ACOUSTICAL AND PERCEPTUAL ANALYSIS OF SPEECH PRODUCED IN NOISE

Swathi S

Reg. No.: 12AUD031

A Dissertation Submitted in part fulfillment of final year

Master of Science (Audiology)

University of Mysore

ALL INDIA INSTITUTE OF SPEECH AND HEARING,

MANASAGANGOTHRI, MYSORE – 570 006

MAY, 2014

DEDICATED TO…

MY PARENTS & MY

GUIDE

Certificate

This is to certify that this dissertation entitled '**Acoustical and Perceptual Analysis of Speech Produced in Noise**' is a bonafide work submitted in part fulfillment for the degree of Master of Science (Audiology) of the student Registration No: 12AUD031. This has been carried out the under guidance of a faculty of this institute and has not been submitted earlier to any other university for the award of any diploma or degree.

Mysore
May, 2014

Dr. S. R. Savithri
DIRECTOR
All India Institute of Speech and Hearing
Manasagangothri, Mysore - 570 006

CERTIFICATE

This is to certify that this dissertation entitled '**Acoustical and Perceptual Analysis of Speech Produced in Noise**' has been prepared under my supervision and guidance. It is also certified that this has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Dr. Ajith Kumar U

Guide

Reader & Head of the Department,
Department of Audiology
All India Institute of Speech and Hearing
Manasagangothri, Mysore -570006

Mysore
May, 2014

Declaration

This is to certify that this master's dissertation entitled '**Acoustical and Perceptual Analysis of Speech Produced in Noise**' is the result of my own study under the guidance of Dr. Ajith Kumar U, Reader in Audiology, Department of Audiology, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier to any other university for the award of any diploma or degree.

Mysore

May, 2014 Register No: 12AUD031

ACKNOWLEDGEMENT

I would like to render my thanks to the Director of AIISH, ***Prof. S. R. Savithri,*** *for permitting me to carry out this study.*

Ajith Sir*…… You are one of the most amazing people I've ever come across... I genuinely believe your caliber is no less than a* ***Superman****…. Your grounded nature, broad horizon of thinking, calm and composed attitude, flexible style, superhuman intelligence and depth of knowledge is truly inspirational…. U were extremely patient with me, always ready to listen and help even in your busy schedule…. Also, for the first time in my life, I was exposed to Self- learning because of you and concluded it's the best form of learning….. Thanks for bearing with my mistakes and my super slow pace of working…….*

Amma and Appa*……. The unconditional love, trust and support from you is like oxygen to me, which keeps me alive every day, every moment of my life… I just want to say one line…. If there exists something called 'rebirth', I'm ready to take any number of rebirths just to be able to born as your daughter once again……… and again……… and again………………*

Usha Akka*…… U r like my elder sister who I always wanted to have… U were always there for me whenever I needed help or advice…. Ur intelligence, dedication and hard work fascinates me….. I can proudly say U were unofficially my second guide for this dissertation..;)*

Deepa Akka*…. U r one of the most caring and helping persons I've ever met… I can't thank U enough for all the help and support… U r a true motivator, taking challenges and viewing things in a complete positive stride (real sportsmanship…:P :P)…. Those badminton days with U r worth cherishing for a lifetime…..*

Nike sir*….. U r a true genius… Your quest to learn, clarity of concepts and your way of analyzing things just amazes me... Sorry for all the trouble I gave u…*

Mammu, Roji, Sharu, Sandu, Jyothi, Roja, Teju, Suhani*…… The absolute craziness that we shared….. those weekend trips, our DJ nights in hostel, horror movie sessions, the never ending philosophy discussions (wid u mammu), gossips, the* *silly teasings, nonsense pjs …… I don't think I'll ever get a chance to do these things in my life again….. Thanks for giving me some of the most happy and memorable moments of my life …..*

Padma**…….** *'Thank You' is a very small word for the amount of help you've given me to do this dissertation…. There was no word called 'NO' in your dictionary whenever I needed help….. This dissertation wouldn't be complete without acknowledging u….*

Sowmya**…….** *U r like a little sister to me….. U r a wonderful person and thanks loads for all the help and support…..*

Zeena**…..** *U r a very sweet girl….. U never complained even once, even though I troubled u so many times for recording…. Thanks so much…..*

Pooja**….** *Thanks so much for proof reading my dissertation with so much of patience and correcting even the minor errors….*

Sangeetha**…..** *Thanks a lot for ur help….. U r a very kind hearted girl….. keep it up….*

I want to thank all my participants for giving me your valuable time…..

I thank the God Almighty for giving me health and will power to give my best to this dissertation.

TABLE OF CONTENTS

LISTS OF TABLES

LISTS OF FIGURES

CHAPTER 1-INTRODUCTION

The environments in modern society consist of different kinds of background noise. It is quite difficult to communicate while being surrounded by noise, especially when the level of noise is high. Speakers tend to automatically increase their vocal intensity in such conditions, which is known as the Lombard Effect. Lombard (1911) demonstrated increase in loudness of speaker's voice when exposed to background noise. A number of studies have demonstrated the same effect (Junqua, 1996).

However, the modifications occurring to speech produced in a noisy background is not restricted only to changes in vocal intensity. A number of studies have reported that speech produced in noise demonstrates acoustic-phonetic modifications in speech such as increase in fundamental frequency (Fo), word duration (or a decrease of speaking rate) and first formant frequency (F1), as well as a shift of spectral energy to higher frequencies (Hanley & Steer, 1949; Korn, 1954; Dreher & O'Neill, 1957; Webster & Klumpp, 1962; Charlip & Burk, 1969; Stanton, Jamieson, & Allen, 1988; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; Bond, Moore, & Gabel, 1989; Howell, Young & Sackin, 1992; Junqua, 1993; Letowski, Frank, & Caravella, 1993; Steeneken & Hansen, 1999; Pittman & Wiley, 2001; Garnier, Bailly, Dohen, Welby, & Loevenbruck, 2006; Varadarajan & Hansen, 2006; Garnier, 2007; Mixdorff, Pech, Davis, & Kim, 2007; Boril, 2008; Patel & Schell, 2008). In addition, changes in consonant-to-vowel energy ratio has also been reported. Junqua (1993) and Womack and Hansen (1996) reported a shift of energy from consonant to vowel for speech produced in noise with respect to quiet. Changes in formant frequencies have been reported, with the consensus that F1 tends to increase

(Summers et al., 1988; Lu & Cooke, 2008) while F2 has been reported to increase (Junqua, 1993) or decrease (Pisoni, Bernacki, Nusbaum, & Yuchtman, 1985).

These acoustic-phonetic modifications are influenced by the type of environment or type of noise. Recent findings show that the Lombard effect is sensitive to frequencies vital for speech and is not a general response to any competing sound in the environment (Stowe & Golob, 2013). In the presence of speech-shaped noise, flattening of spectral tilt and increase in Fo has been found (Lu & Cooke, 2009b). Changes in the speech level, Fo, F1, and spectral center of gravity differed when speech was produced with low and high pass filtered noise backgrounds (Lu & Cooke, 2009a). The effect of noise on speech production increased with the number of background talkers, which increases the energetic masking effect of the noise (Lu & Cooke, 2008).

There is convincing evidence that speech produced in noise has acoustic-phonetic modifications other than mere increase in the loudness. Furthermore, these modifications are shown to be dependent on spectral and temporal characteristics of noise. However, the perceptual benefits of these acoustic modifications of speech produced in noise are not clear. It is interesting to see whether the speech produced in noisy background would be more intelligible when compared to speech produced in quiet due to the acoustic-phonetic modifications that are induced. Studies have shown that for isolated words or continuous speech, speech produced in noise is more intelligible than speech produced in quiet (Dreher & O'Neill, 1957; Summers et al., 1988; Pittman & Wiley, 2001). The improvement in speech intelligibility is attributed to the changes in the spectral and temporal properties of speech which accompany the Lombard effect.

Need for the Study

Previous research on speech produced in noise has mainly concentrated on loudness effects. Though a few studies have looked into the modification of other acoustic features of speech produced in noise, these studies have used only broad band stationary noise. Moreover, these studies have not compared modifications in speech productions in the presence of noises with different spectral and temporal characteristics. Furthermore, perceptual benefits of these speech production modifications are not clear. Therefore, the present study was taken up with the aim to evaluate the speech production modifications and perceptual benefits, if any, of the speech produced in noise.

Aim of the Study

To measure the speech production modifications and perceptual benefits of speech produced in high pass, low pass filtered white noise and temporally modulated noise.

Objectives of the Study

1) To compare the mean fundamental frequency, first three formant frequencies, duration and mean intensity of the speech produced in quiet, high pass and low pass filtered white noise and temporally modulated noise at syllable level.

2) To measure the speech intelligibility of speech produced in quiet, high pass and low pass filtered white noise, and temporally modulated noise in the presence of same and different maskers.

CHAPTER 2-REVIEW OF LITERATURE

Studies on speech production modifications in background noise began with Lombard, who originally demonstrated the effect of background noise on speaker's loudness, popularly known as the 'Lombard Effect' (Lombard, 1911). Older studies mainly concentrated upon the intensity effect and thought that Lombard effect is a very general phenomenon where, upon the presentation of noise, overall vocal effort will increase, resulting in increased intensity. However, recent studies have shown that such effect is not only on the intensity but also on other spectral and temporal parameters.

Acoustic-Phonetic Modifications of Speech Produced in Noise

Table 2.1 summarizes the various studies that have been carried out on acoustic-phonetic modifications of speech produced in noise.

Table 2.1.

