the skull bones that cover them respectively: the frontal bone, the parietal bone, the occipital bone and the temporal bone. The borders between the lobes are beneath the sutures that link the skull bones except for the border between frontal and parietal lobe which is located backward from the corresponding suture [7].

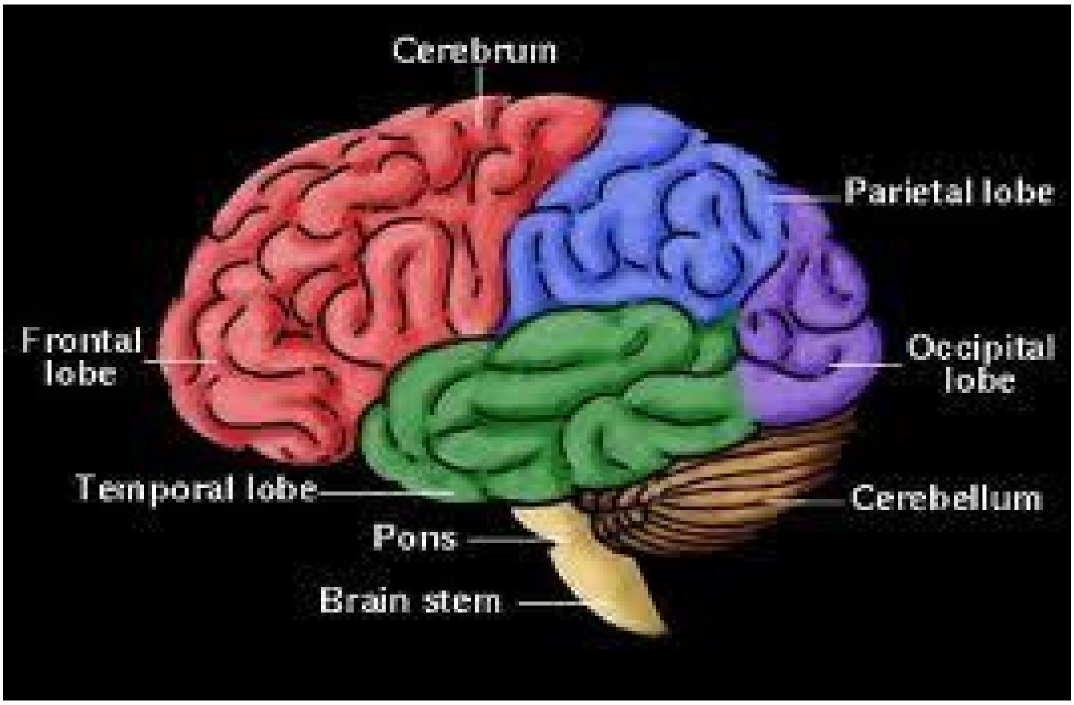

Figure 2.1. Human brain [http://www.holisticeducator.com/brain.htm, 8th Aug 2012].

The dopamine-sensitive neurons are present in the frontal lobe. Dopamine limits the sensory information arriving from the thalamus to the forebrain. The frontal lobe helps in analysing the future consequences from the current action. In general, the frontal lobe deals with the mental functions. The pars triangularis located in the inferior frontal gyrus of the frontal lobe contribute to the propositional language comprehension and linked to speech production.

The sensory information from different modalities is integrated by the parietal lobe. For instance the somatosensory cortex and the dorsal stream of the visual system are present in the parietal lobe which helps in navigation. The multisensory parietal lobe is separated from the frontal lobe by the central sulcus; the parieto-occipital sulcus separates it from the occipital lobe; the sylvian fissure separates it from the temporal lobe. The Wernicke's area surrounding the sylvian fissure is responsible for auditory processing. The lobe itself is again divided into two hemispheres: the right deals with the spatial relationships, and the left with the symbolic functions in language and mathematics [8].

The temporal lobe is responsible for the auditory perception and is also important in semantic processing of both speech and vision. The hippocampus in the temporal lobe plays a vital role in long-term memory. The superior temporal gyrus receives the signal from the cochlea; the medial temporal lobe is involved in memory.

Wernicke's area situated between the temporal lobe and the parietal lobe and is important in language development. If this area of the brain is seriously damaged, it results in the impairment of language and its usage. The heart of the Wernicke's area is the planum temporale, the cortical area posterior to the auditory cortex within the sylvian fissure [9].

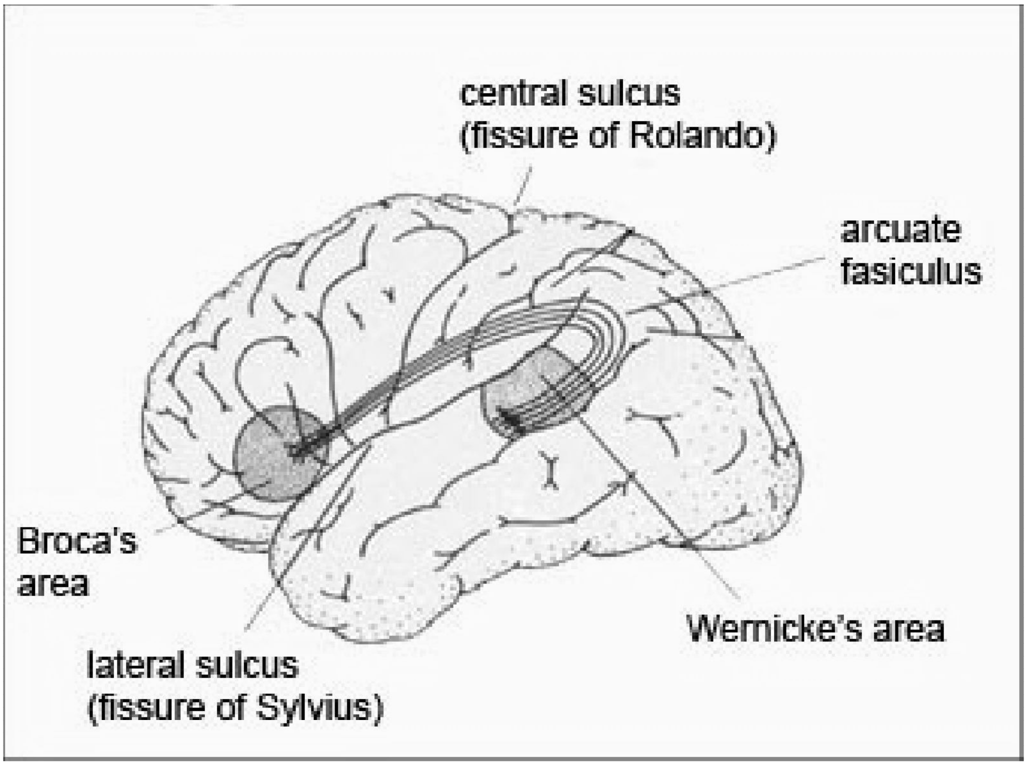

Figure 2.2. Brain: Wernicke's area [http://thebrain.mcgill.ca, 2nd Sep 2012].

