RHYTHM PERCEPTION WITH FINE STRUCTURE CUES - A SIMULATED STUDY

Ramia (Narayanan)
Register No: 06AUD009

A Dissertation Submitted in Part Fulfillment of
Final year M.Sc (Audiology),
University of Mysore, Mysore.

ALL INDIA INSTITUTE OF SPEECH AND HEARING
NAIMISHAM CAMPUS, MANASAGANGOTHRI
MYSORE-570006

This work of mine is dedicated to my dearest parents without whom my very existence is meaningless...

CERTIFICATE

This is to certify that this dissertation entitled ***"Rhythm perception with fine structure cues — A simulated study"*** is the bonafide work submitted in part fulfillment for the degree of Master of Science (Audiology) of the student (Registration No.06AUD009). This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other University for the award of any other Diplomа or Degree.

V. basa

Mysore **Dr. Vijayalakshmi Basavaraj**
Director
April, 2008 ***All India Institute of Speech and Hearing***
Naimisham Campus
Manasagangothri
Mysore-570 006.

CERTIFICATE

This is to certify that the dissertation entitled *"Rhythm perception with fine structure cues - A simulated study"* has been prepared under my supervision and guidance. It is also certified that this has not been submitted earlier in any other University for the award of any Diploma or Degree.

Guide

Mysore

April, 2008

(Dr. K.Rajalakshmi)
Reader in Audiology
Department of Audiology
All India Institute of Speech and Hearing,
Mysore-570006.

DECLARATION

This is to certify that this dissertation entitled *"Rhythm perception with fine structure cues -A simulated study"* is the result of my own study under the guidance of Dr. K. Rajalakshmi, Reader in Audiology, Department of Audiology, All India Institute of Speech and Hearing, Mysore, and has not been submitted in any other university for the award of any diploma or degree.

Mysore

April, 2008 Register No.06AUD009

ACKNOWLEDGEMENT

I express my heartfelt gratitude to my Guide Dr.K.Rajalakshmi, Reader., Department of Audiology, for her guidance and immense support.

I thank the Director, Dr. Vijayalakshmi Basavaraj, for permitting me to complete the study.

I thank Prof Asha Yathiraj, HOD audiology for having permitted me to use the instruments..

I Thank Savithri ma' am for clearing my doubts and queries...

A Heartfelt thanks to Asha ma 'am , Animesh sir, Vanaja ma 'am, Manjula ma 'am and Rajalakshmi ma 'am for the knowledge imparted to me and for what I am today... thank u..

Savithri ma 'am, Yeshoda ma 'am, Manjula ma 'am, Rohini ma 'am,Shymala ma 'am, Geetha ma 'am and Prema ma 'am... thank u...

/ also acknowledge my heartfelt thanks to Mr. Vijaya Kumar Name helping me throughout the course of this dissertation... "Thanks " is not a word adequate for you sir...

My Achan and Amma...you are always my source of inspiration...you are the best.:-)

Kane tan and thakku...my critics.. You have made me what I am today :-)

Sunil... though I haven't known you for long....u have made your presence felt in my life..Be the same...:-)

Kavitha, sheena and Rajani,, though you guys aren 't here,,your trust in me is unwavering..n I miss u loads...

Rubina, Anushya, Selvi, Bindu, Dhanya...unforgettable dinner time...I'll miss all the enjoyment we had...

Sandhya, Shweta, Jijo, Vignesh, Manuj, Chandu (Taklu)...my best posting mates...

Rubina, thanks for all what you have done for me n tolerating me...you made me realize my mistakes and my talents...

Abhijeet (Motta), Vignesh, Rubina, Anushya and Sreejyothi...my internship batch mates...the best time in AIISH.

Noble etan, Subhash ettan, Nishad, And akka... thank you for being there whenever I needed you :-)

A special thanks to Vasanthalakshmi ma 'am... ma 'am you are great at what you are doing...

My dearest classmates ...every day was fun with you all...

A hearty thanks to my juniors and seniors for all the help.. A special thanks to Swapna Raj

Thanks to Shivappa for your timely help in printing and binding works..

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER: 1

INTRODUCTION

Cochlear implants (CI) are devices that have been successful in restoring hearing to profoundly deaf patients through electrical stimulation of the auditory nerve with fine electrodes inserted into the scala tympani of the cochlea. The performance of cochlear implants depends on the speech processor's ability to faithfully decompose speech signals into a number of channels of narrow-band electrical signals that is used to activate the spiral ganglion cells of the auditory nerve. The number of electrodes in modern cochlear implant devices may be different from the number of channels in the speech processor, and may vary from 12 to 22 or more depending on the strategy used or with the monopolar or bipolar stimulation used. The configuration and placement of the electrodes are of importance to the overall performance of these devices.

Components of music perception:

Every person is immersed in an environment full of sound and being able to understand speech is not the only function of hearing. For most people, listening to music is also a significant and enjoyable experience. Natural speech carries abundant acoustic cues in both spectral and temporal domains (Smith, Delgutte & Oxenham, 2002), while music is very complex and wide ranging. Musical perception is based on rhythm, timbre, and pitch. Rhythm is related to the temporal characteristics of the sound. Timbre is more difficult to describe but can be defined as the quality of a sound by which a listener can tell that two sounds of the same loudness and pitch are dissimilar (Fragoulis, 1999). Pitch conveys melody and is strongly related to the spectral content of the sound. For accurate musical perception, all three of these characteristics must be transmitted. Rhythm, pitch, timbre and melody identification requires good functioning of highly specific patterns of discrimination.

Rhythm: temporal patterns in musical sounds along with overall variations in loudness are perceived as the rhythm.

Melody: the ability to recognize melody depends on highly variable factors, such as individual's musical training, the socio-cultural experience, memory of the tune and situational contexts. The ability to perceive accurately the fundamental features of musical sounds, such as pitch and temporal patterns is a pre-requisite for melody recognition.

Timbre: the principal properties of the frequency spectrum and the amplitude envelope of sounds, including changes in those attributes over time and frequency are also relevant.