Studies on speech production modifications in noise

Study (Author, Year)	Method	Acoustic- Phonetic Modifications
1. Dreher and O'Neill, 1957	Words and sentences spoken in background of white noise	Increased intensity and duration
2. Summers et al., 1988	Production of 15 words in the presence of white noise	Increase in rms amplitude, duration of utterances, Fo, a decrease in spectral tilt and increase in F1
3. Junqua, 1993	Recording of monosyllables and words produced in background of white noise	Increase in duration, energy, pitch, F1 and spectral center of gravity
4. Tartter, Gomes, and Litwin, 1993	Production of 14 words in different levels of white noise	Decreased spectral tilt, increased duration, F2 and average speech amplitude. Effects increased with increase in level
5. Pittman and Wiley, 2001	Production of sentences in quiet, wide band noise and meaningful multi-talker babble	Increase in vocal output and increased spectral level at high frequencies
6. Lu and Cooke, 2008	Speech production in quiet and N- talker babble	Significant change in energy and Fo, which increased with increase in N
7. Lu and Cooke, 2009a	Compared read speech in low pass and high pass noise	Spectral parameters did not shift to noise- free regions for speech produced in HPN, unlike LPN
9. Stowe and Golob, 2013	Compared speech produced in broadband noise and notched noise	Increase in duration, intensity, Fo in broadband noise background and no effect of notched noise

Note. Fo = Fundamental frequency; F1 = first formant frequency, F2= second formant frequency; HPN= High pass noise; LPN= Low pass noise.

Summary of each study.

1. Dreher and O'Neill (1957) analyzed the effect of various levels of the back ground noise on the production of words and sentences. They reported that when intensity of the background noise was increased (quiet, 70, 80, 90 and 100 dB SPL), the duration and intensity of words and sentences increased.
2. Summers et al., (1988) conducted acoustic analysis of speech produced in increasing levels of background noise consisting of white noise low pass filtered at 3.5 kHz at 80, 90 and 100 dB SPL for 15 words and found that the mean rms energy of speech increased with the level of noise for every word. There was a consistent increase in word duration with increase in noise level. Fundamental frequency (Fo) was found to be significantly different between noise and quiet conditions. Spectral tilt towards the higher frequencies was also found.
3. Junqua (1993) found that the Lombard effect is highly variable with respect to speaker and significantly different for male and female speakers. The consonant to vowel ratio (CVR) was found to be reduced in noise induced speech.
4. Tartter, Gomes, and Litwin (1993) reported spectral shift and increase in duration of Lombard speech. The tilt was observed towards the high frequencies with noise above 35 dB SPL. One speaker steadily increased the fundamental frequency (Fo) while the other speaker decreased it. In loud noise, there were significant changes in first and second formant frequencies for some words, but not others.

5. Pittman and Wiley (2001) recorded 50 sentences with embedded target word in three conditions- quiet, wide band noise and meaningful multi-talker babble at 80 dB SPL. On an average, vocal levels for wide band noise and meaningful multi-talker babble increased by 14.5 dB SPL. Spectral level increased in high frequencies.

6. Lu and Cooke (2008) recorded sentences in different N- talker speech babble, varying in number of background speakers, at 89 dB SPL. The energetic masking effect was increased (increase in energy, mean Fo) with increase in noise level and as the number of background talkers increased and reached a ceiling at N = 8.

7. Lu and Cooke (2009a) hypothesized that speakers actively shift their spectral energy distribution to regions least affected by noise. To test this, they measured the speech level, Fo, first formant frequency (F1), spectral center of gravity (spectral CoG) for speech produced in the presence of low and high-pass filtered noise at 89 dB SPL. However, they found little evidence for the hypothesis since parameters such as F0 and F1 frequencies, and spectral CoG did not shift downwards but instead increased relative to speaking in quiet conditions, in the presence of high-pass noise.

8. Stowe and Golob (2013) hypothesized that Lombard effect is frequency specific and tested it in two noise conditions- broadband noise and notched noise. Results showed that broadband noise significantly increased intensity, duration, and F0 of speech while notched noise, had no effect.

From the above studies, it is evident that there is clear effect of background noise on spectral and temporal characteristics of speech. However, there is no clear consensus regarding whether this effect is dependent on the spectral and temporal characteristics of background noise. This is because very few studies have used noises with distinctly different spectral and temporal characteristics. Therefore, in the present study noises with different spectral characteristics (high and low pass) and temporal characteristic (temporally modulated speech) was used.

Intelligibility of Noise Induced Speech

Table 2.2 summarizes the various studies that have been carried out on perceptual benefits of speech produced in noise over speech produced in quiet.

Table 2.2.

Studies on perceptual effects of speech produced in noise

Study (Author, Year)	Condition/ Type of noise	Results
1. Dreher and O'Neill, 1957	Compared intelligibility for words and sentences spoken in different background levels of noise	More intelligibility of stimuli produced in noise over quiet
2. Summers et al., 1988	Intelligibility of speech produced in broad band noise at different signal-to-noise ratios	Speech produced in noise had a significant perceptual advantage over quiet
3.Pittman and Wiley, 2001	Speech production in backgrounds of broadband noise and meaningful multi-talker babble. Perception tested with multi-talker babble	Recognition of the speech produced in broadband noise and multi-talker babble was 15% higher than that for the speech produced in quiet
4. Lu and Cooke, 2008	Compared intelligibility of speech produced in quiet and N-talker babble when mixed with N- talker babble	Noise- speech more intelligible than quiet and intelligibility gain increased with increase in N
5. Lu and Cooke, 2009b	Studied the relative effect of flattening of spectral tilt and change in Fo on intelligibility	Flattening of spectral tilt contributed to increase in intelligibility while change in Fo did not
6. MacDonald and Raufer, 2013	Studied the intelligibility of speech produced in broadband noise that was amplitude modulated at different rates, ranging from 1 to 16 Hz	SRTs improved with increase in modulation frequency of noise

Note. Fo = Fundamental frequency; SRT = Speech recognition thresholds

Summary of each study.

1. Dreher and O'Neill, 1957 found that speech produced in 70 dB SPL of white noise gave an intelligibility benefit over quiet for words and sentences, at a constant signal to noise ratio (SNR).

2. Pittman and Wiley (2001) investigated effects of speech produced in backgrounds of broadband noise (WBN) and meaningful multi-talker babble (MTB) on intelligibility. Perception was tested with MTB. Recognition of the speech produced in WBN and MMB was, on an average 15% higher than that for the speech produced in quiet. They conducted two experiments, first with preserved vocal intensity of noise induced speech, and second with equating the vocal intensity to that produced in quiet. The intelligibility gain obtained was 69% and 15% respectively for two experiments. It was also found that the acoustic properties of recorded speech measured did not correlate with the recognition measures.

3. In the study conducted by Summers et al., (1988), speech produced in 90 dB of masking noise was consistently identified more accurately than speech produced in quiet, regardless of talker and signal-to-noise ratios. Furthermore, for speech produced in 100 dB of masking noise, the effect of masking noise on intelligibility increased as signal-to-noise ratios decreased, supporting the hypothesis that acoustic-phonetic modifications of speech produced in noise are dynamic and had reliable advantages over quiet production.

4. Lu and Cooke (2008) compared intelligibility of speech produced in N-talker babble in the presence of quiet and N-talker babble. Speech produced in noise was more intelligible than speech produced in quiet and the intelligibility gain increased with the level of noise and number of background talkers.

5. Lu and Cooke (2009b) studied the relative effect of flattening of spectral tilt and change in F0 on intelligibility, by manipulating the recorded speech.

Flattening of spectral tilt contributed to increase in intelligibility while change in F0 did not.

6. MacDonald and Raufer (2013) investigated the effect of speech produced in broadband noise that was amplitude modulated at different rates, ranging from 1 to 16 Hz. It was found that speech reception thresholds improved with increase in modulation frequency of noise.

From the above studies it is clear that the speech produced in noise is more intelligible compared to the speech produced in quiet. However, it is not clear whether this perceptual benefit is general or specific to noise condition in which they were produced. Therefore, the present study was taken up with the aim to evaluate the benefit of speech produced with different spectral and temporal characteristics when mixed with same or different noise.

CHAPTER 3-METHOD

The method of the study was divided into two phases.

Phase I: Speech Production Modifications of Speech Produced in Noise

This phase involved the acoustic analysis of speech produced in the presence of noises with different spectral and temporal characteristics.

Participants. The speakers consisted of 10 female participants in the age range of 18-25 years (mean age- 21.5 years). All the participants were native speakers of Kannada language, with pure tone thresholds of less than 15 dB HL at octave frequencies between 250 Hz and 8000 Hz. Their speech recognition thresholds were less than 15 dB HL (for bisyllabic words in Kannada used in the Department of Audiology, All India Institute of Speech and Hearing) and speech identification scores were more than 90% (for phonetically balanced word list in Kannada given by Yathiraj & Vijayalakshmi, 2005). All the participants had normal middle ear functioning as reflected by type-A tympanogram and presence of ipsilateral and contralateral acoustic reflexes at normal sensation levels. They did not have any history or presence of gross speech, language, psychological or neurological problems as revealed by a structured interview. Participants were explained about the purpose of the study and prior oral consent was taken before their participation in the study.

Stimuli. Speech material consisted of different monosyllables and sentences. Monosyllables and sentences were produced in quiet and while listening to high pass, low pass and temporally modulated noise at 80 dB SPL. Intensity of the noise was kept at 80 dB SPL as previous studies have shown that Lombard effect is more at high intensity (Dreher and O'Neill, 1957; Summers et al., 1988). Seventeen monosyllables (/k/, /g/, /ʧ/, /ʤ/, /t/, /d/, /ʈ/, /ɖ/, /p/, /b/, /m/, /n/, /j/, /r/, /l/, /s/ and /ʃ/) were produced in combination with vowel /a/ in quiet and while listening to high pass, low pass and temporally modulated noise. Also, one randomly selected speaker from the participants produced sentences in quiet and while listening to high pass, low pass and temporally modulated noise. Sentences were selected from a corpus of Kannada sentences developed by Geetha, Kumar and Pavan (2011). The first 16 lists (10 sentences in each list) were recorded. Monosyllables were used for both acoustic and perceptual analysis while sentences were used only for perceptual analysis. Low pass noise was filtered from white noise and had a cut off frequency of 1000 Hz with a slope of 60 dB/ octave. High pass noise was filtered from white noise and had a cut off frequency of 1000 Hz with a slope of 60 dB/ octave. Finite impulse response filter was used for this purpose. Temporally modulated noise consisted of white noise sinusoidally amplitude modulated at a modulation frequency of 8 Hz with modulation depth of 100%.