The occipital lobe contains the primary visual cortex and is the centre for visual processing. Retinal sensors pass the stimuli to the lateral geniculate bodies, which then passes the optic radiations to the visual cortex. If one of the occipital lobes is damaged it causes homonomous vision loss. The occipital lesion is the reason behind visual hallucination.

Cerebellum is the second largest part of the brain and is partially hidden by the cerebrum. The cerebellum is also divided into two hemispheres which are covered by a layer of gray matter, the cerebellar cortex. The cerebellum adjusts ongoing movements by comparing arriving sensations with previously experienced sensations, allowing us to perform the same movements again and again [7]. The functional and structural link between the two cerebellar hemispheres is the diencephalon. The walls of the diencephalon are made up of the right and left thalamus which contains the processing and relay centres for sensory information. The floor of the diencephalon is the hypothalamus and it contains the centre for the autonomic functions. The pituitary gland is connected to the hypothalamus by the infundibulum.

The brain stem contains a variety of processing centres and nuclei that relay information headed to or from cerebrum or cerebellum. The brain stem consists of mid-brain, pons, and medulla oblongata. The mesencephalon or the mid-brain consists of the nuclei that are responsible for visual and auditory processing. The pons connects the brain stem and the cerebellum and also contains nuclei for somatic and visceral control. The medulla oblongata is the region where the spinal cord is connected to the brain. The medulla oblongata relays sensory information to thalamus and other parts of the brain. It also contains major centres responsible for autonomic functions such as digestion, heart rate and blood pressure [9].

2.3. MORPHOLOGICAL DEFICITS OF CHILDREN WITH SLI

The left cerebral hemisphere of the human brain is considered to be the language dominant hemisphere; this is because of the specialization of perisylvian region of the left hemisphere in language. The individuals with normal language skills have a larger planum temporale and pars triangularis in the left hemisphere than the right hemisphere. But the individual with developmental language disorder reported to have symmetry or asymmetry of the planum [10].

Soriano-Mas et al., [11] analysed the whole brain of the individuals with SLI using voxel-wise analysis technique. It was revealed that the patients showed global increase in gray and white matter volume than the controlled individual. The gray matter volume at the regions of the right perisylvian region and the occipital petalia were significantly larger. The global changes were the outcome of increase in specific brain parenchyma.

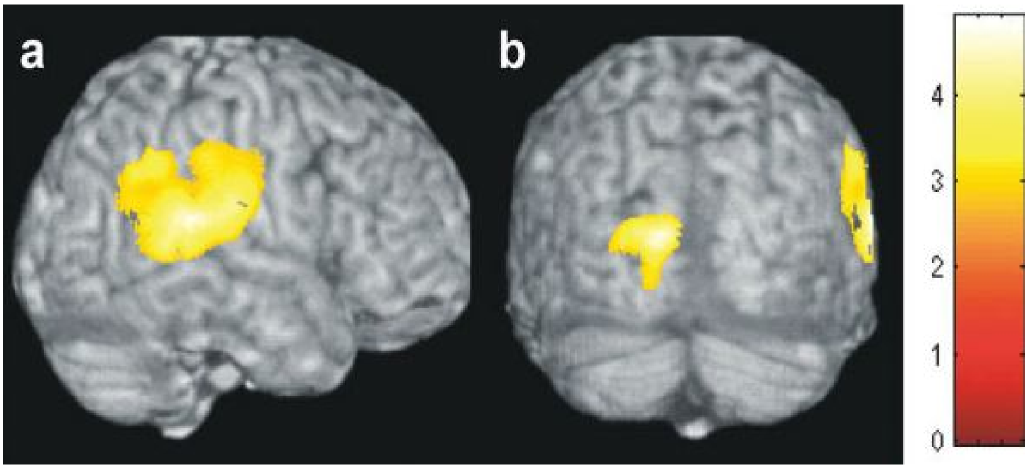

Figure 2.3. Area of gray matter volume increase in the impaired children superimposed on a rendered normalized brain. (a) Right perisylvian region, (b) left middle occipital gyrus. Bar represents the t value (P < 0.001) [11].

Abnormal development of procedural memory system of brain structure is witnessed in the individuals with SLI. The procedural memory system is also associated with grammar aspects. This system is a network of interconnected structures in the frontal lobe of the brain [12].

2.4. HUMAN AUDITORY SYSTEM

Hearing is an essential part in the speech chain. Speech communication was developed with an idea that a speaker wants to convey information to the listener.

The human auditory system converts mechanical signals into complex electrical signals in the central nervous system. Humans can detect sounds of range 20 to 20,000 Hz. The human ear consists of three layers – outer, middle and inner ear, with functions similar to a sensor, amplifier and converter, respectively.

Sounds are generally the pressure waves that impinge on the ear. The outer ear, comprised of the pinna and external auditory meatus, detects the pressure waves and directs them to the tympanic membrane (the ear drum). The sound waves create vibrations on the ear drum, which are then transmitted through the three middle ear ossicles - malleus, incus and stapes to the fluid of inner ear. The stapes fits tightly into the oval window of the bony cochlea. The stapes move in and out and there needs to be a compensatory movement. Since the inner fluid is incompressible, the round window membrane beneath the oval membrane moves in the opposite direction to the stapes [7].

Since the area of the stapes is smaller than the area of the tympanic membrane, there is amplification in the sound waves. The arm of the malleus to which the tympanic membrane is attached is longer than the arm of the incus to which the stapes is attached, therefore the lever action of the stapes also helps in the amplification of sound [13]. The stapes provides a gain of approximately 20 dB, the majority of the total gain.

The semi-circular canals are related to the sense of balance, whereas the cochlear canals are related to the sense of hearing. The cochlea is a long coiled tube with three canals separated by two membranes – scala vestibuli, scala tympani and scala media or cochlear duct which contains the organ of corti. The Reissner’s membrane separates the scala vestibuli and scala tympani whereas the basilar membrane separates the scala tympani and scala media and also determines the wave propagation properties of the cochlear partitions.

Perilymph, rich in sodium, is a cochlear fluid located within the scala vestibuli and scala tympani. Reissner's membrane and basilar membrane together forms a compartment filled with cochlear fluid, endolymph. Endolymph is rich in potassium ions, which produces an electrical potential.

Four rows of sensory hair cells arranged in the organ of corti along the entire length of the cochlea are powered by the potential difference between the perilymph and endolymph. One row consists of inner hair cells (IHC) which provide the neural output of the cochlea whereas the three rows consisting of outer hair cells (OHC) receive neural input from the brain.

The cochlea converts the electrical signal back to mechanical by the reverse transducer property of OHC whenever it is necessary to amplify weak sounds. The movement of the hairs may result in the sound waves being emitted back to the ear canal and this phenomenon is called otoacoustic emissions (OAE). OAE are due to a wave exiting the cochlea via the oval window, and propagating back through the middle ear to the eardrum, and out the ear canal, where it can be picked up by a microphone.

The neural code of the central auditory system is complex. The cells responsible for different frequencies of sound are at different places and there is a logarithmic relationship between the frequency and the positions. Each cell has a characteristic frequency and hence different hearing impairments have different characteristics.