Music perception in Cochlear implants

Investigators have shown that trends in the patterns of correlation between speech and music perception suggests that particular structural elements of music are differently accessible to cochlear implant users (Gfeller & Lansing, 1997; McDermott & Mckay, 1997). Pitch cues can be elicited using two independent mechanisms in cochlear implants, namely place pitch and temporal pitch cues (Tong, Clark & Blamey, 1982; Shannon, 1983; Mckay, McDermott & Clark, 1996). Temporal pitch cues arise when the repetition rate of stimulation on one channel changes. The temporal pitch sensation rises with increasing rate upto 300Hz but saturates at higher rates (Shannon, 1983). Place pitch cues arise when the site of stimulation is changed while keeping the stimulation rate constant, with more basal stimulation eliciting higher pitches. It has been shown that both temporal and place pitch cues enable CI recipients to some degree perceive Fo differences of synthetic harmonic sounds with currently used sound processors (Geurts and Wouters, 2001), and in normal hearing subjects with acoustic models of CI processors (Green et al, 2004). Rhythm perception and melody perception was observed to be poorer than normal hearing counter parts due to inadequate cues provided through the cochlear implant (Gfeller and Lansing, 1991).

Music perception in CI subjects was poor compared to normal hearing subjects (Leal, 2003; Gfeller & Lansing, 1991). Pijl (1997) showed that atleast part of the limited pitch performance of the CI subjects was due to ineffective sound processing as CI subjects were able to estimate musical intervals more accurately for synthetic stimuli(pulse trains) than for real musical sounds. Part of this limited Fo discrimination ability in CI recipients using their speech processors may be due to the limited transmission of the fine temporal information in current speech processors, since most current speech processors only extract the slowly varying envelope and use this to modulate a constant rate pulse (Smith, Delgutte & Oxenham, 2002).

The temporal envelope consists of frequency information in the 2-50Hz range; periodicity consists of frequency information in the 50-500 Hz range and the fine structure consists of frequency information in the 600-10,000 Hz (Rosen, 1992).

It is commonly believed that envelope cues are represented in the auditory system as fluctuations in the short-term rate of firing in auditory neurons, while temporal fine structure(FM) is represented by the synchronization of nerve spikes to a specific phase of the carrier (phase locking) (Shannon, Zeng, Kamath, Wygonski, & Ekelid ,1995).

Current signal processing for cochlear implants allow adequate speech perception in quiet environments for most users. However, their speech recognition performance in noise and music perception is severely limited. Typically, each electrode receives an amplitude modulated pulse train representing the narrow band temporal envelope of a sound from a particular frequency band. Amplitude modulations from low frequencies are delivered to the apical electrodes and amplitude modulations from high frequencies are delivered to the basal electrodes. Though amplitude modulations are sufficient to support sentence recognition in quiet (Shannon, 1995), it is not sufficient for speech recognition in noise and music perception (Dorman, Loizou & Tu, 1998; Freisen, Shannon, Baskent & Wang, 2001). The pitch of the voice can be conveyed by the temporal envelope; however this cue provides a relatively weak representation of pitch especially for the high frequencies. This indicates that pitch information is not effectively coded by the envelope modulations and hence not sufficient for rhythm perception (Green & Rosen, 2004).

Need for the study:

The speech processing strategy in most modern cochlear implants extracts and encodes only amplitude modulation in a limited number of frequency bands and hence the place pitch coding is impaired. To overcome the inherent limitations of temporal pitch coding in cochlear implants, techniques may be needed to provide perceptual information about the "fine structure" of the acoustic signals. The fine structure contains rapidly varying components of sounds that are not present in the envelope coding strategies that are being used. Amplitude modulations alone do not provide good rhythm perception as described above; hence it is proposed to study the importance of adding fine structure cues in rhythm perception. .

In addition, investigators demonstrate that number of channels also affects music (rhythm) perception. Here, we propose to understand how a novel speech processing strategy that encodes frequency modulations with varying number of channels improves cochlear implant performance in rhythm perception.

Aim of the study:

The study aims to understand the

(1) Effect of number of channels on rhythm perception with envelope cues (amplitude modulation).

(2) Effect of number of channels on rhythm perception with fine structure and envelope (amplitude and frequency modulation) cues.

(3) To compare rhythm perception with the fine structure and envelope (AM + FM) condition and envelope only (AM) condition across different channels.

CHAPTER: 2

REVIEW OF LITERATURE

Leal, Shin, Laborde (2003) compared music perception and speech perception in noise in 29 multichannel CI users. The study concluded that individuals who have good speech perception in the presence of noise had better rhythm identification and discrimination scores (60-70%), but those individuals with poor speech identification scores in the presence of noise had poor rhythm identification and discrimination scores. In addition, investigators have also demonstrated that increase in number of channels improves the speech identification scores in quiet and noise (Stickney, Nie & Zeng, 2005; Freisen, 2001; Carroll & Zeng, 2007).

Kong, Cruz, Jones and Zeng (2004) showed differences in performance between four normal-hearing and five cochlear implant listeners in the melody identification task with and without rhythmic cues. While the normal hearing listeners achieved near perfect performance for the rhythm (98.3%) conditions, the cochlear implant listeners performed less accurately than the normal-hearing listeners in the rhythm condition (63.2%).

Generally, the number of frequency bands equals the number of implanted electrodes that is used to present electric stimulus to the cochlea depending on the mode of stimulation. With such rough estimation of the short term spectrum, certain acoustic features of complex signals are degraded or eliminated in the processing, making it difficult or impossible for implant users to perceive some subtle characteristics of sound. McDermott (2005) demonstrated in 10 adult cochlear implant users that on average, the performance and the quality of musical sounds are generally inferior to that of normally hearing individuals.

Laneau, Woulters & Moonen (2004) demonstrated the utility of place and temporal pitch cues for Fo discrimination. Fo discrimination based on purely place pitch cues (cutoff frequency of the low pass filter was set to 10 Hz) are weak but when temporal cues (cutoff frequency of the low pass filter was set to 200Hz) were added to the stimuli, the performance improved significantly and the jnd's for Fo discrimination ranged from 6% - 60% depending on the subject, processing condition and the reference Fo. The performance of cochlear implant subjects is still very poor compared to the performance of normal hearing subjects where jnd's for Fo are typically less than 1%.

Laneau, Woulters & Moonen (2006) demonstrated that a new processing scheme for CI, (Fo mod) enhanced pitch and music perception in six experienced postlingually deafened CI recipients. In this strategy all channels were modulated sinusoidally at F0 with 100% modulation depth and in phase across all channels to maximize temporal envelope pitch cues.