Test environment. The recording was done in an acoustically treated room with noise levels within permissible limits (ANSI S3.1, 1991).

Instrumentation. Estimation of pure tone thresholds was done using a calibrated diagnostic audiometer (GSI-61). Tympanogram and acoustic reflexes were recorded using a calibrated middle ear analyzer (GSI Tympstar). A condenser microphone was used for the recording, along with an audio interface, MOTU Microbook II. Recording was done on a desktop computer using Adobe Audition 3 software. Masking noise was presented using Sennheiser HDA 200 headphones connected to MOTU Microbook II audio interface. Figure 3.1 illustrates instrumentation.

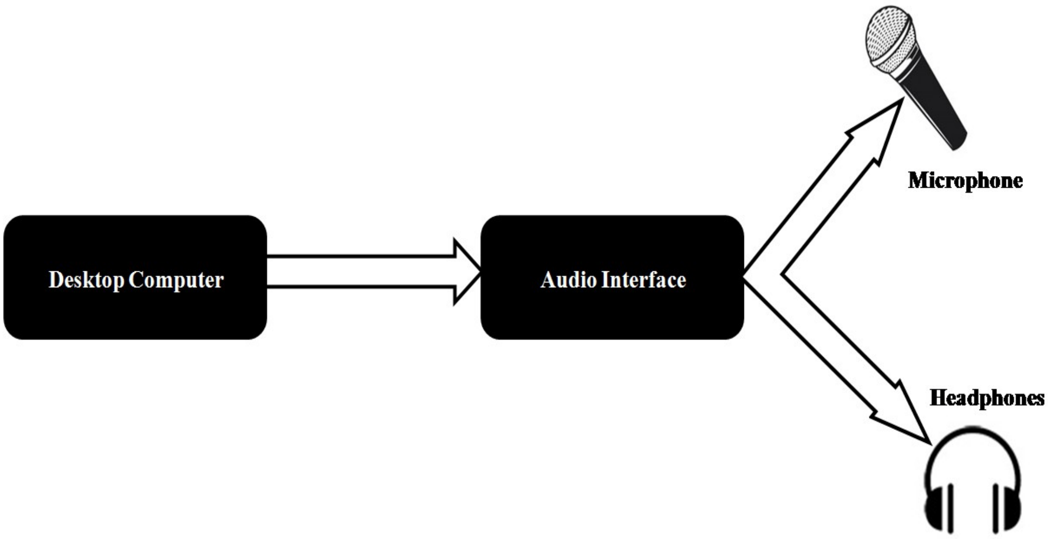


Figure 3.1. Block diagram of instrumentation for recording.

Procedure. Speakers were asked to produce the given monosyllables and sentences in quiet and while listening to 80 dB SPL low pass, high pass and temporally modulated noises. The participants were seated comfortably in a chair with appropriate head and neck support to avoid any unnecessary movement of head or neck during recording. The recording microphone was placed 2-3 inches from the mouth of the speaker with the help of a microphone stand.

Spoken sentences and monosyllables were recorded using Adobe Audition 3 software at a sampling frequency of 44100 Hz. The gain in MOTU Microbook II audio interface was set at +36 dB so that the speech would not overshoot the clipping level while recording. Recording conditions (quiet and three different noise conditions) were counter balanced across participants. The order of recording of monosyllables versus sentences was also randomized across participants. Practice trial was given before the actual recording. The speakers were asked to keep the headphones on, even in the recording for quiet condition.

For monosyllables, a list of 17 monosyllables was given to the participants to read aloud. Background noise was presented continuously through Windows Media Player. During recording of the sentences, the sentences were visually displayed in Kannada script on a computer screen using DMDX software (Forster & Forster, 2003). Simultaneously, noise was presented binaurally so that the participants read aloud the sentences simultaneously with noise in their ears. Noise was automatically turned off after the completion of each sentence and turned on again for the next sentence to avoid any effect of adaptation.

Acoustic analysis. Acoustic analysis was carried out only on monosyllables. The recorded monosyllables were analyzed using Praat software (version 5.3.75, Paul & David, 2014). Seventeen monosyllables produced by 10 speakers in four conditions (quiet and three noise conditions) were analyzed for the following parameters:-

- Fundamental frequency (Fo)

  The mean fundamental frequency for each speaker was measured by extracting the value of Fo from the software. In other words, pitch was extracted for each monosyllable.

- Analysis of formants

  The first three formant frequencies were measured by extracting the formants from the software. The steady state portion of formants from the vowel part of the spectra was considered.

- Duration of monosyllables

Duration of the monosyllable was calculated by measuring the total duration from the onset to the offset of the monosyllable.

- Mean intensity

The mean intensity for each monosyllable over the complete duration of the monosyllable was extracted.

Figure 3.2 depicts the screen shot of the analysis window in Praat software, showing Fo, the formants, duration and the mean intensity for the monosyllable /ʤa/, which was produced by one of the participants while listening to low pass noise.

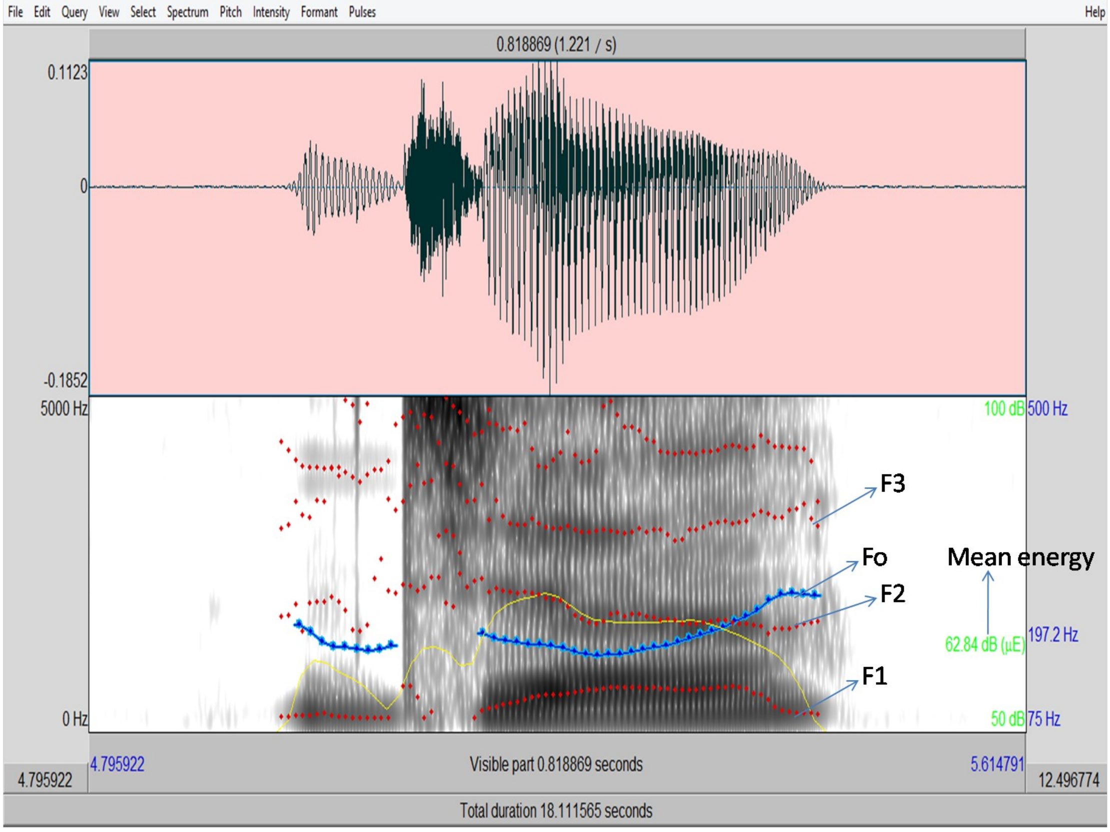


Figure 3.2. Screen shot of the analysis window of /ʤa/ produced in low pass masking noise. Fo = Fundamental frequency; F1 = first formant frequency; F2 = second formant frequency; F3= third formant frequency.

Phase II: Perceptual Analysis of Speech Produced in Noise

Participants. The listeners consisted of 10 female participants in the age range of 18-25 years (mean age- 22 years). Participants who had participated in Phase I did not participate in Phase II. Other selection criteria were same as that mentioned in Phase I.

Stimuli. Monosyllables and sentences produced while listening to different background noises were used as the stimuli. Monosyllables recorded from one of the speakers in Phase I, who also produced the sentences was used for perceptual analysis. The sentences and monosyllables were processed and edited by using Adobe Audition 3 as follows:

- The leading and the trailing silent intervals for each sentence and monosyllable were removed.
- Sentences and monosyllables were separately normalized to rms level.
- The processed sentences and monosyllables were mixed with three different types of noise used in Phase I at -5 dB signal to noise ratio (SNR), using a custom written MATLAB code (Gnanateja & Pavan, 2013).

Thus, sentences and monosyllables were mixed with the noise which the talker heard while producing it and remaining other two types of noise. The control conditions consisted of original recordings (produced in four conditions) without mixing with any type of noise. This led to a total of 16 conditions. The 16 sentence lists were used for 16 conditions, respectively (one list per condition). All 17 monosyllables were tested in all 16 conditions. Table 3.1 explains the 16 different conditions.

Table 3.1.