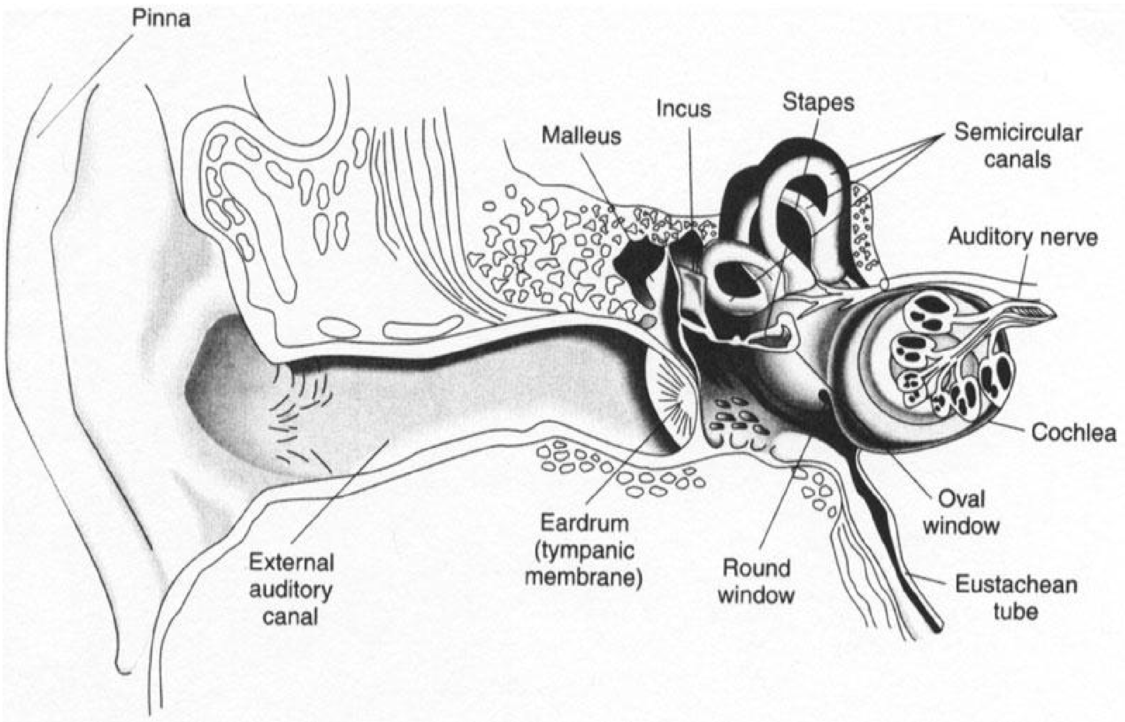

Figure 2.4. Human auditory system [http://www.skidmore.edu/~hfoley/Perc9.html, 2nd Sep 2012].

2.5. AUDIOMETRY

Audiometry is the science of measuring hearing accuracy for variations in sound intensity and pitch. The hearing level of an individual can be determined using audiometric tests with the help of an audiometer. Acoustic reflex and otoacoustic emissions may also be measured. There are different types of audiometry such as pure tone audiometry, speech audiometry and immittance audiometry.

2.5.1. Pure tone audiometry

Pure tone audiometry helps in identifying the hearing threshold levels of an individual. This audiometry consists of two parts, air and bone conduction. The air conduction tests evaluate all parts of the auditory system. A tone is presented at decreasing intensities through a headphone until the patient can no longer hear the sound. A sequence of decreasing and increasing intensities is then applied, and the threshold is determined [14]. The bone conduction tests evaluate only the inner ear and the hearing nerve and are done only if hearing loss is detected in air conduction tests. The bone conduction test is conducted using a bone-conduction vibrator placed behind the ear that transmits the sound. It also follows the same procedure as air-conduction as mentioned above.

2.5.2 Speech audiometry

Speech audiometry assists in measuring the intelligibility at different intensity levels. During the test, certain numbers of words are presented via headphone and the patient is asked to repeat these words. At the end of the test, percentage of words repeated correctly can be derived. This type of audiometry is suitable for the assessment of SLI.

2.5.3 Immittance audiometry

The measurements can be taken directly without the voluntary responses from the patient. This generally consists of three separate tests – tympanometry, acoustic reflex thresholds and reflex decay. Tympanometry is the test that evaluates the movement of eardrum and status of the middle ear. In acoustic reflex thresholds test, the slight changes in eardrum due to the contraction of stapes and tympani in response to loud sounds are assessed. This helps in evaluating if the hearing loss is conductive or sensorineural. The reflex delay evaluates if the hearing loss is due to the problem in cochlea or the acoustic nerve. A loud tone is transmitted for 10 sec and the changes in eardrum due to the contraction of stapes are monitored [15].

2.6. TEXT-TO-SPEECH SYNTHESIS

The task of a text-to-speech synthesis (TTS) system is to mimic what human readers do. TTS is a type of speech synthesis application used to create a spoken version of the text in a computer page. The TTS system consists of two processes, text and linguistic analysis, and speech synthesis. The first main task breaks down into further subtasks. A block diagram of a TTS system is shown in Figure 2.5.

2.6.1. Text and linguistic analysis

Text and linguistic analysis is the process by which the input text is converted into a form of linguistic representation that includes information on the phonemes (sounds) to be produced, their duration, the locations of any pauses and the frequency contours to be used [16]. The text and linguistic analysis may be further broken down as follows:

- Text normalization.
- Phonetic representation.
- Prosodic analysis.

2.6.2. Text normalization

The input to a TTS system is text, encoded using an electronic coding scheme appropriate for the language, such as ASCII. The tokens in the raw text are identified and are divided into reasonable size. This process is also called text tokenization.

2.6.3. Phonetic representation

After the tokenization of input text, the parts-of-speech tags, also called grammatical tags and phonetic transcriptions (the visual representation of speech sounds) are assigned. The process of assigning phonetic transcriptions to words is called text-to-phoneme (TTP) conversion.