Smith, Delgutte & Oxenham (2002) showed a new dichotomy in auditory perception between temporal envelope and fine structure cues in six native English speakers with normal hearing. They first digitally filtered a wideband signal 80-8820 Hz into 1-64 frequency bands and then used the Hilbert transform to decompose the band-limited signal into a slowly varying envelope component and a fast varying fine-structure component. To assess relative contributions of temporal envelope and fine structure to auditory perception, they produced "auditory chimaeras" by mixing one sound's envelope with another sound's fine structure. By conducting listening tests with the chimaeric sounds, they found that the envelope is most important for speech reception, and the fine structure is most important for pitch perception which in turn also helps in rhythm perception.

Smith, Delgutte & Oxenham (2002) found that, when 4 to 16 frequency bands were used, recognition of English speech was dominated by the envelope, whereas recognition of melody was dominated by the fine structure.

Gfeller & Lansing (1991) demonstrated that rhythmic and melodic elements of music are differently accessible to cochlear implant users and that notable qualitative differences between different processors may exist. The subjects of the study were 18 postlingually deafened cochlear implant users (Ineraid and Nucleus), the evaluative measures used in the study was primary measures of music audition (PMMA).

Lan, Nie, Gao & Zeng (2004) developed an algorithm to extract the fundamental frequency and identified the general pattern of pitch variations of four typical tones in Chinese language. To improve the performance in cochlear implant users who speak tonal language, the strategy extracted both the envelopes of the narrow band signals and the fundamental frequency of the speech signal, and used them to modulate both the amplitude and the frequency of the electrical pulses delivered to the stimulation electrodes. A comparison was made between novel strategy and CIS strategy in subjects with normal hearing. The novel strategy produced significant improvement in perception of Chinese tones, phrases and sentences.

Lorenzi, Gilbert, Garnier, and Moore (2006) hypothesized that the normal auditory system decides whether a speech signal in the dips of a background sound is produced by the target speech or is simply part of the background sound by using information derived from neural phase locking to temporal fine structure; changes in phase locking of auditory nerve discharges when a valley occurs indicate that a signal is present in the valley. This aids in speech perception in noise. Modifying cochlear implant processors to deliver fine-structure information might improve the pitch perception of implant patients, thus facilitating music perception and lexical-tone perception.

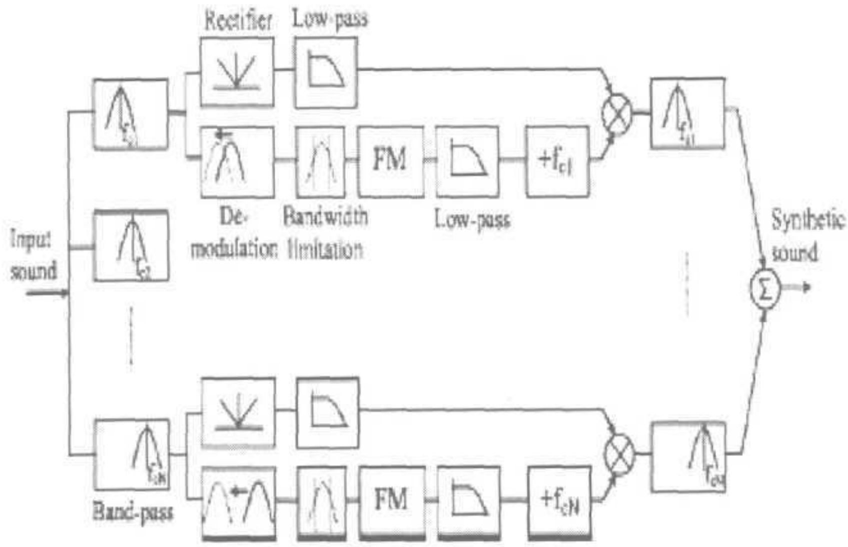
Formant transitions and voice pitch can be useful for segregating competing speech sounds. However, these cues are not adequately coded in current cochlear implant speech processing algorithms. Stickney, Nie & Zeng (2005) evaluated the use of FM cues in 24 normal hearing subjects and concluded that the addition of FM could potentially provide the transition and pitch cues. The slowly varying FM cue can be readily extracted from the temporal fine structure and may enhance cochlear implant performance for speech perception in noise as well as music perception.

Previous research has demonstrated that with cochlear implant users, speech recognition performance improves as the number of electrodes increase upto about 8 channels in quiet (Fishman, Shannon, Slattery, 1997; Friesen, Shannon, Baskent, Wang, 2001; Dorman, Loizou, Rainey, 1997). It remains unclear why the improvement in performance is limited upto 8 channels in cochlear implant users, while in normal hearing subjects with similar processing performance continues to improve as the number of channels increases (Friesen, Shannon, Baskent, Wang, 2001).

Zeng, Nie, Stickney, Kong (2004) evaluated thirty-four normal-hearing and 18 cochlear-implant subjects with 1, 2, 4 , 8 ,16, 32 bands with only AM cues and AM + FM cues. The study concluded that the utility of the AM cues are limited to ideal conditions, while the FM cues have a significant role in perception of tonal languages and rhythm perception. This finding further highlights the limitation of current cochlear implant speech processing strategies, as well as the need to encode the FM cues to improve speech recognition in noise, speaker identification, and tonal language perception.

Stickney, Nie & Zeng (2005) demonstrated in 24 subjects with normal hearing that AM+FM information on 5 frequency bands produced similar performance as 8 AM+FM bands.

Nie, Stickney, and Zeng (2005) evaluated the contribution of the additional FM information to speech recognition in noise in 40 subjects with normal hearing. The study suggests that frequency modulation be extracted and encoded to improve cochlear implant performance in realistic listening situations. They also developed the FAME algorithm to derive slowly varying amplitude and frequency modulations in speech sounds.



FIG, 1, Diagram of the speech processing strategy which combines AM and FM inforraatioQ, This speech processing strategy was used for the simula-doas,

The FAME algorithm has two independent parallel pathways used to extract the amplitude modulations (AM) and the frequency modulations (FM) in each band. The AM pathway extracts the slowly varying envelope, while the FM pathway extracts the slowly varying frequency modulation using a demodulator to remove sub band's center frequency. FAME algorithm is used in this study to segregate the envelope (AM) and fine structure cues (FM).