Sixteen conditions taken for perceptual analysis

Production Condition	Perception Condition			
	Quiet	Low pass noise	High pass noise	Temporally modulated noise
Quiet	QQ	QL	QH	QT
Low pass noise	LQ	LL	LH	LT
High pass noise	HQ	HL	HH	HT
Temporally modulated noise	TQ	TL	TH	TT

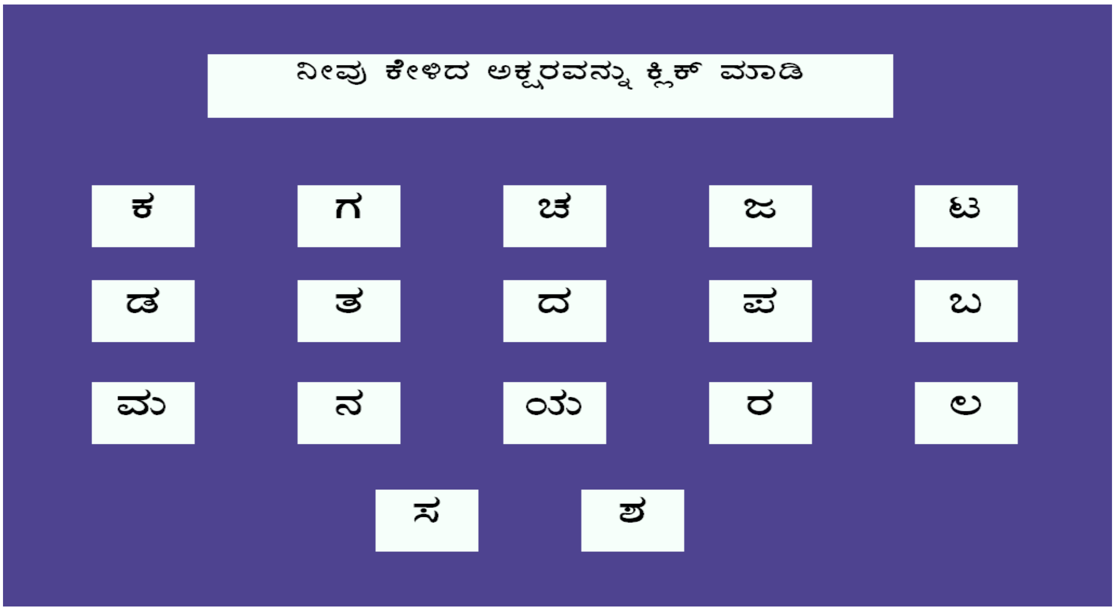
Note. Q=quiet; L=low pass noise; H= high pass noise; T= temporally modulated noise. First and second letters in the abbreviations correspond to production and perception conditions, respectively. For example, LH corresponds to the condition where the speech produced with low pass noise as background, was mixed with high pass noise for perception testing.

Test environment. The testing was done in an acoustically treated room with noise levels within permissible limits (ANSI S3.1, 1991).

Instrumentation. Presentation of stimuli was done using Laptop computer-Sony Vaio F14212. Stimuli was presented through Sennheiser communication Headset PC 320 G4ME headphones. Paradigm software (Paradigm v2.2.0.197) was used for presentation of stimuli and also for recording of responses for monosyllables.

Procedure. Stimuli were presented diotically using Sennheiser communication Headset PC 320 G4ME headphones at comfortable level of listeners (75-85 dB SPL). Identification of the sentences was done in open set while monosyllables were done in closed set using Paradigm v2.2.0.197 software through laptop computer. Order of identification task with different noise conditions was counter balanced. The order of presentation of monosyllables versus sentences was also randomized across participants.

Monosyllables. Participants were asked to click the correctly heard stimulus from a choice of 17 monosyllables displayed in a rectangular block as shown in Figure 3.3. Complete randomization was done across different monosyllables and conditions. The participants' responses were automatically saved by the computer program. Each monosyllable was presented five times in each condition which resulted in a total of 1360 stimuli per participant.



ನೀವು ಕೇಳಿದ ಅಕ್ಷರವನ್ನು ಕ್ಲಿಕ್ ಮಾಡಿ

Figure 3.3. Screen-shot of monosyllable perception testing. 17 monosyllables are represented in Kannada alphabets, for the participants to click the heard stimulus.

Sentences. Participants were asked to repeat the sentence heard verbatim, and the responses were manually recorded in a scoring sheet. Complete randomization was done across 16 sentence lists during presentation.

Analysis. ***Monosyllables.*** Confusion matrices were drawn for each condition using a custom written MATLAB code (Gnanateja, 2014) for each participant. Later, consonants were classified with respect to their features- place, manner and voicing, according to the classification given by Schiffman (1979) and Shridhar (1990) as shown in Table 3.2.

Table 3.2.

Feature matrix for 17 monosyllables

	Monosyllables																
Feature	/k/	/g/	/ʧ/	/ʤ/	/t/	/d/	/ʈ/	/ɖ/	/p/	/b/	/m/	/n/	/j/	/r/	/l/	/s/	/ʃ/
Place	vel	vel	pal	pal	den	den	ret	ret	lab	lab	lab	alv	pal	alv	den	alv	pal
Manner	st	st	af	af	st	st	st	st	st	st	na	na	gl	li	li	fr	fr
Voicing	u	v	u	v	u	v	u	v	u	v	v	v	v	v	v	u	u

Note. v= voiced; u= unvoiced; st= stop; af= affricate; na= nasal; fr= fricative; li= liquid; gl= glide; vel= velar; pal= palatal; den= dental; ret= retroflex; lab= labial; alv= alveolar.

Sequential information analysis was done for all three features, for 12 conditions (SINFA, Wang & Bilger, 1973). Transmitted information was calculated using FIX software (Feature Information Xfer) given by Department of Phonetics & Linguistics, University College London.

Sentences. Analysis of sentences for each participant was done by calculating total word correct scores for each condition.

CHAPTER 4-RESULTS

Results are reported separately for Phase I (Production analysis) and Phase II (Perception analysis).

Production Analysis Results

The results of acoustic analysis of monosyllables are analyzed and reported with respect to six parameters. They are duration of the monosyllable (in seconds), mean intensity (in dB SPL), fundamental frequency (Fo in Hz), first formant (F1 in Hz), second formant (F2 in Hz), and third formant (F3 in Hz) frequencies. These parameters are analyzed for four conditions in which speech was produced - quiet, low pass noise (LPN), high pass noise (HPN) and temporally modulated noise (TMN).

Duration. The mean and standard deviations for duration of the monosyllables are given in Figure 4.1.

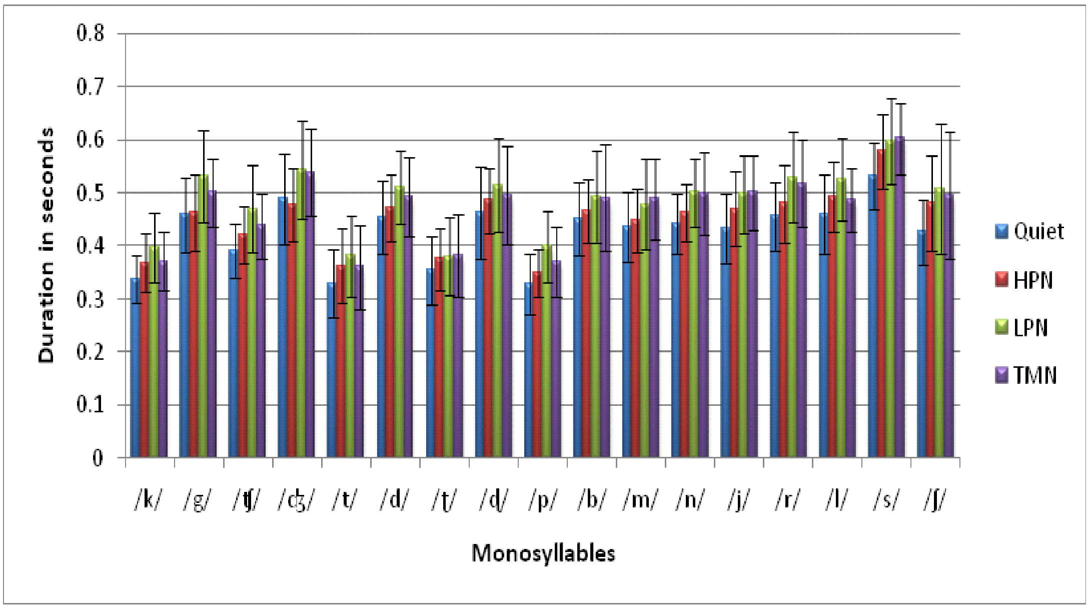


Figure 4.1. Duration of monosyllables across conditions.

From Figure 4.1, it is clear that monosyllables produced in noise conditions had longer durations compared to quiet condition. Among noise conditions, monosyllables produced in LPN condition had maximum duration, followed by TMN, HPN.

Two-way repeated measures ANOVA was done with production condition and monosyllable as within subject factors. Results showed that there was significant main effect of production condition, F [3, 27] = 14.370, $p < 0.001$ and significant main effect of monosyllable, F [16, 144] = 56.863, $p < 0.001$. There was no significant interaction effect between production condition and monosyllable, F [48, 432] = 1.035, $p > 0.05$. Pairwise comparisons were done using Bonferroni's adjusted multiple pairwise comparisons and the results are given in Table 4.1.

Table 4.1.

Pairwise comparisons for duration

	HPN	LPN	TMN
Quiet	NS	**	*
HPN		*	NS
LPN			NS

Note. * denotes $p < 0.05$; ** denotes $p < 0.01$; NS denotes no significance, $p > 0.05$.

Intensity. The mean and standard deviations for intensity of the monosyllables are given in Figure 4.2.