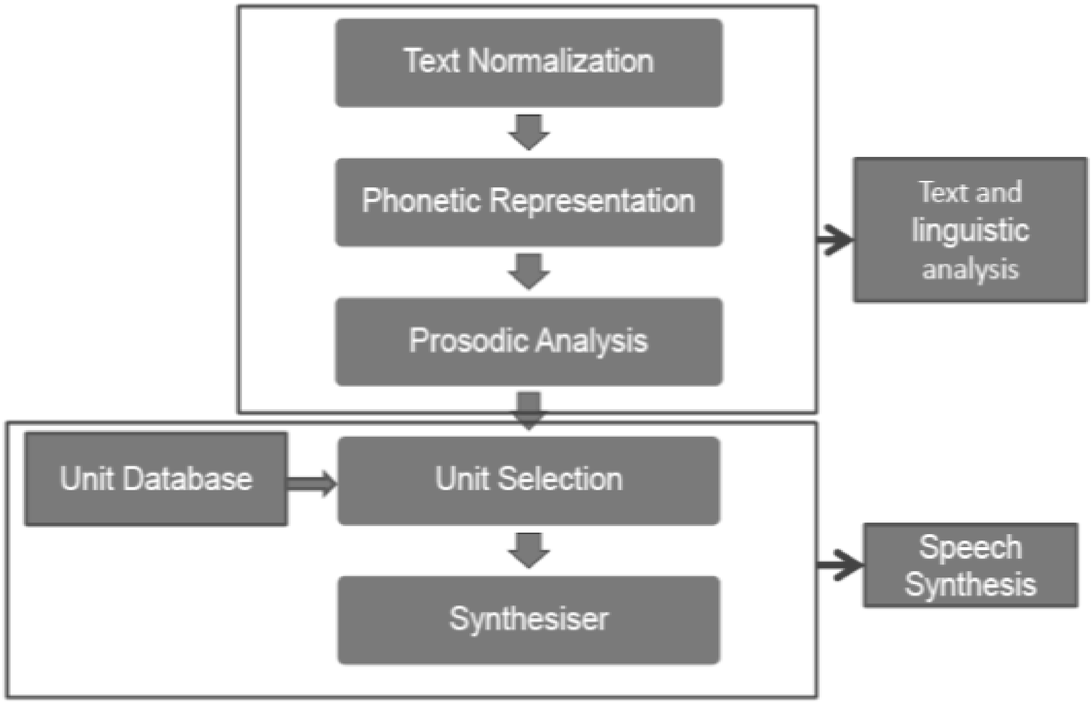

Figure 2.5. Block diagram of a TTS system.

2.6.4. Prosodic analysis

Prosody is an inherent supra-segmental feature of human's speech that can express attitude, emotion, intent and attention [17]. Prosody may reflect various features of the pronunciation other than the grammar of the language. The prosody generator needs to analyse the text and overlay appropriate prosodic information [14].

2.6.5. Speech synthesis

Speech synthesis is a process in which the linguistic representation of the input text is transformed into an audible speech waveform. The heart of a TTS system is speech synthesis. Speech synthesis can be classified into four groups – articulatory, formant, concatenative and statistical synthesis. Articulatory synthesiser produces speech waveform based on the model of human vocal tract and articulation process. Formant synthesis is a process in which parameters such as fundamental frequency, voicing and noise levels are varied over time to create the speech waveform [18].

In concatenative synthesis, the speech waveform is produced by concatenating various segments of recorded speech. It produces the most natural sounded speech waveform. This is again subdivided into three types – unit selection, diphone and domain-specific synthesis. The unit selection utilises large database of recorded speech. Unit selection produces the greatest natural sounding waveform because it involves only a minimum of digital signal processing on recorded speech. Diphone synthesis uses speech database containing all diphones occurring in a language. Domain-specific synthesis makes use of a database which has only limited words and phrases; these are not for general purpose and can synthesis only words and phrases with which it was programmed [19]. The domain-specific synthesis has been in commercial purpose for a long time (e.g. talking clocks, calculators).

Statistical synthesis is based on hidden Markov models (HMM) and hence it is also called HMM synthesis. In this particular method, the frequency spectrum, the fundamental frequency and the prosody are modelled simultaneously by HMMs.

3. METHODOLOGY

The audiometer developed has two components. First and foremost is the text-to-speech system which generates audible speech from the arbitrary text. The second is a user interface which helps in specifying the text to be spoken, the delay speed at which the text to be spoken and other general information regarding the user.

3.1. SPEECH SYNTHESISER

Initially, a Harmonic plus noise (HNM) synthesiser was developed using the Simulink® environment. Simulink® is a platform for stimulation and model-based designs. The speech signal generated using this harmonic noise model would have been composed of a harmonic part, the periodic component of a speech signal and a noise part, the non-periodic component of a speech signal. The two components are separated by a time-varying parameter, maximum voiced frequency (F_m) in the frequency domain. The lower band of the spectrum below F_m is the harmonic component and the upper band above F_m is the noise component [20]. The harmonic part is given as a sum of harmonics:

$$s_h(t) = \sum_{k=-L(t)}^{L(t)} A_k(t) \exp\left(jk\omega_0(t)t\right) \qquad (3.1)$$

where $L(t)$ is the number of harmonics in the harmonic part, ω_0 denotes the fundamental frequency and $A_k(t)$ can take one of the following forms:

$$A_k(t) = a_k\left(t_a^i\right) \qquad (3.2)$$

$$A_k(t) = a_k\left(t_a^i\right) + tb_k\left(t_a^i\right) \qquad (3.3)$$

$$A_k(t) = a_k\left(t_a^i\right) + t\Re\left\{b_k\left(t_a^i\right)\right\} + t^2\,\Re\left\{c_k\left(t_a^i\right)\right\} \qquad (3.4)$$

where $a_k(t_a^i)$, $b_k(t_a^i)$ and $c_k(t_a^i)$ are complex numbers denoting the amplitude of the kth harmonic, its first derivative and its second derivative, respectively. $\Re$ denotes the real part of that particular complex number. These parameters are measured at $t = t_a^i$ referred to as analysis time instants. The $L(t)$ depends on the fundamental frequency ω_0 and the maximum voiced frequency F_m. For $|t - t_a^i|$ small, HNM assumes that $\omega_0(t) = \omega_0(t_a^i)$ and $L(t) = L(t_a^i)$. The noise part of the synthesiser is described in frequency by a time-varying autoregressive model $h(\tau, t)$, and its time domain is imposed by a parametric envelope, $e(t)$ that modulates the noise component. The noise part is given by

$$s_n(t) = e(t)[h(\tau, t) * b(t)] \tag{3.5}$$

where $*$ denotes convolution and $b(t)$ is white Gaussian noise. Finally, the speech signal $s(t)$ is given by:

$$s(t) = s_h(t) + s_n(t) \tag{3.6}$$

The basic block diagram of the harmonic part of a HNM synthesiser is shown below in the Figure 3.1.

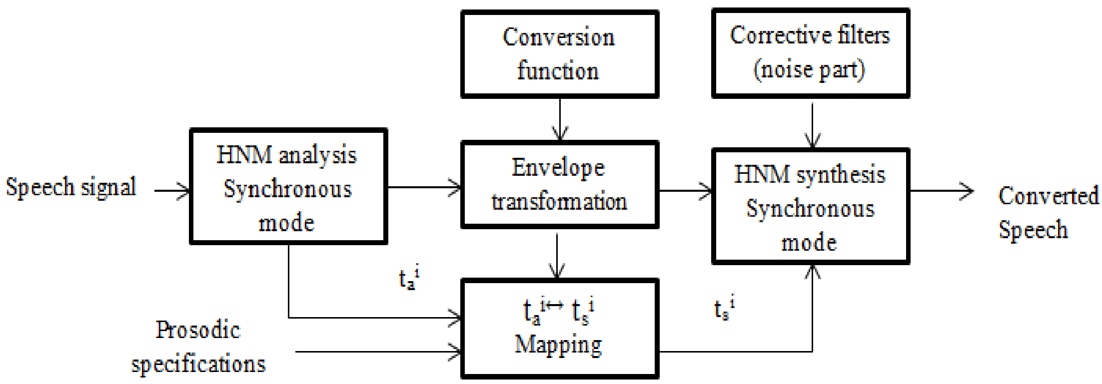

Figure 3.1. Harmonic plus noise model synthesiser [21]; t_a^i: analysis time instants; t_s^i: synthesis time instants.