CHAPTER 3

METHOD

A. Subjects:

The study consists of 30 subjects with normal hearing in the age range of 18-25years (15 females and 15 males with the mean age of 21.5years). Pure tone audiometry thresholds at all frequencies were within 20 dBHL. No indication of middle ear pathology as shown by tympanometry and presence of acoustic reflexes at 500 Hz and 1000 Hz.

B. Instrumentation:

- A calibrated dual channel clinical audiometer and a calibrated Immittance meter were used to recruit subjects.
- A dual channel clinical audiometer was used for the testing.
- The stimulus generated using MATLAB 6.5 software was played using an Intel Pentium processor computer.

All the testing, both for selecting subjects and for experimental purposes were conducted in an air conditioned, acoustically treated double room set-up. The ambient noise levels inside the test room were within the permissible limits (re: ANSI S3.1 1991, as cited in Wilber 1994).

C. Stimulus and signal processing:

The stimulus consists of thirty nursery rhymes (30 rhymes; 5 rhymes in each of the six conditions). The simulation was done using MATLAB 6.5 software. The stimulus contains cues in the frequency range (100-10,000Hz). The stimulus was separated into its envelope (AM) and fine structure (FM) using FAME (frequency and amplitude modulation encoder) software. Frequency modulation (FM) was used to code the instantaneous frequency, or temporal fine structure of the speech waveform, while the envelope (AM) coded the instantaneous amplitude of the signal.

The stimulus was filtered into 4,8,12 narrow bands with third order Butterworth filters respectively. Each of the narrow bands was then subjected to full-wave rectification to obtain the slowly varying envelope followed by a low pass cut off filter at 500Hz(first order Butterworth filter) to control the amplitude modulation rate, to form the AM component of the synthesized signal. The envelope thus extracted was then modulated with sinusoids of 100 - 10,000Hz frequency range to produce the AM signal. The fine structure (FM) was extracted using FAME software. This was followed by low pass filtering (500Hz) to limit the FM depth and rate to relatively slowly varying components that can be perceived by the cochlear implant users, an optimum of 500Hz bandwidth and 400Hz rate was used (Stickney, Nie & Zeng, 2005). The processed signal was then band pass filtered to remove any frequency components that fall outside the analysis filters bandwidth.

Finally, the band passed signals were summed to form the synthesized signals that contain AM and FM components. The slowly varying envelope modulated with the sinusoids of 100Hz-10,000Hz formed the AM component of the signal.

D. Procedure:

The stimulus was presented in a sound treated room in the free field condition at 40 dBSL. The subjects were asked to tap the rhythm of the rhymes heard (Schutz & Kerber, 1993). All the thirty rhymes were presented to the thirty subjects randomly in the six conditions. The conditions include

(a) Amplitude modulated stimulus with 4 channels,

(b) Amplitude modulated stimulus with 8 channels,

(c) Amplitude modulated stimulus with 12 channels,

(d) Amplitude modulated and Frequency modulated stimulus with 4 channels,

(e) Amplitude modulated and Frequency modulated stimulus with 8 channels,

(f) Amplitude modulated and Frequency modulated stimulus with 12 channels.

The stimulus was presented and the participants were asked to tap out the rhythm heard. The rhythm pattern tapped was scored on a sheet for all the thirty (single foot) rhymes. This pattern was then compared with a model created from the tap pattern of three speech language patholo gists who were well versed in music. Each correct tap was scored one and the incorrect tap was scored zero.

Statistical analysis:

Repeated measure AN OVA with gender as the independent variable was administered to determine the main and interaction effects of all the variables in this study. To study the interaction effects paired t test was carried out.

The present study was designed to investigate the effect of adding fine structure cues (FM) on rhythm perception. The measurements were carried out in 30 individuals whose hearing was within normal limits. The subjects were asked to tap out the rhythm of the rhymes and the percentage of correct taps was calculated. The data was appropriately tabulated and statistically analyzed using SPSS (10.0) version. Repeated measure ANOVA was carried out.

CHAPTER 4

RESULTS AND DISCUSSION

The present study aimed at investigating the effect of adding fine structure cues in rhythm perception. All the speech coding strategies used with cochlear implants extract the envelope of the signal and code the information while missing out information on the fine structure cues. In this study, the fine structure cues were provided along with the envelope and the rhythm perception enhancement was studied. A group of 30 individuals with hearing within normal limits served as subjects for the study. The stimulus consisted of thirty rhymes which were modified by the MATLAB 6.5 software into envelope only (AM) and envelope + fine structure (AM+FM) conditions with 4, 8, 12 channels respectively.

The following effects were analyzed using the repeated measure ANOVA with age as the independent variable

(a) The effect of using only envelope (AM) cues in rhythm perception with respect to 4,8,12 channels.

(b) The effect of using fine structure (FM) cues along with envelope (AM) cues in rhythm perception with respect to 4,8,12 channels.

(c) Paired t-test was done to compare the effects of envelope only and envelope and fine structure cues with respect to 4,8,12 channels.

The results are as follows:

(a) Effect of only envelope cues on rhythm perception with 4, 8,12 channels:

A repeated measure ANOVA was performed with the number of channels and envelope cues (AM). There was a main effect of the number of channels. As expected the 12 channel condition produced higher scores than 8 channel condition, however there was no significant difference between 4 and 8 channels.

*AM represents envelope only condition.

*AMFM represents the condition where both envelope and fine structure cues are presented.

Figure 2: comparison of percentage correct scores across gender and 4,8,12 channels in AM and AM+FM condition.

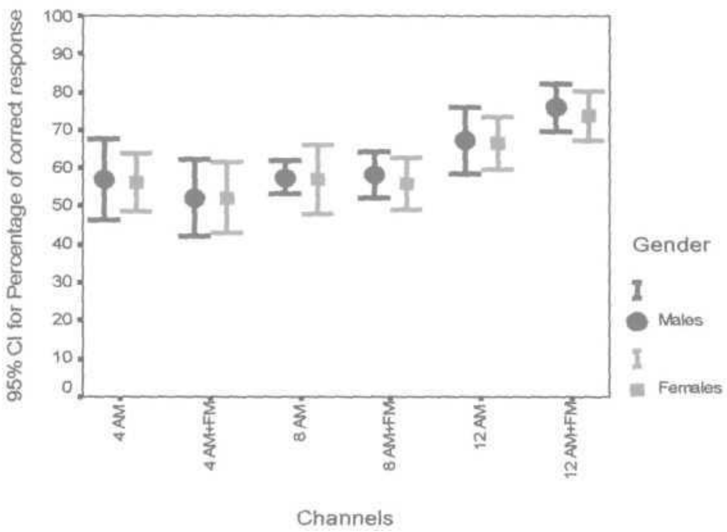


Figure 2 depicts the mean and the standard deviation as error bars for different channels across gender. From this figure it is evident that there is significant difference between the channels (4, 8,12 channels) in the envelope only condition.