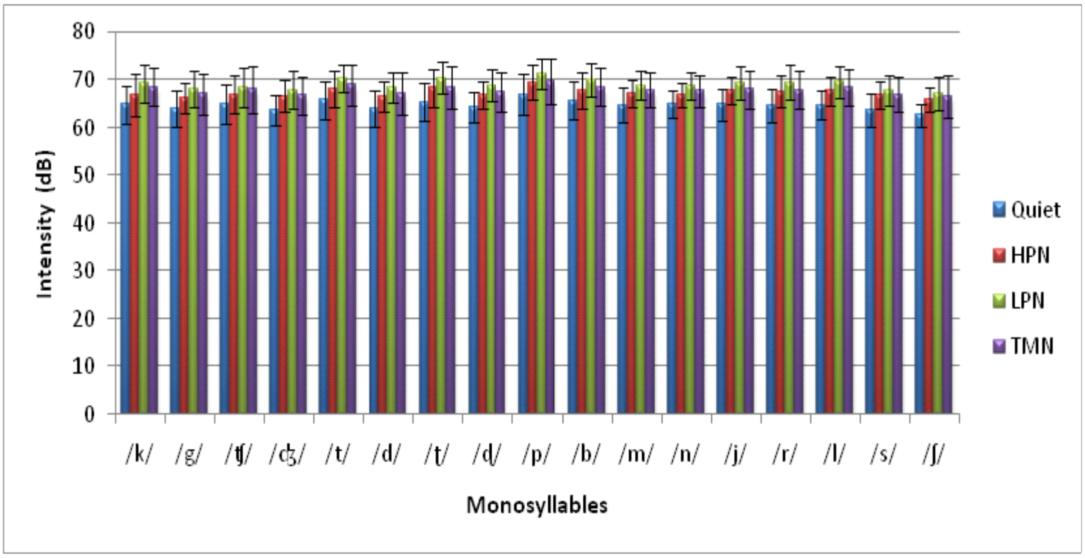


Figure 4.2. Intensity of monosyllables across conditions.

From Figure 4.2, it is clear that monosyllables produced in noise condition had higher intensity as compared to quiet condition. Among the noise conditions, monosyllables produced in LPN condition had maximum intensity. Monosyllables produced in HPN and TMN conditions had almost similar intensity.

Two-way repeated measures ANOVA was done with production condition and monosyllable as within subject factors. Results showed that there was significant main effect of production condition, *F* [3, 27] = 12.521, $p < 0.001$ and significant main effect of monosyllable, *F* [16, 144] = 9.953, $p < 0.001$. There was no significant interaction effect between production condition and monosyllable, *F* [48, 432] = 0.848, $p > 0.05$. Pairwise comparisons were done for production conditions using Bonferroni's adjusted multiple pairwise comparisons and the results are given in Table 4.2.

Table 4.2.

Pairwise comparisons for intensity

	HPN	LPN	TMN
Quiet	NS	**	*
HPN		*	NS
LPN			NS

Note. * denotes $p < 0.05$; ** denotes $p < 0.01$; NS denotes no significance, $p > 0.05$.

Fundamental frequency. The mean and standard deviations for Fo of the monosyllables are given in Figure 4.3.

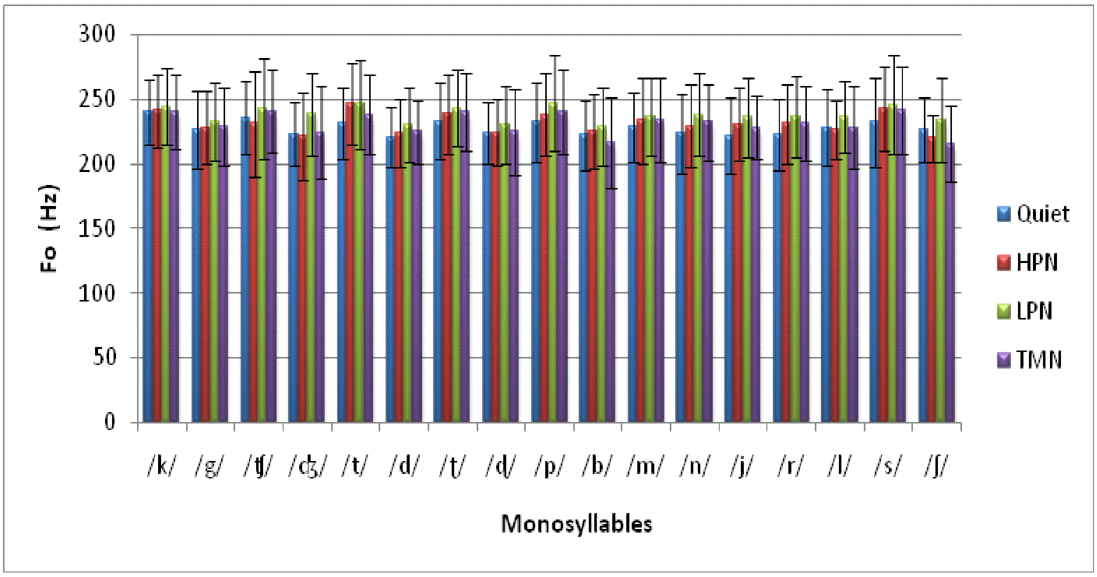


Figure 4.3. Fo of monosyllables across conditions.

From Figure 4.3, it is clear that monosyllables produced in quiet had minimum Fo. Monosyllables produced in LPN condition had maximum Fo. Monosyllables produced in HPN and TMN conditions had almost similar Fo.

Two-way repeated measures ANOVA was done with production condition and monosyllable as within subject factors. Results showed that there was significant main effect of production condition, *F* [3, 27] = 6.709, $p < 0.01$ and significant main effect of monosyllable, *F* [16, 144] = 7.761, $p < 0.001$. There was significant interaction effect between production condition and monosyllable, *F* [48, 432] = 1.500, $p < 0.05$. Pairwise comparisons were done using Bonferroni's adjusted multiple pairwise comparisons and the results are given in Table 4.3.

Table 4.3.

Pairwise comparisons for Fo

	HPN	LPN	TMN
Quiet	NS	NS	NS
HPN		*	NS
LPN			**

Note. * denotes $p < 0.05$, ** denotes $p < 0.01$, NS denotes no significance, $p > 0.05$.

First formant frequency. The mean and standard deviations for F1 of the monosyllables are given in Figure 4.4.

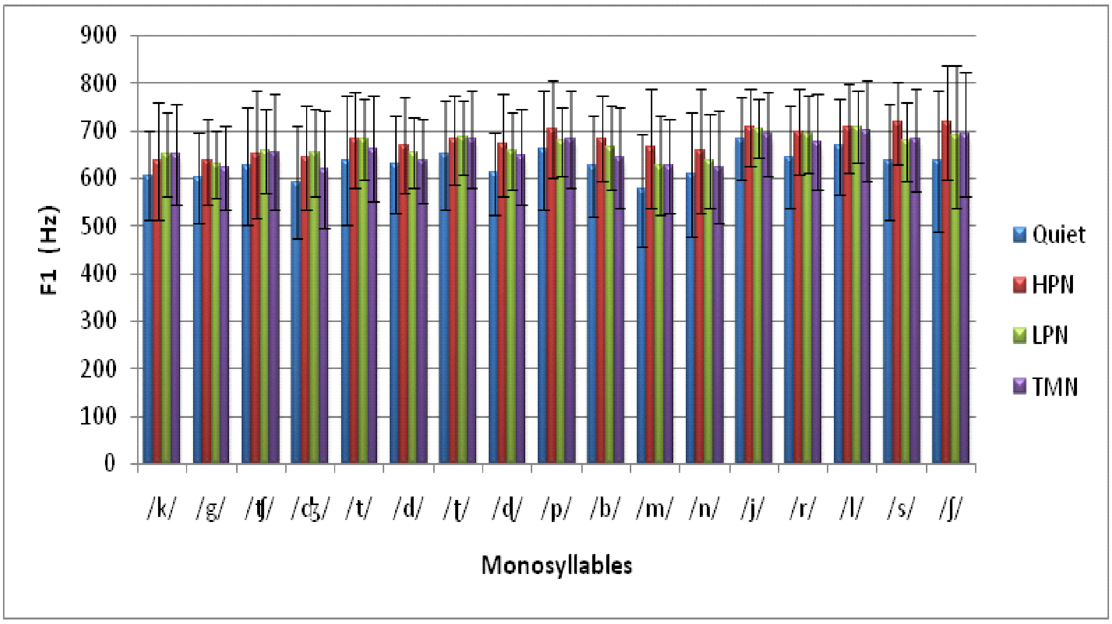


Figure 4.4. F1 of monosyllables across conditions.

From Figure 4.4, it is clear that monosyllables produced in quiet condition had minimum F1. Monosyllables produced in HPN condition had maximum F1, followed by LPN, TMN.

Two-way repeated measures ANOVA was done with production condition and monosyllable as within subject factors. Results showed that there was significant main effect of production condition, *F* [3, 27] = 5.446, *p* < 0.01 and significant main effect of monosyllable, *F* [16, 144] = 4.217, *p* < 0.001. There was no significant interaction effect between production condition and monosyllable, *F* [48, 432] = 0.962, *p* > 0.05. Pairwise comparisons were done using Bonferroni's adjusted multiple pairwise comparisons and the results showed no significant difference across any condition.

Second formant frequency. The mean and standard deviations for F2 of the monosyllables are given in Figure 4.5.

