Introduction

Understanding speech in noise is one of the most complex activities encountered in everyday life, it is an important task for successful participation in educational and social environments. It poses particular challenge for children with hearing impairment. Successful perception of speech in noise is dependent on cognitive factors as well as sound processing at peripheral, subcortical and cortical level.

Despite the sophisticated algorithms in the hearing aids and advances in C.Is technologies to help alleviate the difficulty of understanding speech in background noise, individuals with hearing loss continue to experience problems.

There has been considerable controversy in the literature about the reasons for the difficulties in speech understanding. Some studies have suggested that the difficulties arise primarily from reduced audibility. Others argued that the difficulty understanding speech arises, at least partly, from deficits in the ability to discriminate sounds that are well above absolute threshold. Examples of such deficits include reduced frequency selectivity

(Papakonstantinou et al., 2011) and deficits in temporal processing abilities **(e.g., Lorenzi et al., 2006).**

To the extreme end of poor speech perception lies the auditory neuropathy spectrum, where the signal from the inner hair cells or the neural connections from them are compromised. Perceptual abnormalities seen in people with auditory neuropathy include severe temporal processing impairments **(Starr et al., 1996)** and extremely impaired Frequency discrimination abilities **(Rance et al., 2004)** leading to severe difficulty with speech in background noise **(Zeng & Liu, 2006)**.

To better understand why individuals with a hearing loss struggle with speech perception in background noise, we must first investigate the effects of background noise on a normally functioning auditory system.

Extracting vital information from complex auditory signals like speech relies on multi-staged processing within the peripheral and central auditory system **(Pichora-Fuller and Singh, 2006).** Therefore, any abnormalities in these mechanisms and pathways can lead to communication breakdown **(Schneider et al., 2010).**

Electrophysiological responses recorded in response to speech stimuli; provide an objective tool to evaluate the neural encoding along the auditory pathway, and localize any potential disorder. By analysing recordings made with different objective tools (e.g. c-ABR & P1-CAEP) it may be possible to determine whether deficits are due to abnormal neural encoding in ascending auditory pathways or difficulties with higher order processes needed for performing speech perception in noise **(Billings et al., 2013)**. Moreover, these measures could provide predictive valuable information about neuronal signal encoding in noise **(Billings et al., 2009)**.

The c-ABR is well suited for assessing hearing loss effects on neural processing of speech. Its fidelity to the stimulus can be seen in representation of timing features (onsets, offsets), pitch (encoding of the fundamental frequency (F0) of the stimulus), and timbre (representation of formants) through cycle-by-cycle neural phase-locking **(Skoe and Kraus, 2010).**

On the other hand, cortical responses reflect the summation of excitatory post-synaptic potentials originating from multiple generator sites in response to stimulus onset of the stimulus. These slow dendritic events can be separated by several milliseconds thus still can sum constructively even with variable degrees of neural dys-synchorony.

This work is to highlight the impact of hearing loss (either cochlear or auditory neuropathy spectrum) on neural encoding of speech. Also to study the effect of background noise on subcortical and cortical speech encoding in both normal and hearing impaired children and correlate these neurophysiological results with behavioural measures. This may provide better insight towards rehabilitation philosophies for overcoming difficulties associated with hearing loss, as well as shaping counseling and intervention.

Aims of the Work

1. To explore the effect of hearing loss (both cochlear & auditory neuropathy spectrum) on neural encoding of speech by measuring p1 and speech ABR.
2. To investigate the effects of background noise on both evoked responses (p1 and speech ABR).
3. To study the relationship between speech ABR and the behavioural word recognition test (WRT) both in quiet and in noise.

Chapter (1):

Neural Encoding of Speech

One of the most valued uses of hearing is for communication. Oral communication is dependent on correctly received and understood information. This speech understanding process depends on a subject’s ability to detect, discriminate and perceive the individual sounds of the language. This involves the peripheral sense organ as well as central auditory pathways and cognitive functions such as working memory capacity and selective attention.

Basic Acoustics of sounds

Basically starting, Acoustic signals are made up of sound waves, which are variations in air pressure over time. Waveforms can be described in terms of complexity, simple or complex, or in terms of periodicity, periodic or aperiodic.

An example of the simple waveform is a pure tone which consists of a single sine wave. Complex waveforms are comprised of multiple simple waveforms and can be further classified by their spectro-temporal characteristics as steady-state or time-varying, as seen in Figure 1. Both steady-state and time-varying stimuli can further be characterized by their linguistic relevance, as speech or non-speech entities.