Table: 1 mean and standard deviation values for the envelope only (AM) condition with 4, 8, 12 channels.

	Gender	Mean	Std. Deviation
P4AM	Males	57.0333	18.9374
	Females	56.2600	13.9391
	Total	56.6467	16.3427
P8AM	Males	57.4467	7.8940
	Females	57.1133	16.6533
	Total	57.2800	12.8061
P12AM	Males	67.2933	15.5387
	Females	66.5800	12.6156
	Total	66.9367	13.9114

Table: 2 pair wise comparison across channels (4, 8, 12 channels) in the AM (envelope) condition.

(I) AM	*(J)AM*	*Mean difference (I-J)*	*Sig.*
1	2	-.633	1.000
	3	-10.290	.079
2	1	.633	1.000
	3	-9.657	.029*
3	1	10.290	.079
	2	9.657	.029*

* The mean difference is significant at the .05 level.

It is evident from table 1 and 2 that F (2, 56) = 4.104, $p<0.05$ hence there is significant difference between channels (4, 8, 12 channels) in envelope only condition (AM).

From Boniferrin's multiple comparison tests there is significant difference between 8 and 12 channels but no significant difference between the other channels in the envelope only condition.

F(l,28) = 0.052,$p>0.05$ hence there is no significant difference between gender in envelope only (AM) condition with 4, 8, 12 channels.

F(2,56) = 0.002, $p>0.05$ hence there is no significant interaction between gender and channels (4,8,12 channels) in envelope only (AM) condition.

To summarize, there was no significant difference between 4 and 8 channels but there was significant difference between 8 and 12 channels.

The above statistical analysis shows that there is a significant difference between 8 and 12 channels when only amplitude cues are provided, but there is no significant difference between 4 and 8 channels. The absence of significant difference between 4 and 8 channels could be due to the large standard deviation as shown in table-1. This is in consonance with the earlier studies (Freisen, 2001; Stickney, 2006) which showed that increase in number of channels improves the speech identification scores in quiet and in the presence of noise in cochlear implant users. Leal et al (2003) demonstrated a significant correlation between speech perception in the presence of noise and rhythm identification and discrimination.

In addition, Nie et al (2006) demonstrated a significant improvement in all speech tests with increase in number of stimulating electrodes from 4 to 12. The improvement in the scores was more significant in the quiet than in the presence of noise.

Brill et al (1997) & Fishman et al (1997) demonstrated that 4-6 channels are adequate to support high levels of speech reception in quiet , they also report that further increase in number of channels do not produce an increase in speech test scores. This is true for speech in quiet, but doesn't hold good for rhythm perception, as a significant improvement in scores is seen with increase in number of channels. This improvement in the scores with increase in number of channels can be attributed to the increased spectral information provided through the channels.

In addition, Freisen et al (2001) demonstrated that though the number of channels was increased CI users weren't able to utilize the spectral information provided by the additional channels. The reason for this limitation was quoted as the electrode interactions and possible tonotopic shifts and warping in the frequency to place mapping of spectral information. No improvement was observed in speech recognition as the number of channels was increased from 7 to 20 channels for vowel and consonant recognition and increased from 10 to 20 channels for sentence recognition. However in this study, the effect of number of channels can be attributed to the normal physiological system of the subjects due to which they are able to make use of the fine spectral and temporal cues unlike the compromised physiological system in cochlear implantees (Souza & Boike, 2006).

Gfeller and Lansing (1991) examined the performance of 18 adult cochlear implant users on rhythm perception with the "primary measures of music audition"

(PMMA). The speech coding strategies used were SPEAK and F0F1F2 and the mean score for both schemes on the PMMA rhythm subtest was approximately 84% correct which is very close to the average scores obtained by a control group of 35 subjects with normal hearing (Gfeller et al, 1997).

However, Leal et al (2003) assessed rhythm perception by 29 recipients of the nucleus 24 electrode device. Twenty subjects used ACE while nine of them used SPEAK strategy. Rhythm discrimination and identification tasks were used. 24 of the 29 subjects obtained 75%-90% scores in the rhythm discrimination test while only 12 subjects obtained good scores in rhythm identification.

The statistical analysis also showed that there was no significant difference between gender and no significant interaction between the number of channels and gender. It is evident from the above studies, that rhythm perception is average in cochlear implant users and addition of fine structure cues could result in better rhythm perception due to the additional spectral and temporal cues available.

(b) Effect of number of channels (4, 8, 12 channels) on rhythm perception with envelope and fine structure cues:

Repeated measure ANOVA was performed with the number of channels and envelope + fine structure cues (AM+FM) conditions. There was a main effect of the number of channels. As expected the 12 channel condition produced higher scores than 8 channel condition, however there was no significant difference between 4 and 8 channels.

Table: 3 mean and standard deviation values for the fine structure + envelope (AM+FM) condition.

	Gender	*Mean*	*Std. Deviation*
P4AMFM	Males	52.2067	18.2989
	Females	52.1800	16.9078
	Total	52.1933	17.3107
P8AMFM	Males	58.1933	11.0864
	Females	55.9267	12.2017
	Total	57.0600	11.5125
P12AMFM	Males	75.8600	11.1537
	Females	73.8333	11.8526
	Total	74.8467	11.3552

Table: 4 pair wise comparison across channels (4, 8, 12 channels) in the AM+FM (envelope + fine structure) cues condition.

(1) AMFM	*(J) AMFM*	Wean *Difference (I-J)*	*Sig.*
1	*2*	-4.867	.223
	3	-22.653	.000*
2	1	4.867	.223
	3	-17.787	.000*
3	1	22.653	.000*
	2	17.787	.000*

* The mean difference is significant at the .05 level.

From table 3 and table 4 it is evident that $F(2,56) = 28.696$, $p<0.001$ hence there is significant difference between channels in the envelope with fine structure cues(AM+FM) condition.