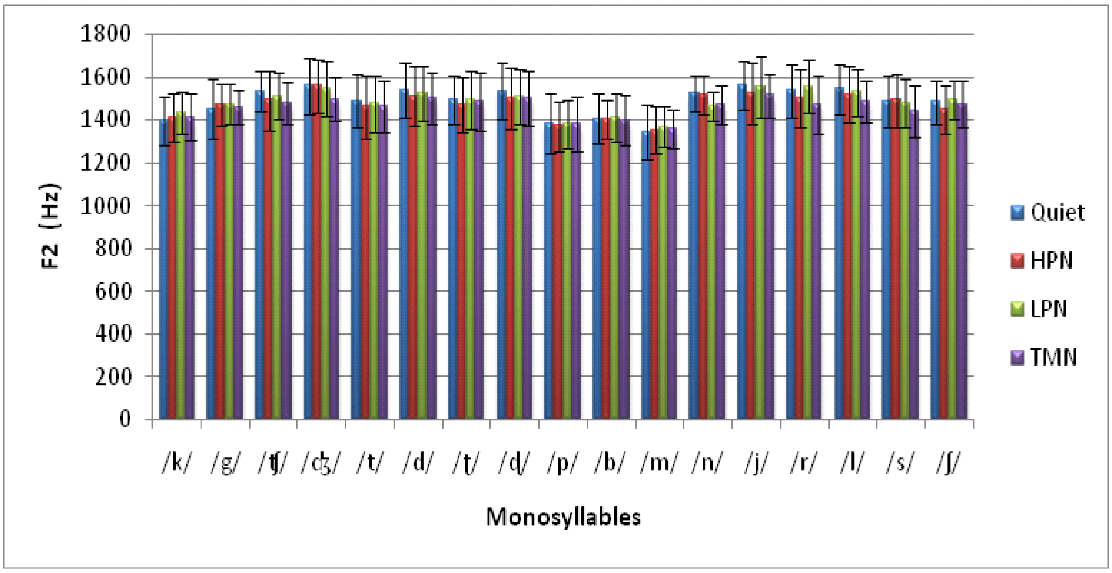


Figure 4.5. F2 of monosyllables across conditions.

Two-way repeated measures ANOVA was done with production condition and monosyllable as within subject factors. Results showed that there was no significant main effect of production condition, *F* [3, 27] = 1.581, *p* > 0.05. There was significant main effect of monosyllable, *F* [16, 144] = 15.763, *p* < 0.001. There was significant interaction effect between production condition and monosyllable, *F* [48, 432] = 1.525, *p* < 0.05.

Third formant frequency. The mean and standard deviations for F3 of the monosyllables are given in Figure 4.6.

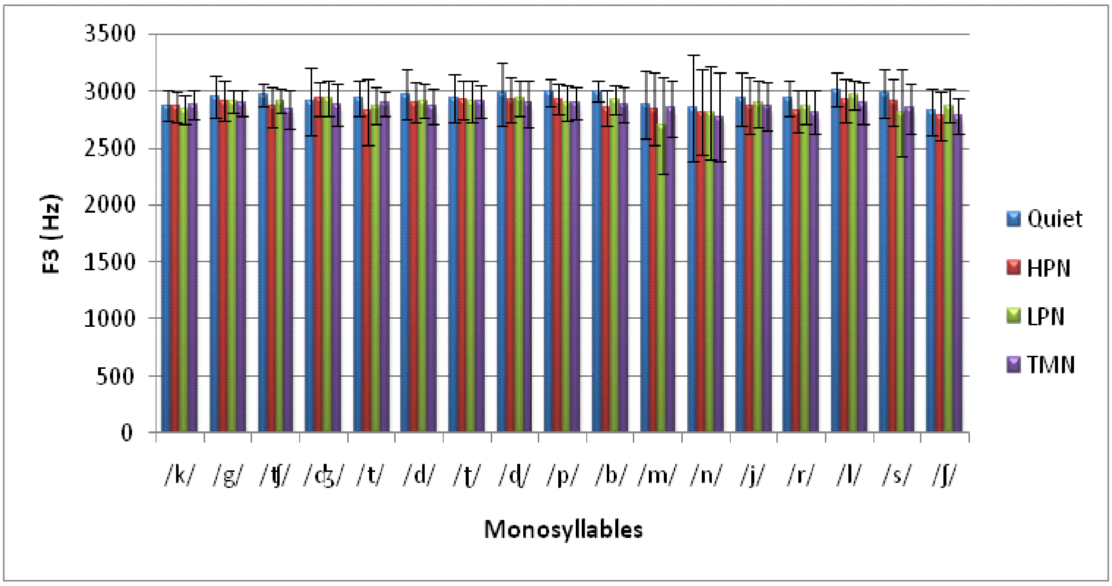


Figure 4.6. F3 of monosyllables across conditions.

Two-way repeated measures ANOVA was done with production condition and monosyllable as within subject factors. Results showed that there was significant main effect of production condition, *F* [3, 27] = 5.503, $p < 0.01$. There was no significant main effect of monosyllable, *F* [16, 144] = 0.865, $p > 0.05$. There was significant interaction effect between production condition and monosyllable, *F* [48, 432] = 1.442, $p < 0.05$. Pairwise comparisons were done using Bonferroni's adjusted multiple pairwise comparisons and the results showed no significant difference across any condition.

Perceptual Analysis Results

Results were reported separately for monosyllables and sentences.

In control conditions, (i.e. perceived in quiet condition) the identification scores of all the participants were approaching 100% and were at the ceiling. Hence they were not further analyzed. These results reiterate the fact that in quiet condition individuals with normal hearing can perceive the speech near to perfection effortlessly. The remaining 12 conditions were subjected to various statistical analysis.

These 12 conditions were abbreviated for the purpose of simplicity and understanding, according to the following rule- the first letter in the abbreviation corresponds to the condition in which the stimulus was produced and second letter corresponds to the condition in which it was heard/ perceived. Table 4.4 shows abbreviations and their meaning.

Table 4.4.

Abbreviations of different conditions analyzed

Abbreviated Condition	Expansion
LH	Speech produced in low pass noise and perceived in high pass noise
LT	Speech produced in low pass noise and perceived in temporally modulated noise
LL	Speech produced in low pass noise and perceived in low pass noise
HL	Speech produced in high pass noise and perceived in low pass noise
HT	Speech produced in high pass noise and perceived in temporally modulated noise
HH	Speech produced in high pass noise and perceived in high pass noise
TL	Speech produced in temporally modulated noise and perceived in low pass noise
TH	Speech produced in temporally modulated noise and perceived in high pass noise
TT	Speech produced in temporally modulated noise and perceived in temporally modulated noise
QL	Speech produced in quiet and perceived in low pass noise
QH	Speech produced in quiet and perceived in high pass noise
QT	Speech produced in quiet and perceived in temporally modulated noise

Monosyllables. Monosyllables were analyzed by formulating the confusion matrices and calculating the information transmitted for the features of voicing, manner and place. Analysis was carried out separately for all 12 conditions listed in Table 4.4.

Speech produced in low pass noise and perceived in high pass noise (LH).

Table 4.5 shows the confusion matrix for the LH condition. In all the confusion matrices presented here, the stimuli are represented in columns and the participants responses are represented in rows. The principal diagonal of the matrix gives the number of correct responses.

Table 4.5.

Confusion matrix for LH condition

	/b/	**/ʧ/**	**/ɖ/**	**/d/**	**/ʤ/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/t/**	**/ʃ/**	**/ʈ/**	**/j/**	**/s/**
/b/	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	42	0	0	0	0	0	0	0	0	0	0	1	8	0	0	24
/ɖ/	0	0	49	40	0	0	0	0	0	0	0	0	0	0	0	0	0
/d/	0	0	1	10	0	1	0	0	0	0	0	0	0	0	0	0	0
/ʤ/	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0
/g/	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	0	0
/k/	0	1	0	0	0	0	46	0	0	0	0	0	0	0	0	0	3
/l/	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0
/p/	0	0	0	0	0	0	4	0	0	0	50	0	0	0	0	0	0
/r/	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/t/	0	0	0	0	0	0	0	0	0	0	0	0	7	0	1	0	1
/ʃ/	0	3	0	0	0	0	0	0	0	0	0	0	0	42	0	0	8
/ʈ/	0	0	0	0	0	0	0	0	0	0	0	0	42	0	49	0	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0
/s/	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14

In this condition, the maximum confusion occurred between /t/ and /ʈ/ for 42 times, also between /d/ and /ɖ/ for 40 times. /s/ was perceived as /ʧ/ for 24 times and as /ʃ/ for 8 times. /ʃ/ was perceived as /ʧ/ for 8 times.

The proportion of information transmitted in bits with respect to input was calculated by sequential information analysis (SINFA), which revealed that in LH condition, voicing feature was maximally transmitted (1.000), followed by manner (0.898) and place (0.855). Proportion value 1 indicates there were no errors in voicing. The total information transmitted was 3.605 bits.

Speech produced in low pass noise and perceived in temporally modulated noise (LT).

Table 4.6

Confusion matrix for LT condition

	/b/	**/ʧ/**	**/ɖ/**	**/d/**	**/ʤ/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/ʃ/**	**/t/**	**/ʈ/**	**/j/**
/b/	50	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ɖ/	0	0	45	40	0	0	0	0	0	0	0	0	0	0	0	0	0
/d/	0	0	4	10	0	1	0	0	0	0	0	0	0	0	0	0	0
/ʤ/	0	1	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0
/g/	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	0	0
/k/	0	0	0	0	0	0	47	0	0	0	0	0	0	0	0	0	0
/l/	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	50	0	0	1	0	0	0	0
/p/	0	0	0	0	0	0	3	0	0	0	50	0	0	0	0	0	0
/r/	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/s/	0	1	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0
/ʃ/	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0
/t/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	1	0
/ʈ/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	42	49	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

Table 4.6 shows the confusion matrix for the LT condition. In this condition, the maximum confusion occurred between /d/ and /ɖ/ for 44 times, also between /t/ and /ʈ/ for 42 times.

SINFA revealed that in LT condition, manner feature was maximally transmitted (0.991), followed by voicing (0.986) and place (0.875). The total information transmitted was 3.803 bits.

Speech produced in low pass noise and perceived in low pass noise (LL).

Table 4.7.

Confusion matrix for LL condition

	/b/	**/ʤ/**	**/ʧ/**	**/ɖ/**	**/d/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/ʃ/**	**/ʈ/**	**/t/**	**/j/**
/b/	49	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
/ʤ/	0	50	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ɖ/	0	0	0	50	0	0	0	0	0	0	0	0	0	0	26	0	0
/d/	0	0	0	0	45	0	0	0	0	0	0	0	0	0	0	0	0
/g/	1	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0
/k/	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0
/l/	0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	1	0	50	0	0	0	0	0	0	0
/p/	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	0	0
/r/	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/s/	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0
/ʃ/	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0
/ʈ/	0	0	0	0	2	0	0	0	0	0	0	0	0	0	24	0	0
/t/	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	50	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

Table 4.7 shows the confusion matrix for the LL condition. In this condition, the maximum confusion occurred between /ʈ/ and /ɖ/ for 26 times. /d/ was perceived as /t/ for 3 times and /ʧ/ was perceived as /ʤ/ for 2 times.