An example of a complex steady-state non-speech stimulus might be a musical note. In contrast, a linguistically relevant complex steady-state stimulus is a synthesized vowel consisting of several formants or a natural vowel that has been resynthesized so that its formants are flattened to a specific frequency. Complex time-varying, non-speech stimuli include noise, while linguistically relevant time-varying stimuli include formants, formant transitions, vowels, diphthongs, CV syllables, and speech in general **(Somers, 2016).**

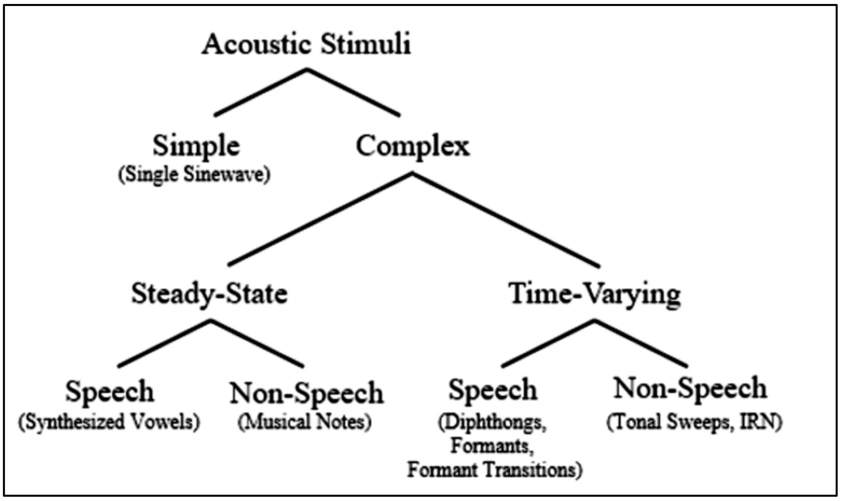

Figure (1): Flowchart showing the differences between simple and complex waveforms and their various subtypes **(Somers, 2016).**

Acoustical properties of speech

Speech is an inherently complex auditory signal and is composed of numerous acoustic features which influence how a particular stimulus is neurally encoded in the auditory system. Non-linguistic acoustic features of speech include the fundamental frequency, harmonics, formants, formant transitions, acoustic onsets, periodicity and the speech envelope **(Abrams & Kraus, 2009).**

The **fundamental frequency F0** may be defined in one of two ways. Physically, the fundamental frequency is the frequency of repetition of the lowest frequency component of a given complex waveform **(Abrams and Kraus, 2015)**. In the field of acoustic phonetics, the fundamental frequency F0 is determined by rate of vocal fold vibration. All voiced speech has energy at integer multiples of the fundamental frequency (determined by glottal pulse rate) called **harmonics (Liberman 1954).**

Psychoacoustically, the F0 correlates with pitch while its harmonics underlie the perception of timbre **(Fant, 1960).** Pitch cues are used to extract prosodic information from speech, enabling the listener to perceive emotional affect and linguistic meaning (i.e. questions and statements) while timbre gives a speech stimulus its characteristic quality i.e. raspy, nasal, etc **(Russo et al., 2009).**

Moreover, F0 has an important role in speech-in-noise perception, which is mediated by speaker identification and object formation (**Oxenham, 2008; Shinn-Cunningham & Best, 2008). Brokz and Nooteboom (1982**) suggested that listeners tend to group together components that arise from the same source (same F0) in a process called **source segregation.**

Another mechanism called **Stream segregation**, or the ability to separate sounds that arise from different channels (different F0) **(Oxenham, 2008).**

After being produced by vocal fold vibration, the sound is then filtered by the vocal tract. The vocal tract has its own characteristic resonant frequencies based on its length and shape. As the sounds passed through the vocal tract filter, certain frequencies in the source signal will be enhanced (Figure 2). These enhanced resonant frequencies are termed **formants.** The lowest frequency formant is known as the first formant and is notated F1, while subsequent formants are notated F2, F3, etc **(Abrams & Kraus, 2009).**

Formants and formant transitions are critical for speech sound identification, with the lowest three formants capable of conveying enough information for successful vowel and consonant identification. Formant also has an important role in speech identification in noise. As formant frequencies shape the speech spectrum creating "**spectral peaks**" provide the listener with phonetic information. These spectral peaks become essential cues especially when listening in background noise or when frequency resolution is degraded **(Assmann & Summerfield, 1989).**