Bonferroni's multiple comparison test showed a significant difference between 4 and 12 and 8 and 12 channels and no significant difference between 4 and 8 channels at 0.01 significance level in the AMFM condition.

$F(l,28) = 0.166$, $p>0.05$ hence there is no significant difference between gender in the AMFM condition.

$F(2,56) = 0.076$, $p>0.05$ hence there is no significant interaction between gender and channels in the AMFM condition.

It is evident that there is significant improvement in the scores as the numbers of channels were increased from 4 to 12 and 8 to 12 channels. But no significant differences in scores were obtained between 4 and 8 channels. There was no significant interaction between the number of channels and gender and no significant effect of gender on the scores as expected.

Similar to the previous condition, increase in the number of bands and addition of fine structure (FM) cues resulted in improved performance in rhythm perception task.

Frequency modulations derived from the temporal fine structure is important to support speech recognition in noise and other critical functions such as speaker identification, music perception, tonal language perception and sound localization (Smith, Delgutte & Oxenham, 2002). Similarly it was stated that additional FM(fine structure) cues helps the listener to better segregate the envelope of the target signal and hence helps in improved performance (Nie et al, 2005; Zeng et al, 2005).

In addition, Nie et al (2005) demonstrated improved performance in speech tests with increase in number of channels and addition of fine structure (FM cues).This improvement was significant in the presence of noise condition and hence applies for rhythm perception. Significant correlation was obtained between speech in noise and rhythm perception (Leal et al, 2003).

The results of this study are in support with the above studies which conclude that increase in the number of channels with the fine structure cues (FM) results in improved overall performance. This improvement in performance can be attributed to the increased spectral and temporal resolution provided in the envelope and fine structure (AM+FM) condition.

(c) To compare the effect of envelope cues (AM) only and envelope along with fine structure cues (AMFM) across 4,8,12 channels.

To compare the effect of additional FM cues (AM + FM) in rhythm perception with 4, 8, 12 channels paired t-test was done.

Table: 5 mean and standard deviation values across channels with AM only and AM+FM condition.

		Mean	*Std. Deviation*
Pair 1	P4AM	56.6467	16.3427
	P4AMFM	52.1933	17.3107
Pair 2	P8AM	57.2800	12.8061
	P8AMFM	57.0600	11.5125
Pair 3	P12AM	66.9367	13.9114
	P12AMFM	74.8467	11.3552

Table: 6 Paired Samples t Test across AM and AM+FM condition.

		T	Sig. (2-tailed)
Pair1	P4AM - P4AMFM	1.333	.193
Pair 2	P8AM - P8AMFM	.077	.939
Pair 3	P12AM-P12AMFM	2.912	.007 *

From table 5 and 6 it is evident that 4AM + 4(AM+FM) $t(29) = 1.333$, $p>0.05$ hence there is no significant difference between the given two conditions i.e., envelope cues only and envelope and fine structure cues (4AM, 4AMFM condition).

8AM + 8(AM+FM) $t(29) = 0.077$, $p>0.05$ hence there is no significant difference between the two given conditions (8 AM and 8AMFM).

12AM + 12(AM+FM) t (29) = 2.912, $p<0.01$, hence there is significant difference between the two given conditions (12AM and 12AMFM).

The statistical analysis shows that there is a significant improvement in the 12 AM and 12 AM+FM condition, this can be attributed to the increased temporal and spectral information provided in this condition.

The results of this study are in accordance with the following studies:

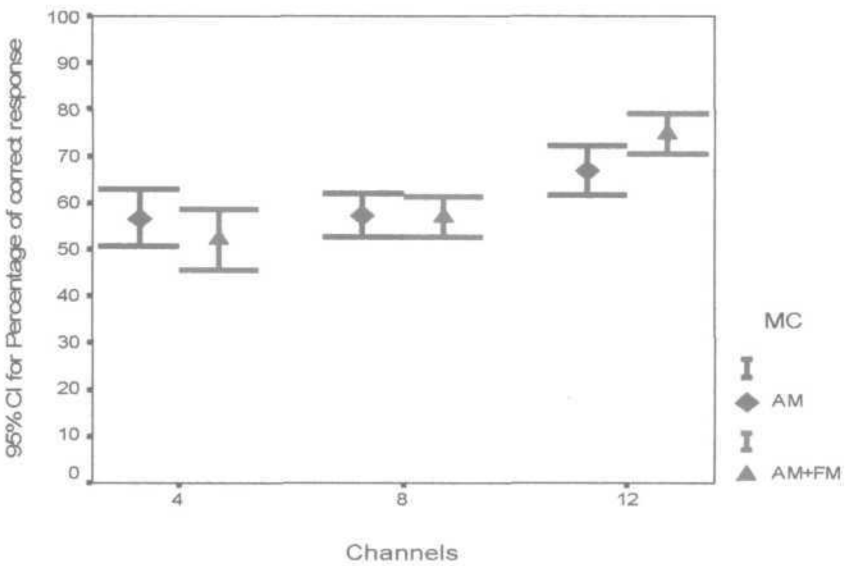
Nie, Stickney & Zeng (2005) demonstrated that fine structure (FM) cues are crucial for improving cochlear implant performance under realistic conditions and also helps in music perception in normal hearing subjects. The subjects obtained large improvement in sentence recognition scores in the presence of noise. Speech perception in noise and rhythm perception are significantly correlated (Leal et al, 2003). Hence, the improvements in the scores in this study are attributed to the better temporal and spectral representation of the signal.

Lan, Nie, Gao & Zeng (2004) demonstrated that the novel speech processing strategy with dynamic modulations of both frequency and amplitude is encouraging as it resulted in significant improvement in perception of Chinese tones, phrases and sentences. Experimental results reveal that tonal information in Chinese speech was encoded primarily in the fine details in the spectrum of the speech signal. It is inferred from this study that if the perception of tonal languages improves with the novel processing strategy, then rhythm perception which depends on the overall loudness variations of the signal will also be perceived better.

Stickney, Nie & Zeng (2005) demonstrated a significant improvement in the speech perception scores when fine structure (FM) cues were added to the envelope (AM) information especially in realistic listening environments. The addition of FM cues would provide information about pitch and formant transitions which will help in better speech and music perception, in rum rhythm perception.