SINFA revealed that in LL condition, manner feature was maximally transmitted (0.996), followed by place (0.980) and voicing (0.821). The total information transmitted was 3.934 bits. From these results, it is clear that LL condition had less errors compared to LH and LT conditions.

Speech produced in high pass noise and perceived in low pass noise (HL).

Table 4.8.

Confusion matrix for HL condition

	/b/	**/ʧ/**	**/ɖ/**	**/d/**	**/ʤ/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/ʃ/**	**/ʈ/**	**/t/**	**/j/**
/b/	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ɖ/	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/d/	0	0	0	39	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʤ/	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0
/g/	0	0	0	0	0	50	0	0	0	0	2	0	0	0	0	0	0
/k/	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0
/l/	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0
/p/	2	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	0
/r/	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/s/	0	0	0	0	0	0	0	0	0	0	0	0	50	1	0	0	0
/ʃ/	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0
/ʈ/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	2	0
/t/	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	48	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

Table 4.8 shows the confusion matrix for the HL condition. In this condition, the maximum confusion occurred between /d/ and /t/ for 11 times. No other prominent errors were noticed.

SINFA revealed that in HL condition, manner information was maximally transmitted (1.000), followed by place (0.968) and voicing (0.898). The total information transmitted was 3.988 bits.

Speech produced in high pass noise and perceived in temporally modulated noise (HT).

Table 4.9

Confusion matrix for HT condition

	/b/	**/ʧ/**	**/d/**	**/ɖ/**	**/ʤ/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/ʃ/**	**/t/**	**/ʈ/**	**/j/**
/b/	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/d/	0	0	15	17	0	0	0	0	0	0	0	0	0	0	0	0	0
/ɖ/	0	0	33	30	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʤ/	0	0	0	3	36	3	0	0	0	0	0	0	0	0	0	0	0
/g/	0	0	2	0	0	47	0	0	0	0	0	0	0	0	0	0	0
/k/	0	0	0	0	0	0	40	0	0	0	17	0	0	0	0	0	0
/l/	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0
/p/	0	0	0	0	0	0	3	0	0	0	33	0	0	0	1	0	0
/r/	0	0	0	0	2	0	0	0	0	0	0	50	0	0	0	0	0
/s/	0	1	0	0	0	0	4	0	0	0	0	0	50	1	0	0	0
/ʃ/	0	4	0	0	0	0	1	0	0	0	0	0	0	49	1	0	0
/t/	0	0	0	0	0	0	2	0	0	0	0	0	0	0	27	4	0
/ʈ/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	46	0
/j/	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	50

Table 4.9 shows the confusion matrix for the HT condition. In this condition, the maximum confusion occurred between /d/ and /ɖ/ for 50 times. /t/ was perceived as /ʈ/ for 21 times. /p/ was perceived as /k/ for 17 times. /ʤ/ was perceived as /j/ for 12 times.

The SINFA revealed that in HT condition, voicing feature was maximally transmitted (1.000), followed by manner (0.902) and place (0.754). The total information transmitted was 3.590 bits.

Speech produced in high pass noise and perceived in high pass noise (HH).

Table 4.10.

Confusion matrix for HH condition

	/b/	**/t/**	**/ʧ/**	**/ɖ/**	**/d/**	**/j/**	**/ʤ/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/ʃ/**	**/ʈ/**	**/s/**
/b/	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/t/	0	16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	17
/ʧ/	0	0	47	0	0	0	0	0	0	0	0	0	0	0	15	0	9
/ɖ/	0	0	0	50	43	0	0	0	0	0	0	0	0	0	0	0	0
/d/	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0
/j/	0	0	0	0	0	50	42	0	0	0	0	0	0	0	0	0	0
/ʤ/	0	0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0
/g/	0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0
/k/	0	0	0	0	0	0	0	0	44	0	0	0	0	0	0	0	2
/l/	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/p/	0	0	0	0	0	0	0	0	1	0	0	0	50	0	0	0	2
/r/	0	0	0	0	0	0	3	0	0	0	0	0	0	50	0	0	0
/ʃ/	0	0	2	0	0	0	0	0	1	0	0	0	0	0	27	0	2
/ʈ/	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0
/s/	0	0	0	0	0	0	0	0	4	0	0	0	0	0	8	0	18

Table 4.10 shows the confusion matrix for the HH condition. In this condition, the maximum confusion occurred between /d/ and /ɖ/ for 43 times, also between /t/ and /ʈ/ for 34 times. Also, /ʤ/ was perceived as /j/ for 42 times. /s/ was perceived as /t/ for 17 times and as /ʧ/ for 9 times. /ʃ/ was perceived as /ʧ/ for 15 times and as /s/ for 8 times.

SINFA revealed that in HH condition, voicing feature was maximally transmitted (1.000), followed by place (0.836) and manner (0.765). The total information transmitted was 3.484 bits. From these results, it is clear that HH condition had more errors compared to HL and HT conditions.

Speech produced in temporally modulated noise and perceived in low pass noise (TL).

Table 4.11.

Confusion matrix for TL condition

	/p/	**/b/**	**/ʧ/**	**/ɖ/**	**/d/**	**/ʤ/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/r/**	**/s/**	**/ʃ/**	**/ʈ/**	**/t/**	**/j/**
/p/	49	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/b/	1	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	0	50	0	0	2	0	0	0	0	0	0	0	0	0	0	0
/ɖ/	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0
/d/	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0
/ʤ/	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	0	0
/g/	0	0	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0
/k/	0	0	0	0	0	0	7	50	0	0	0	0	0	0	0	0	0
/l/	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0
/r/	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/s/	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0
/ʃ/	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0
/ʈ/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	4	0
/t/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

Table 4.11 shows the confusion matrix for the TL condition. In this condition, the maximum confusion occurred between /b/ and /p/ for 17 times. /g/ was perceived as /k/ for 7 times.

SINFA revealed that in TL condition, manner feature was maximally transmitted (1.000), followed by place (0.978) and voicing (0.806). The total information transmitted was 3.942 bits.

Speech produced in temporally modulated noise and perceived in high pass noise (TH).

Table 4.12.

Confusion matrix for TH condition

	/b/	**/ʧ/**	**/ɖ/**	**/d/**	**/ʤ/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/t/**	**/ʃ/**	**/ʈ/**	**/j/**	**/s/**
/b/	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	47	0	0	0	0	0	0	0	0	0	0	0	18	0	0	10
/ɖ/	0	0	50	17	0	0	0	0	0	0	0	0	0	0	0	0	0
/d/	0	0	0	33	0	2	0	0	0	0	0	0	0	0	0	0	0
/ʤ/	0	0	0	0	50	5	0	0	0	0	0	0	0	0	0	0	0
/g/	2	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0
/k/	0	0	0	0	0	0	50	0	0	0	2	0	0	0	0	0	18
/l/	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0
/p/	0	0	0	0	0	0	0	0	0	0	48	0	5	0	0	0	0
/r/	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/t/	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	4
/ʃ/	0	3	0	0	0	0	0	0	0	0	0	0	0	29	0	0	2
/ʈ/	0	0	0	0	0	0	0	0	0	0	0	0	32	0	50	0	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0
/s/	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	16

Table 4.12 shows the confusion matrix for the TH condition. In this condition, the maximum confusion occurred between /t/ and /ʈ/ for 32 times, also between /d/ and /ɖ/ for 17 times. /s/ was perceived as /ʧ/ for 10 times and as /k/ for 18 times. /ʃ/ was perceived as /ʧ/ for 18 times.

SINFA revealed that in TH condition, voicing feature was maximally transmitted (1.000), followed by manner (0.869) and place (0.818). The total information transmitted was 3.590 bits.

Speech produced in temporally modulated noise and perceived in temporally modulated noise (TT).

Table 4.13

Confusion matrix for TT condition

	/b/	**/ʧ/**	**/ɖ/**	**/d/**	**/ʤ/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/ʃ/**	**/ʈ/**	**/j/**	**/t/**
/b/	46	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	46	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
/ɖ/	0	0	49	29	0	0	0	0	0	0	0	0	0	0	0	0	0
/d/	1	0	1	17	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʤ/	0	0	0	0	49	10	0	0	0	0	0	0	0	0	0	0	0
/g/	2	0	0	4	0	40	0	0	0	0	0	0	0	0	0	0	0
/k/	0	0	0	0	0	0	50	0	0	0	7	0	0	0	0	0	0
/l/	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0
/p/	0	0	0	0	0	0	0	0	0	0	43	0	0	0	0	0	4
/r/	1	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/s/	0	4	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0
/ʃ/	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0	0
/ʈ/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	34
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0
/t/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12

Table 4.13 shows the confusion matrix for the TT condition. In this condition, the maximum confusion occurred between /t/ and /ʈ/ for 34 times, also between /d/ and /ɖ/ for 29 times. /g/ was perceived as /ʤ/ for 10 times. /p/ was perceived as /k/ for 7 times.

SINFA revealed that in TT condition, voicing feature was maximally transmitted (1.000), followed by manner (0.928) and place (0.823). The total information transmitted was 3.690 bits. From these results, it is clear that TT condition had less errors compared to TH condition but more errors compared to TL condition.

Speech produced in quiet and perceived in low pass noise (QL).

Table 4.14.