The Simulink® model for a harmonic plus noise synthesiser is a complex open-loop design and it was decided to build an audiometer using the Matlab® development environment.

3.2. TEXT-TO-SPEECH SYSTEM

The text-to-speech part of the audiometer is designed in such a way that Google text-to-speech engine already available online was used in combination with the ActiveX controls using Matlab®. The audiometer initially developed was without specifying the speech delay and is shown below.

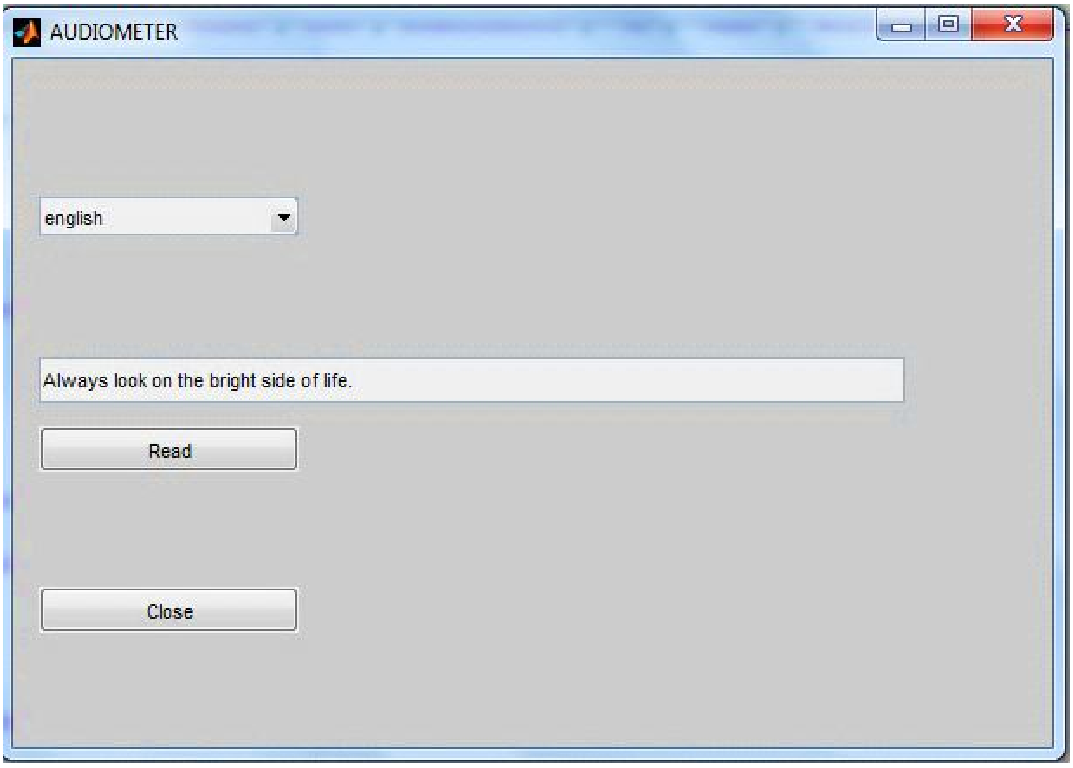

Figure 3.2. Audiometer: initial stage.

The speed of a sound wave is given by the following equation:

$$v = f\lambda \tag{3.7}$$

where v is the speed of the sound, f is the frequency and λ is the wavelength of the sound.

The above equation relates the speed, frequency and wavelength. Though the speed is calculated using frequency and wavelength, speed is independent of the other two factors. In contrast, the frequency and wavelength are inversely proportional; if the wavelength is doubled, the frequency is halved. In order to slow down the sound wave, the wavelength has to be increased which in turn results in the decrement of frequency.

The normal audible frequency range for human is from 20 Hz to 20,000 Hz. The frequency is also called pitch in terms of sound wave. As the audiometer is being developed for children with SLI, the pitch factor cannot be compensated. Hence an algorithm was developed to adjust the speed of the sound wave without changing the pitch. This was implemented by using windows media player as the source of audio output.

The media player has in-built play speed settings and the pitch can be maintained at the same level with the changing play speed. The media player usually plays the sound waves from the library containing the playlists. The output from the online text-to-speech engine is integrated as a playlist in the player. Each time the text is modified a new playlist is created and it contains only a single track. This can be achieved by commanding the autoloop property of the player to return to zero.

The play speed settings are modified using the default commands that control the media player. The different speeds are selected and applied accordingly to the playlist using the switch function. Assuming that the default speed is 1, it is reduced by factors of 3/4, 1/2 and 1/4 to slow down the speech. It can also be further reduced by factors such as 1/8, 1/16. The online text-to-speech engine is combined with the

ActiveX media player plugin using the callback command. The text-to-speech module developed using a simple Matlab® algorithm was not as effective as expected. Each time the text has to be changed, the user should make a change in the Matlab® m-file and hence the online text-to-speech engine was implemented.

3.3. GUI

The Graphical user interface (GUI) is a graphical display that allows the user to perform interactive tasks. Here, the GUI generated was based on a programmatic, GUI-building approach. A code file was created that defined all properties and behaviours of the components. When the file is executed, a figure is created in which the components and the user interactions are loaded. A new figure is created each time the algorithm is executed.

The audiometer was designed to have a listbox from which a particular play speed can be selected. On selecting a speed, the speed of the sound is assigned to the respective play speed. An edit field is also been created in which the text to be entered is limited to 100 strings. There are also separate fields to enter the general information about a patient. The GUI also has two push buttons, one that commands the audiometer to speak is labelled “read” and the one to stop and close the audiometer window is labelled “close”.