Zeng, Nie, Liu, Stickney et al (2004) demonstrated that envelope cues are important for speech perception whereas the fine structure cues are critical for pitch perception which in turn enhances rhythm perception.

Figure 3: comparison of percentage correct scores across number of channels



To summarize, from the figure 3 it is evident that subjects showed improved performance with fine structure cues (AM+FM) and with increased number of channels. This increase in the scores can be attributed to the increased spectral and temporal resolution provided with 12 channels and the additional fine structure cues.

CHAPTER 5

SUMMARY AND CONCLUSIONS

The present study aimed at investigating the effect of adding fine structure cues in rhythm perception. All the speech coding strategies used with cochlear implants extract the envelope of the signal and code the information while missing out information on the fine structure cues. In this study, the fine structure cues were provided along with the envelope and the rhythm perception enhancement was studied. A group of 30 individuals with hearing within normal limits served as subjects for the study. The stimulus consisted of thirty rhymes which were modified by the MATLAB 6.5 software into envelope only (AM) and envelope + fine structure (AM+FM) conditions with 4, 8, 12 channels respectively.

The investigation was carried out to address the following research goals:

1. To determine effect of number of channels on rhythm perception with envelope cues (AM).
2. To determine the effect of number of channels on rhythm perception with envelope and fine structure cues (AM+FM).
3. To compare rhythm perception in the AM+FM condition and AM condition across different channels.

The stimulus was presented at 40 dB SL in the free field condition and the subjects were asked to tap the rhythm of the rhymes heard in six conditions

1. Amplitude modulated stimulus with 4 channels,
2. Amplitude modulated stimulus with 8 channels,
3. Amplitude modulated stimulus with 12 channels,
4. Amplitude modulated and Frequency modulated stimulus with 4 channels,
5. Amplitude modulated and Frequency modulated stimulus with 8 channels,
6. Amplitude modulated and Frequency modulated stimulus with 12 channels.

All the subjects were presented all the thirty rhymes randomly in the six conditions. The rhythm pattern tapped was scored on a sheet for all the thirty (single foot) rhymes. A model was created from the tap pattern of three speech language pathologists who were all well versed in music. The pattern tapped by the subjects was compared with the model developed based on the tap pattern of three speech language pathologists who were well versed in music. The tap pattern was scored. Each correct tap pattern was scored one and incorrect tap was scored zero.

The results obtained are given below:

- In the first condition, i.e., rhythm perception with only envelope cues (AM) with 4, 8, 12 channels, significant improvement was seen between 8 and 12 channels, however no significant difference was measured between 4 and 8 channels. Hence it can be concluded that this improvement in scores with 12 channels could be due to the increased spectral information provided. However, the absence of improvement between 4 and 8 channels could be due to the large standard deviation obtained in this condition

- In the second condition, i.e., rhythm perception with envelope and fine structure cues (AM+FM) with 4, 8, 12 channels, significant improvement was seen between 8 and 12 channels, however no significant difference was measured between 4 and 8 channels. This improvement in scores with 12 channels could be attributed to both the increased spectral information and temporal information provided by the fine structure cues. However, the absence of improvement between 4 and 8 channels could be due to the large standard deviation as mentioned.

- In the third condition, i.e., rhythm perception with envelope cues (AM) only and envelope + fine structure cues (AM+FM) with 4, 8, 12 channels, significant difference was seen between 12 AM and 12 AM+FM condition. However, no significant difference was seen between 4 AM and 4AM+FM condition, 8AM and 8 AM+FM condition. Thus it can be concluded that significant difference between AM and AM+FM condition is seen only in the optimal condition, i.e., the condition with the maximum spectral and temporal cues (12 AM and 12 AM+FM condition).

This study focuses on the importance of adding fine structure cues in the cochlear implant processing strategy to enhance the performance of cochlear implant users in both speech and music perception. Further studies need to be done to understand the effect of fine structure cues on melody and timbre perception, which will provide in depth information about the music perception. Studies should also be replicated in cochlear implant users to understand if similar results are obtained in them, as many other factors determine the performance of cochlear implant users unlike a simulated study on subjects with normal hearing.

BIBLIOGRAPHY

Brill, S., Gstottner, W., Helms, J., Iiberg, C. (1997). Optimization of channel number and stimulation rate for the fast continuous interleaved sampling strategy in the COMBI40+, *American Journal of Otolaryngology Supplement. 18,* S104-S106.

Dorman, M.F., Basham, K., McCandless, G., Dove, H. (1991). Speech understanding and music appreciation with the Ineraid cochlear implant, *The Hearing Journal. 44,* 34-37.

Fishman, K.A., Shannon, R.V., Slattery, W.H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor, *Journal of Speech and Hearing Research. 40,* 1150- 1163.

Fragoulis, D.K., Avaritsiotis, J.N., and Papaodysseus, C.N. (1999). Timbre Recognition of Single Notes Using an Artmap Neural Network. *National Technical University of Athens Dept of Electrical and Computer Engineering.*

Freisen, L., Shannon, R., Cruz, R.J. (2005). Effect of stimulation rate on speech recognition with cochlear implants, *Audiology Neurootology. 10,* 168-184.

Freisen, L., Shannon, R., Baskent, D., Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants, *Journal of Acoustical society of America. 110,* 1150-1163.

Geurts, L., Wouters, J. (2001). Coding of fundamental frequencies in continous interleaved sampling processors for cochlear implants, *Journal of Acoustical society of America. 109,* 713-726.

Gfeller, K., Lansing, C.R. (1991). Melodic, rhythmic and timbral perception of adult cochlear implant users, *Journal of Speech and Hearing Research. 34,* 916-920.

Gfeller, K., Woodworth, G. (1997). Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users, *Ear and Hearing. 18,* 252-260.

Green, T., Faulkner, A., Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants, *Journal of Acoustic Society of America. 116,* 2289-2297.

Kong. Y.Y., Cruz. R., Jones.A.J., Zeng. F.G. (2004). Music Perception with Temporal Cues in Acoustic and Electric Hearing, *Ear and Hearing. 25,* 173-186.

Lan, N., Nie, K., Zeng, F.G. (2004). Novel speech processing strategy incorporating tonal information for cochlear implants, *IEEE Trans. Biomed.Eng.51 (5),* 752-760.

Laneau, J., Wouters, J., Moonen, M. (2006). Improved music perception with explicit pitch coding in cochlear implants, *Audiology Neurootology. 11,* 38-52.