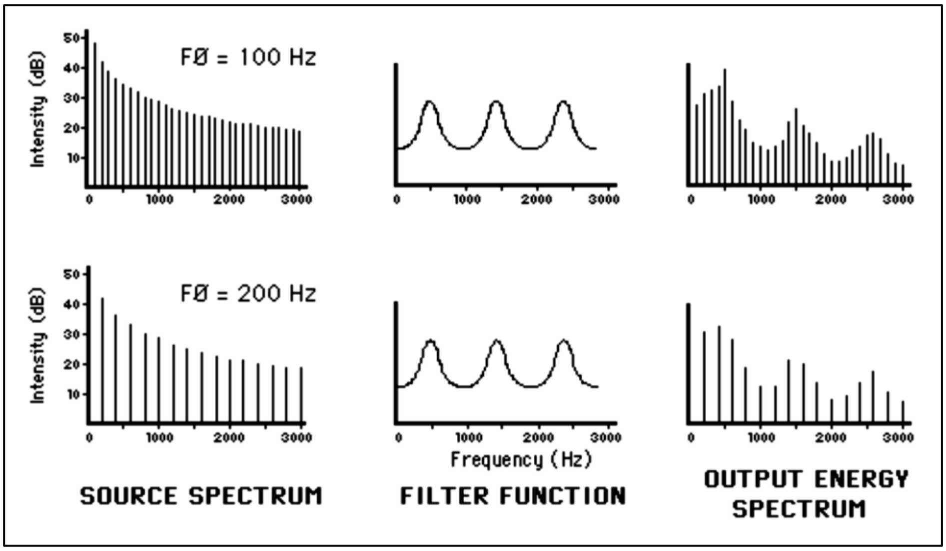

Figure (2): Source-filter model of speech production **(Fant, 1960).**

Speech perception

Speech perception is a complex process that involves multiple stages of signal processing. The cochlea plays a vital role in speech perception. Once the cochlea is damaged, the ability to process speech in noise is seriously degraded.

The main functionalities of the cochlea are to separate the frequency content of the speech stimuli from the movement of the hair cells that correspond to specific regions of the basilar membrane. Also to compress the large acoustic intensity range into the much smaller mechanical and electrical dynamic range of the inner hair cell. The auditory neurons then convert the signal into neural spikes and send them to the central auditory system.

Cochlear encoding of speech sounds:

The encoding of any sound stimulus, whether simple or complex begins in the peripheral auditory system, at the cochlea, specifically at the level of the basilar membrane. The basilar membrane is composed of a series of tonotopically organized overlapping band pass filters **(Moore, 2008).** These band pass filters consist of a lower and upper frequency cut-off, which allow a certain frequency or group of frequencies to pass through these cut-offs. Further, filters located at the base of the cochlea are wider in range then the filters located at the apex of the cochlea (Figure 3).

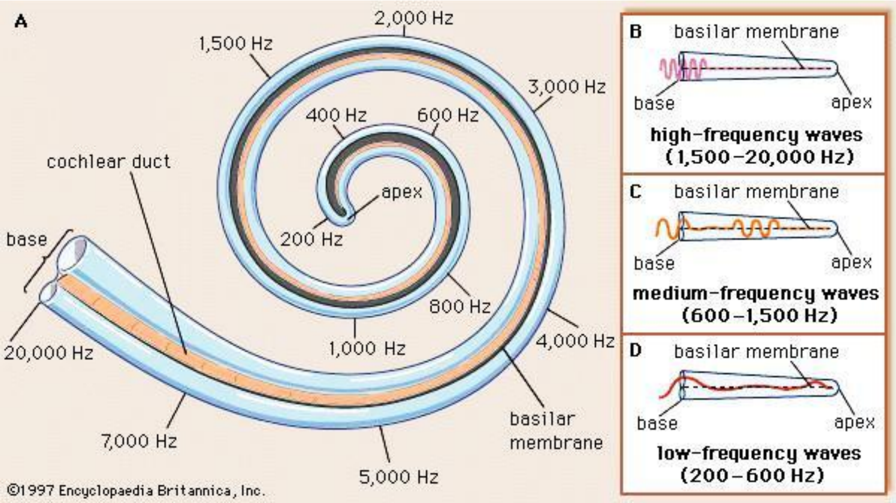

Figure (3): Schematic illustration of the analysis of sound frequencies by the basilar membrane. (a) The basilar membrane becomes progressively wider and more flexible from the base of the cochlea to the apex. So, each area of the basilar membrane vibrates preferentially to a particular sound frequency.

These cochlear filters are logarithmically spaced, but the spacing of the harmonics of complex stimuli is linear. This results in lower harmonics each passing through individual cochlear filters while multiple higher harmonics may pass through a single cochlear filter in the high-frequency region Figure 4 **(Sayles & Winter, 2008)**.