4. RESULTS

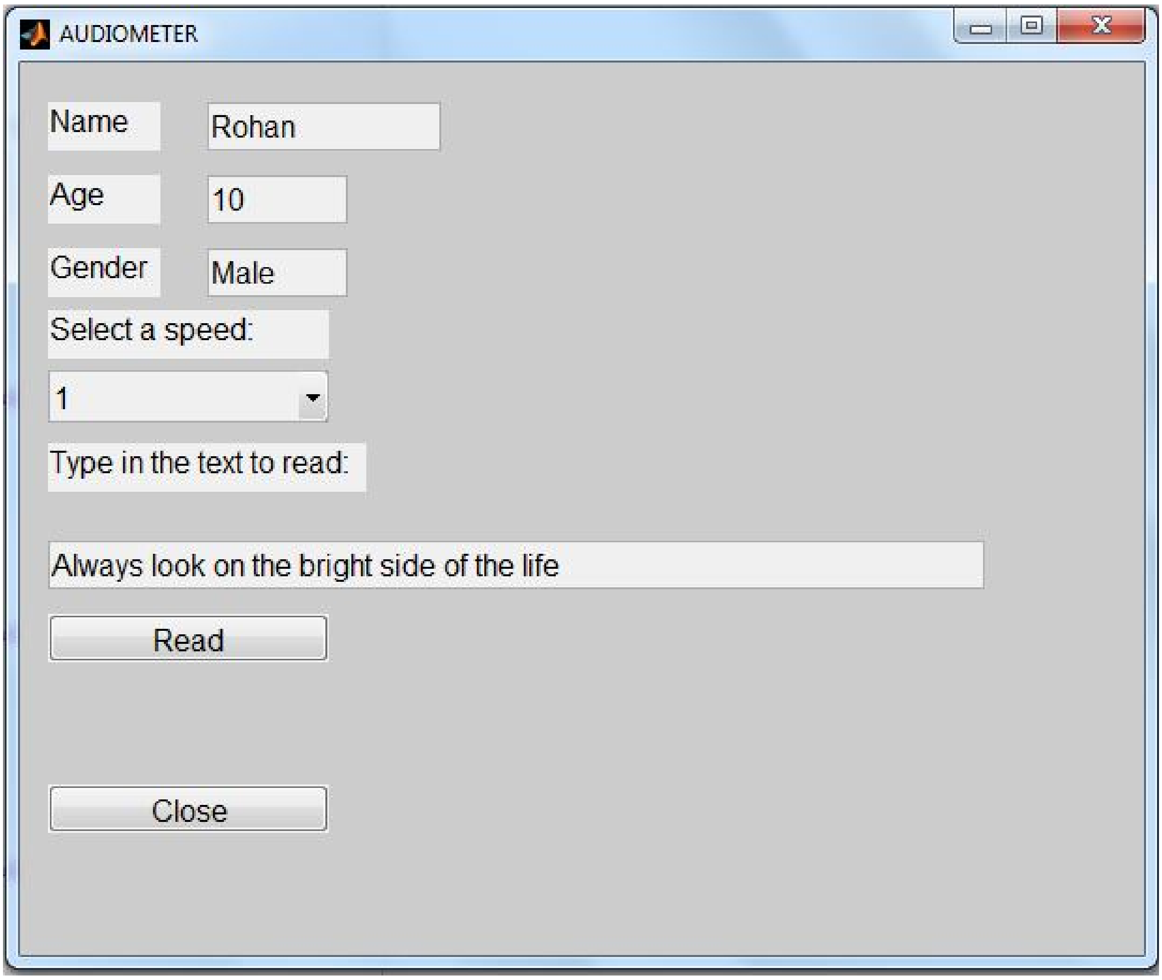

Figure 4.1. GUI of the audiometer developed.

4.1. PROCEDURE

1. Enter a word formed from simple syllables that is to be listened by the patient.
2. Select a play speed and ask the patient to respond by uttering the word that was heard.
3. Increase the play speed until the patient utters the word appropriately.
4. Make a note of the play speed that the patient responded.
5. Repeat the experiment by entering another word with typical syllables.
6. From the results we can know the auditory processing speed of the patients for both simple and typical syllables.

4.2. EXPECTED RESULT

The expected result was to create an audiometer that can generate speech from the text entered at a specified playback speed, so that the auditory processing speed of the children with impairment can be diagnosed. The relation between the auditory processing and the duration of the tone has been explained by Corriveau et al. [6]. It was believed that the same principle can be used to determine the language ability and the auditory processing acuity of the children by delaying the speed of speech.

4.3. ACTUAL RESULT

A clearly audible speech is generated by means of a text-to-speech engine through the media player, but the playback speed control is malfunctioning. The audiometer could not be tested until the above mentioned technical issue is solved. Once the issue is solved, the reliability of the audiometer in diagnosing the auditory processing of children with impairment could be identified.

5. DISCUSSION

The audiometer developed in the previous work by Larsen [14] was based on the principle of immitance audiometry, while the currently developed audiometer relies on speech audiometry. The otoacoustic emissions (OAE) generated from within the inner ear of the children were to be analysed to determine the location of the defect in the auditory system by Larsen [14]. However recording of OAEs from the children required authorisation from the Ethics Committee and it was decided to generate the auditory stimuli that were to be used in the audiometer.

The purpose of the currently developed audiometer is to diagnose the auditory processing acuity of impaired children. I was motivated to perform this experiment from the work done by Benasich and Tallal [22] in which it has been hypothesised that “the processing brief rapid successive auditory cues could impair or delay the formation of the distinct, phonological representations, leading to delay or disruption of language acquisition”. The role of auditory processing in language acquisition is inevitable and for that reason it is significant to determine the processing acuity of children with language impairment.

The absolute cause for the language impairment is still a debatable subject. There are two theories under consideration, a remarkable theory which states that the impaired children have difficulty in processing the brief, rapidly changing auditory stimuli and the second theory falls under two categories. First category involves theories that believe in specific deficits in language knowledge, for instance knowledge of marking tense. The second category involves theories that assume deficits in language processing, for instance the way an individual use the words to communicate.

The audiometer designed is based on the first category that deals with the auditory processing of the language. Moreover, the speech generated is more like a natural speech, whereas in the previous audiometer it's a combination of just a constant and vowel. The current work concentrates in determining the processing speed while the previous one concentrated in diagnosing the defect in auditory system.

The technical difficulty faced when developing the audiometer was in delaying the sound wave without decreasing the frequency. To achieve this requirement, windows media player was used as mentioned in Section 3.2. Only some features of the media player are controllable by the Matlab® algorithm and playback speed could not be controlled. In case of the play speed control functioning appropriately, the audiometer would have been tested and the results would have been prominent to prove that the processing speed can be identified for each individual.

From the knowledge obtained from this research, I would suggest that diagnosing the processing speed of the auditory cortex plays a vital role in understanding the behavioural nature, language ability of children with SLI. It tells you how fast a particular child can process the language.

6. CONCLUSIONS AND FUTURE WORK

This project systemised how to develop an audiometer to determine the auditory processing speed of children with SLI. The development of auditory processing in relation to the language acquisition in children with SLI is not normal as it is in controlled children. The degree of perception is highly correlated with the degree of language impairment and hence it was decided to diagnose the perception acuity. The ideology was implemented with the help of a PC-based audiometer program developed under the Matlab® environment, from which children can hear speech at a manually controlled speed. As a result, an audiometer that can generate speech was developed but the speed could not be controlled because of a software fault. The previously developed audiometer models were designed to determine the hearing level of the auditory system while this was aspired to diagnose the processing speed of the auditory cortex.

An algorithm was developed to combine the text-to-speech engine with the media player. The media player was adopted as it has in-built play speed control which does not influence the pitch of the sound wave when altered. The output from the text-to-speech engine was integrated as a playlist in the player. Assuming the default speed to be 1, the speed was decremented by multiples of 1/2. The combination of text-to-speech engine with the media player commanded by Matlab® environment was not as persuasive as expected. The usage of Simulink® would have provided a better result and the aspect of speed control might have been better. However, the speech developed was flawless and it was more like natural speech than synthetic speech.

A GUI was also developed for the user to enter the text to be heard by the patient and also to enter the patient details. The GUI developed is a programmatic GUI which creates a new display each time it is executed.