Leal, M.C., Shin, Y.J., Laborde, M.L., Calmela, N.M., Verges, S. (2003). Music perception in adult cochlear implant recipients, *Acta Otolaryngologica. 123,* 826-835.

McDermott, H. (2004). Music perception with cochlear implants - A review, *Trends in Amplification. 8(2),* 49-81.

McDermott, H. (2005). Perceiving melody remains challenging for implant wearers, *The Hearing Journal. 58(11),* 65- 74.

Mckay, CM., McDermott, H., Clark, G.M. (1994). Pitch percepts associated with amplitude modulated current pulse trains in cochlear implantees, *Journal of Acoustic Society of America. 96(5),* 2664- 2671.

Nie, K., Barco, A., Zeng, F.G. (2006). Spectral and temporal cues in cochlear implant speech perception, *Ear and Hearing. 27(2),* 208-217.

Nie, K., Stickney, G.S., Zeng, F.G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise, *IEEE Trans.Biomed.Enng.52(l),* 64-73.

Pijl. S. (1997). Labeling of musical interval size by cochlear implant subjects and normally hearing subjects, *Ear and Hearing. 18,* 364 - 372.

Pijl, S., Schwartz, D.W. (1995). Melody recognition and music interval perception by deaf subjects stimulated with electrical pulse rains through single cochlear implant electrodes, *Journal of Acoustic Society of America. 98,* 886 -895.

Rosen, S. (1992). Temporal information in speech: Acoustic, Auditory and Lingiustic aspects, *Philosophical Transactions: Biological Sciences. 336(1278),* 367-373.

Schulz. E., Kerber. M. Music perception with the MEDEL implants. Cited in *Advances in Cochlear implants.* 326-332. Vienna: Hochmair ES (eds).

Smith, Z.M., Delgutte, B., Oxenham, A.J.(2002). Chimaeric sounds reveal dichotomies in auditory perception, *Nature. 416,* 87-90.

Souza, S.E., Boike, K.T. (2006). Combining temporal-envelope cues across channels: effects of age and hearing loss, *Journal of speech, language and hearing research.* 49,138-149.

Stickney, G.S., Nie, K., Zeng, F.G. (2005). Contribution of frequency modulation to speech recognition in noise, *Journal of Acoustical society of America. 118(4),* 2412-2420.

Wilber, L.A. (1994). Calibration, pure tone, speech and noise signals. Cited in J. Katz (Eds.). *Handbook of Clinical Audiology (4th Edn.) 13-19.* Baltimore: Williams and Wilkins.

Zeng, F.G., Nie, K., Liu, S., Stickney, G., Del, R., Kong, Y.Y.,Chen, H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues, *Journal of Acoustical society of America. 116,* 1351- 1354.

Zeng, F.G. (2004). Trends in cochlear implants, *Trends in amplification. 8.*

Zeng, F.G., Nie, K., Stickney,G.S., Kong, Y.Y.(2004) Speech recognition with amplitude and frequency modulations, www.pnas.org cgi doi 10.1073 pnas.0406460102.

APPENDIX

A B C D E F G
H I J K L M N O P
L M N O P Q R S T
U V W X Y Z.

Baa Baa Black Sheep
Have u any wool?
Yes sir yes sir three bags full
One for my master, one for my dame,
One for the little boy who lives down the lane.

Bits of paper, bits of paper
Lying on the ground, lying on the ground
Make the place untidy make the place untidy
Pick them up, pick them up.

Chubby cheeks, dimple chin
Rosy lips, teeth within
Curly hair, very fair
Eyes are blue, lovely too
Teacher's pet is that you
Yes yes yes.

Ding dong bell
Pussys's in the well
Who put her in?
Little Tommy thin
Who pulled her out?
Little Tommy stout
What a naughty boy was that
To trample pussy cat.

Jack and Jill went up the hill
To fetch a pail of water
Jack fell down and broke his crown
And Jill came tumbling after.

Hickory Dickory dock
The mouse ran up the clock
The clock struck one
The mouse ran down
Hickory Dickory dock.

Hot cross buns, hot cross buns,
One a penny two a penny

If you have no daughters give it to your sons
One a penny two a penny hot cross buns.

Humpty dumpty sat on a wall
Humpty dumpty had a great fall
All the kings, horses and all the king's men
Could not put humpty together again.

I hear thunder, I hear thunder
Oh don't you, oh don't you
Pitter patter raindrops, pitter patter rain drops
I am wet through, I am wet through.

Johnny Johnny yes papa
Eating sugar no papa
Telling lies, no papa
Open your mouth ha ha ha.

Little Miss Muffet sat on a tuffet
Eating her curds away
There came a big spider who sat down beside her
And frightened Miss Muffet away.

London bridge is falling down
Falling down, falling down,
London Bridge is falling down my fair lady
Build it up with iron rods
Iron rods iron rods
Build it up with iron rods my fair lady.

Mary had a little lamp, little lamp, little lamp
Mary had a little lamp, its skin was bright as snow
Everywhere that Mary went, Mary went, Mary went
And everywhere that Mary went, the lamp was sure to go.

One two buckle my shoe
Two three knock the door
Five six pick up sticks
Seven eight lay them straight
Nine ten a big fat Hen.

Three blind mice, three blind mice
See how they run. See how they run
They all ran after the farmer's wife
Who cut off their tails with the carving knife?
Did you ever see such a thing in your life?
Three blind mice.

Twinkle twinkle little star

How I wonder what you are
Up above the world so high
Like a diamond in the sky.

Pussy cat pussy cat where have you been
I have been to London to visit the Queen
Pussy cat pussy cat what did u do there?
I frightened the little mouse under the chair.

Rain rain go away, come again another day
Rain rain go away come again another day
Little Tommy wants to play
Rain rain go to Spain, do not show your face again.

Row row row your boat
Gently down the stream
Merrily, merrily, merrily, merrily
Life is but a dream.

Bits of paper, bits of paper
Lying on the ground, lying on the ground
Make the place untidy make the place untidy
Pick them up, pick them up.

Hot cross buns, hot cross buns,
One a penny two a penny
If you have no daughters give it to your sons
One a penny two a penny hot cross buns.

I hear thunder, I hear thunder
Oh don't you, oh don't you
Pitter patter raindrops, pitter patter rain drops
I am wet through, I am wet through.