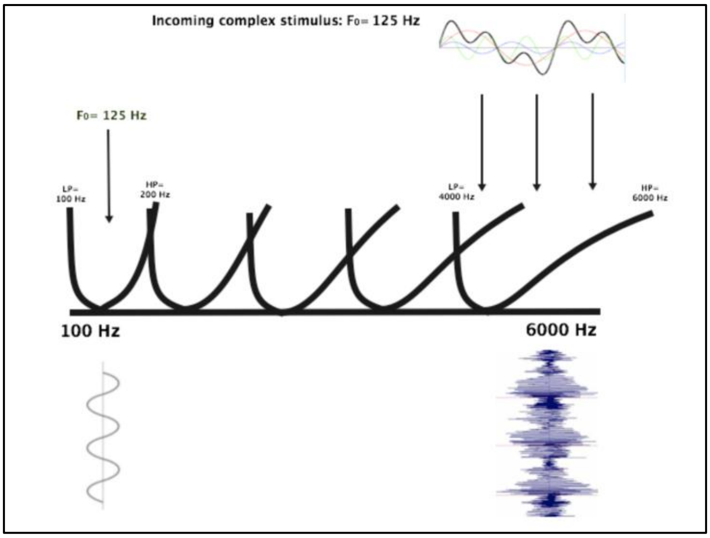

Figure (4): A complex stimulus as it is separated and sent through their respective filters **(Somers, 2016).**

This complex waveform output is composed of a slow-varying envelope superimposed on a rapidly-varying fine structure as depicted in Figure 5 below.

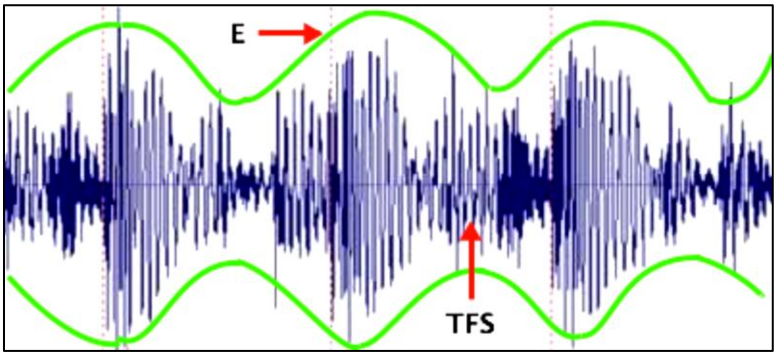

Figure (5): Envelope (E) as it modulates the complex temporal fine structure (TFS) waveform **(Somers, 2016).**

This **temporal fine structure (TFS)**; representing the high frequency content of the stimulus such as the harmonics and frequency formants is modulated by the overall rate of the **envelope (E)** representing the low frequency fundamental of the stimulus **(Moore, 2008).**

Envelope (E) & Temporal Fine Structure (TFS) Cues:

In **1981 Remez et al.** suggested a new classification of acoustic cues essential for speech perception based on spectral and temporal properties of the acoustic signal. This classification contains two types of temporal information: (TFS) and envelope (E) **(Lorenzi et al., 2006).**

The E is related to the fundamental frequency of the stimulus (F0), while TFS is related to the formants of the stimulus **(Rosen, 1992).** The TFS is often described as a ‘‘carrier’’ while the E is described as an amplitude modulator applied to the carrier.

The speech envelope provides phonetic and prosodic cues to the duration of speech segments, manner of articulation, the presence (or absence) of voicing, syllabication, and stress. E cues alone can lead to high intelligibility for speech in quiet both for normal-hearing people **(Shannon et al., 1995, Loizou et al., 1999 and Smith et al., 2002)** and for hearing-impaired people **(Turner et al., 1995 and Souza and Boike, 2006).**

However, the E cues alone do not allow effective listening in the dips of a background sound **(Lorenzi et al. 2006).**

On the other side, the TFS is linked to listening in the background dips, melody identification and pitch perception as well as sound localization **(Qin and Oxenham, 2003, 2006; Nelson et al., 2003; Stickney et al., 2005; Fullgrabe et al., 2006).**

These two spectral properties, the envelope and the temporal fine structure, are further encoded at the level of the 8th nerve and brainstem where a phenomenon called “phase locking” takes places.

8th nerve encoding of speech sounds:

When stimulated by an acoustic signal, the auditory nerve activity becomes phase-locked to certain components of the stimulus. Phase-locking is the ability of a neuron to fire at intervals corresponding to the fundamental and formant frequencies of a stimulus **(Hall, 2007).**

Recall that the period of the stimulus is the reciprocal of the fundamental frequency. For example if the frequency of the stimulus is 500 Hz, the period of the stimulus (1000 ms/ 500 Hz) is 2 ms. Therefore, groups of neurons will discharge at 2 ms intervals (2 ms, 4 ms, 6 ms, 8 ms, etc….) (Figure 6).