6.1. FUTURE WORK:

The audiometer developed was expected to have a manual control over the speed of auditory stimuli. This can be re-implemented having a control on the auditory stimuli speed as well as a control on the interstimulus interval (ISI). This will provide time for the appropriate processing of the previous stimulus. In contrary to this current model, the text-to-speech engine can be designed using the control system in Simulink®. The media player is not needed if a Simulink® model is created. This may produce better results than the current version.

The developed synthesiser must be able to control the speed of the speech without altering the pitch of the sound. It is significant that the sound wave should be of quality before the software being tested on children.

The GUI developed offers limited options where the values can be assigned and the diagnosis takes place. It can be provided with an option of saving, and printing the test result for a particular patient.

REFERENCES

1. Warwick, L., *Promising Genetic link for specific language impairment.* Bulletin for the centre of excellence for early childhood development, 2003. **2**(3): p. 11.
2. De Groot, A.M.B., *Bilingualism and the brain,* in *Language and cognition in bilinguals and multilinguals: an introduction*. Psychology Press, East Sussex, 2010. p. 418-435.
3. Tallal, P., *Improving language and literacy is a matter of time*. Nat Rev Neurosci, 2004. **5**(9): p. 721-728.
4. van der Lely, H.K.J., *Specific language impairment in children: research findings and their therapeutic implications.* Eur J Dis Comm, 1993. **28**(3): p. 247-261.
5. Tallal, P., and Piercy, M., *Developmental aphasia: Impaired rate of non-verbal processing as a function of sensory modality*. Neuropsychologia, 1973. **11**: p. 389-398.
6. Corriveau, K., Pasquini, E., Goswami, U., *Basic auditory processing skills and specific language impairment: a new look at an old hyhypothesis.* J Speech Lang Hear Res, 2007. **50**: p. 647-666.
7. Martini, F.H., Nath, J.L., Bartholomew, E.F, 9th ed. *The brain and cranial nerves,* in *Fundamentals of Anatomy & Physiology*. Pearson Education Ltd., London, 2011. p. 448-451.
8. Kinser, A.P., *Serendip.* URL: http://serendip.brynmawr.edu/bb/kinser /Structure1.html

9. Bear, M.F., Connors, B.W., Paradiso M.A., 3rd ed. *Language*, in *Neuroscience: exploring the brain*. Lippincott Williams & Wilkins, Philadelphia, 2007. p. 479-643.
10. Gauger, L.M., Lombardino, L.J., Leonard, C.M., *Brain morphology in children with specific language impairment.* J Speech Lang Hear Res, 1997. **40**(6): p. 1272-1284.
11. Soriano-Mas, C., Pujol, J.S., Ortiz, H., Deus, J., López-Sala, A., Sans, A., *Age-related brain structural alterations in children with specific language impairment.* Hum Brain Mapp, 2009. **30**: p. 1626-1636.
12. Ullman, M.T., and Pierpont, E.I., *Specific language impairment is not specific to language: the procedural deficit hypothesis.* Cortex, 2005. **41**: p. 399-433.
13. Gray, L., *Auditory system: structure and function*, in *Neuroscience.* The University of Texas Medical School: Houstan, 1997.
14. Larsen, M.E., *An audiometer for children with a specific language impairment*, in *Department of Electronic and Electrical Engineering.* 2003, University of Strathclyde.
15. Stach, B.A., 2nd ed. *The audiologist's assessment tools,* in *Clinical audiology: an introduction. Cengage learning products.* New York, 2010. p. 236-333.
16. Sproat, R.W. and Olive, J.P., *Text-to-speech synthesis.* AT&T Tech J, 1999. **74**: p. 35-44.
17. Tian, J., *Modular design for Mandrain text-to-speech synthesis,* in *TC-STAR Workshop on Speech-to-Speech Translation*. Barcelona, Spain, 2006.

18. Deller, J.R., Proakis, J.G., Hansen, J.H.L., *Fundamentals of speech science*, in *Discrete-time processing of speech signals*. Macmillan Publishing Company, 1993.

19. van Santen, J.P.H., Sproat, R.W., Olive, J.P., Hirschberg, J., *Concatenative synthesis and automated segmentation*, in *Progress in speech synthesis*. Springer, 1996. p. 259-261.

20. Stylianu, Y., *Applying the harmonic plus noise model in concatenative speech synthesis.* IEEE Trans Speech audio processing, 2001. **9**(1): p. 21-29.

21. Stylianu, Y., and Capé, O., *A system for voice conversion based on probabilistic classification and a harmonic plus noise model.* ICASP, 1998. **1**: p. 281-284.

22. Benasich, A.A., and Tallal, P., *Infant discrimination of rapid auditory cues predicts later language impairment.* Behav Brain Sci, 2002. **136**: p. 31-49.

APPENDIX A

USER'S MANUAL

Requirements

- The software requires MathWorks' Matlab Release 2011b or later.
- An internet connection to call the online text-to-speech engine.
- A media player needed to be installed. If the operating system is windows, then the media player will be already available.

The software has been tested only on windows operating system.

Starting the audiometer

To run the audiometer, insert the supplied CD into the drive, and:

- From My computer, right click the CD drive. Select the "Audiometer\speech" from the pop-up menu.

or

- From the Run option in the Windows Start Menu, type H:\Audiometer\speech.m where H is the CD drive.

The GUI will then be loaded in the Matlab.

Using the audiometer

The following window will appear:

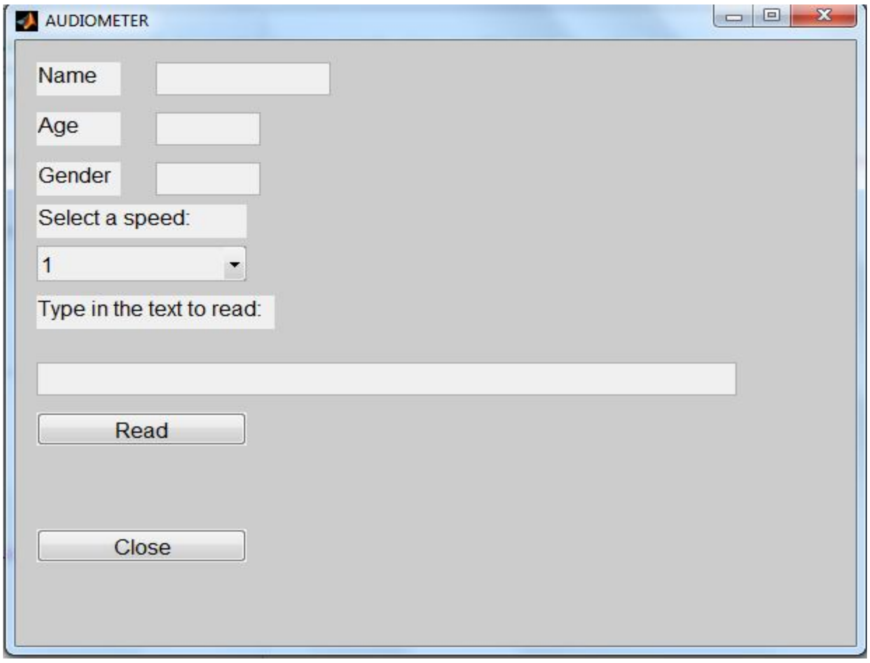

Figure A.1

Select a speed

The play speed can be selected from the pop-up menu. The menu lists the speed only the audiometer has programmed to delay.

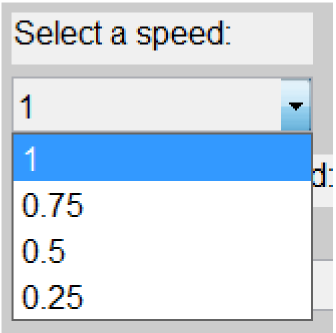

Figure A.2

Enter a text and read

Enter the text to be heard in the text edit field and press the read button to hear the text.

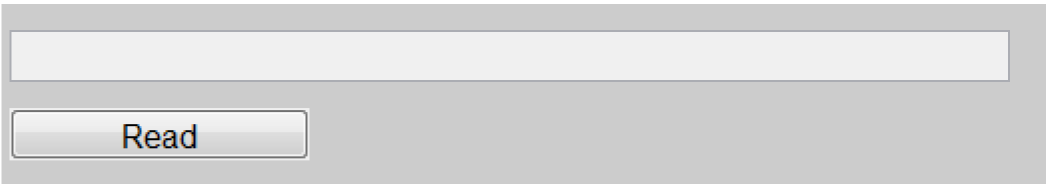

Figure A.3

Close the audiometer

Press the close button to quit the application.

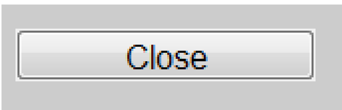

Figure A.4

APPENDIX B

ALGORITHM

```
function audiometer
% This is a project on how to use the ActiveX VideoLAN.VLCPlugin.2 in
combination with the unofficial google text
% to speech (TTS) engine to generate speech from text.
% You need to install the Media-Player and have a internet connection to call the
Google service.

% Programmed by Annie Pushpam Packiaraj
% Date: 2012/08/31
clear all;

% Create a small figure
fig = figure('units','characters','position', [100 20 120 30]
,'Toolbar','none','DoubleBuffer', 'on', 'Name', 'AUDIOMETER', 'NumberTitle',
'off','MenuBar','none');

% insert the mediaplayer invisible to figure
actx = actxcontrol('WMPlayer.OCX.7', [1 1 0 0], fig);


% insert an speed popup
mh = uicontrol(fig,'Style','popupmenu',...
        'units','characters',...
        'String',{'1','0.75','0.5', '0.25'},...
        'Value',1,'Position',[3 22 30 2]);
set(mh,'FontSize',12);

% insert Name label
st = uicontrol(fig,'Style','text',...
        'units','characters',...
        'String','Name',...
        'Position',[3 33 12 2], 'HorizontalAlignment','left');
set(st,'FontSize',12);

% insert an edit box for name
zt = uicontrol(fig,'Style','edit',...
        'units','characters',...
        'String','',...
        'Position',[20 33 25 2],'HorizontalAlignment','left');
set(zt,'FontSize',12);

% insert Age label
ft = uicontrol(fig,'Style','text',...
        'units','characters',...
```

```
            'String','Age',...
            'Position',[3 30 12 2], 'HorizontalAlignment','left');
set(ft,'FontSize',12);

% insert an edit box for age
xt = uicontrol(fig,'Style','edit',...
            'units','characters',...
            'String','',...
            'Position',[20 30 15 2], 'HorizontalAlignment','left');
set(xt,'FontSize',12);

% insert Gender label
pp = uicontrol(fig,'Style','text',...
            'units','characters',...
            'String','Gender',...
            'Position',[3 27 12 2], 'HorizontalAlignment','left');
set(pp,'FontSize',12);

% insert an edit box for gender
kt = uicontrol(fig,'Style','edit',...
            'units','characters',...
            'String','',...
            'Position',[20 27 15 2], 'HorizontalAlignment','left');
set(kt,'FontSize',12);


% insert label for popup
sth = uicontrol(fig,'Style','text',...
            'units','characters',...
            'String','Select a speed:',...
            'Position',[3 24.5 30 2], 'HorizontalAlignment','left');
set(sth,'FontSize',12);

% insert an edit-box
eth = uicontrol(fig,'Style','edit',...
            'String','',...
            'units','characters',...
            'Position',[3 15 100 2], 'HorizontalAlignment','left');
set(eth,'FontSize',12);

% insert label for editbox
sth2 = uicontrol(fig,'Style','text',...
            'units','characters',...
            'String','Type in the text to read:',...
            'Position',[3 19 34 2], 'HorizontalAlignment','left');
set(sth2,'FontSize',12)
```

```
% Insert Read Button
pbhr = uicontrol(fig,'Style','pushbutton','String','Read',...
        'units','characters',...
        'Position',[3 12 30 2], 'Callback', {@read_callback,mh, eth, fig});
  set(pbhr,'FontSize',12);

% Insert close Button
pbhc = uicontrol(fig,'Style','pushbutton','String','Close',...
        'units','characters',...
        'Position',[3 5 30 2],'Callback','close');
  set(pbhc,'FontSize',12);



function read_callback(hObject, eventdata, mh, eth, fig)
% Get Language
val = get(mh,'Value');
 lang = 'en';
 switch val
   case 1
      actx = actxcontrol('WMPlayer.OCX.7', [1 1 0 0], fig);
      actx.speed = 1;
      media = actx.newMedia(['http://translate.google.com/translate_tts?tl=',lang
,'&q=', str] );  % Unofficial text to speech  by google
      actx.CurrentMedia = media;
      actx.Controls.play;
   case 2
      actx = actxcontrol('WMPlayer.OCX.7', [1 1 0 0], fig);
      actx.speed = 0.75;
      media = actx.newMedia(['http://translate.google.com/translate_tts?tl=',lang
,'&q=', str] );  % Unofficial text to speech  by google
      actx.CurrentMedia = media;
      actx.Controls.play;
   case 3
      actx = actxcontrol('WMPlayer.OCX.7', [1 1 0 0], fig);
      actx.speed = 0,5;
      media = actx.newMedia(['http://translate.google.com/translate_tts?tl=',lang
,'&q=', str] );  % Unofficial text to speech  by google
      actx.CurrentMedia = media;
      actx.Controls.play;
   case 4
      actx = actxcontrol('WMPlayer.OCX.7', [1 1 0 0], fig);
      actx.speed = 0.25;
      media = actx.newMedia(['http://translate.google.com/translate_tts?tl=',lang
,'&q=', str] );  % Unofficial text to speech  by google
      actx.CurrentMedia = media;
      actx.Controls.play;
end;
```

```
% Get String to Read
str_read = get(eth,'String');

% Prepare String:
% limit string to at least 100 characters and replace blanks
try
    str = strrep(str_read(1:100),' ','+');
catch
    str = strrep(str_read,' ','+');
end;
```