Confusion matrix for QL condition

	/b/	**/ʤ/**	**/ʧ/**	**/ɖ/**	**/d/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/ʃ/**	**/ʈ/**	**/t/**	**/j/**
/b/	49	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
/ʤ/	0	50	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	0	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ɖ/	0	0	0	49	0	0	0	0	0	0	0	0	0	0	0	0	0
/d/	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0
/g/	0	0	0	0	0	50	1	0	0	0	0	0	0	0	0	0	0
/k/	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	0
/l/	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	1	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	49	0	0	0	0	0	0	0
/p/	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0
/r/	1	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/s/	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0
/ʃ/	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0
/ʈ/	0	0	0	1	0	0	0	0	0	0	0	0	0	0	50	0	0
/t/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

Table 4.14 shows the confusion matrix for the QL condition. In this condition, the maximum confusion occurred between /ʧ/ and /ʤ/ for 8 times. /p/ was perceived as /b/ for 3 times.

SINFA revealed that in QL condition, manner feature was maximally transmitted (0.995), followed by place (0.994) and voicing (0.875). The total information transmitted was 3.996 bits.

Speech produced in quiet and perceived in high pass noise (QH).

Table 4.15.

Confusion matrix for QH condition

	/b/	**/k/**	**/d/**	**/ɖ/**	**/ʧ/**	**/ʤ/**	**/g/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/ʈ/**	**/t/**	**/s/**	**/ʃ/**	**/j/**
/b/	49	0	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0
/k/	0	49	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0
/d/	0	0	27	1	0	1	0	0	0	0	0	0	0	0	0	0	0
/ɖ/	0	0	17	49	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	0	0	0	2	1	0	0	0	0	0	0	0	0	6	11	0
/ʤ/	0	0	0	0	0	38	1	0	0	0	0	0	0	0	0	0	0
/g/	0	0	3	0	0	10	48	0	0	0	0	0	0	0	0	0	0
/l/	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	2	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0
/p/	0	1	0	0	0	0	0	0	0	0	50	0	0	0	0	9	0
/r/	1	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/ʈ/	0	0	0	0	0	0	0	0	0	0	0	0	49	17	1	1	0
/t/	0	0	0	0	2	0	0	0	0	0	0	0	1	33	4	1	0
/s/	0	0	0	0	27	0	0	0	0	0	0	0	0	0	39	9	0
/ʃ/	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	19	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

Table 4.15 shows the confusion matrix for the QH condition. In this condition, the maximum confusion occurred between /ʧ/ and /s/ for 33 times. /ʧ/ was perceived as /k/ for 17 times. /t/ was perceived as /ʈ/ for 17 times and /d/ was perceived as /ɖ/ for 17 times. /ʤ/ was perceived as /g/ for 10 times. /ʃ/ was perceived as /ʧ/ for 11 times, as /s/ for 9 times and as /p/ for 9 times.

SINFA revealed that in QH condition, voicing feature was maximally transmitted (0.987), followed by manner (0.833) and place (0.792). The total information transmitted was 3.468 bits.

Speech produced in quiet and perceived in temporally modulated noise (QT).

Table 4.16

Confusion matrix for QT condition

	/b/	**/ʧ/**	**/d/**	**/ɖ/**	**/ʤ/**	**/g/**	**/k/**	**/l/**	**/m/**	**/n/**	**/p/**	**/r/**	**/s/**	**/ʃ/**	**/t/**	**/ʈ/**	**/j/**
/b/	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʧ/	0	42	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
/d/	0	0	31	1	0	0	0	0	0	0	0	0	0	0	0	0	0
/ɖ/	0	0	14	49	0	0	0	0	0	0	0	0	0	0	0	0	0
/ʤ/	0	1	4	0	50	0	0	0	0	0	0	0	0	0	0	0	0
/g/	1	0	1	0	0	50	0	0	0	0	0	0	0	0	0	0	0
/k/	0	3	0	0	0	0	45	0	0	0	0	0	0	0	0	0	0
/l/	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0
/m/	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0
/n/	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0
/p/	0	0	0	0	0	0	5	0	0	0	50	0	0	0	0	0	0
/r/	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
/s/	0	2	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0
/ʃ/	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0
/t/	0	2	0	0	0	0	0	0	0	0	0	0	0	0	50	3	0
/ʈ/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0
/j/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

Table 4.16 shows the confusion matrix for the QT condition. In this condition, the maximum confusion occurred between /d/ and /ɖ/ for 14 times. /k/ was perceived as /p/ for 5 times.

SINFA revealed that in QT condition, voicing feature was maximally transmitted (0.988), followed by manner (0.954) and place (0.902). The total information transmitted was 3.870 bits. From these results, it is clear that maximum transmission occurred in QL condition, followed by QT and then QH condition.

Summary of monosyllable perception. The results are summarized with respect to the features transmitted. The total information transmitted is also given.

Place. Figure 4.7 shows the summary of place information transmitted in different conditions. From Figure 4.7 it can be inferred that place feature was transmitted maximally when monosyllables were perceived in LPN irrespective of the conditions in which they were produced. However, when comparisons were made between monosyllables perceived in high pass and temporally modulated noise, there was marginal advantage when the production and perception noise condition was same.

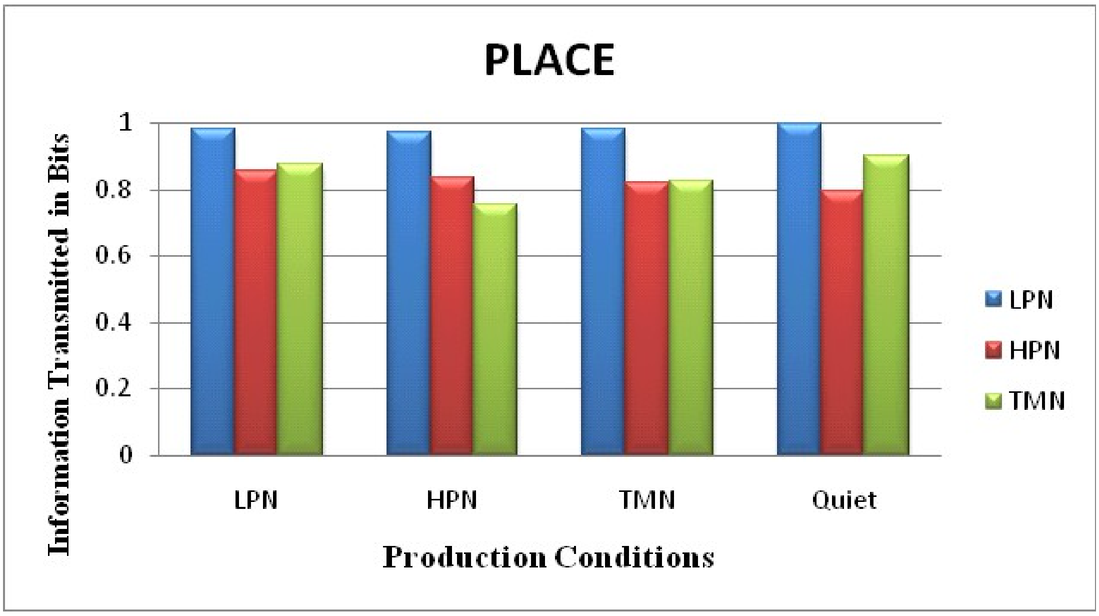


Figure 4.7. Information transmitted with respect to place feature across different conditions. Different perception conditions are indicated by different colors.

Manner. Figure 4.8 shows the summary of manner information transmitted in different conditions. From the Figure 4.8 it is clear that maximum manner information was transmitted when participants perceived speech in LPN condition.

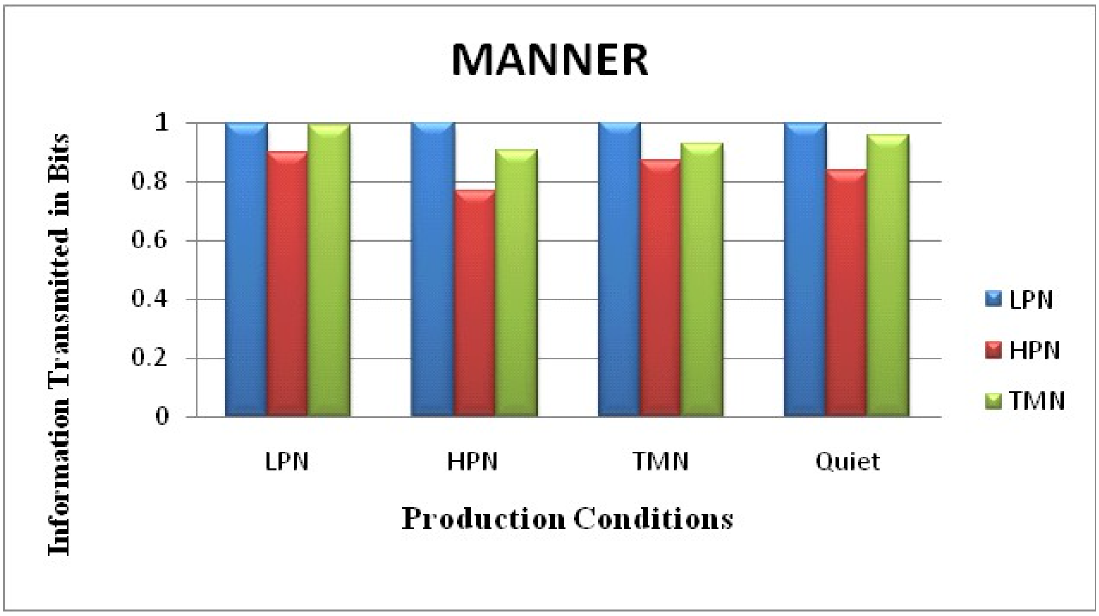


Figure 4.8. Information transmitted with respect to manner feature across different conditions. Different perception conditions are indicated by different colors.

Voicing. Figure 4.9 shows the summary of voicing information transmitted in different conditions. From Figure 4.9 it is clear that voicing feature was transmitted maximally for both HPN and TMN perception conditions (maximum score of 1). Voicing was perceived poorly in LPN perception condition, irrespective of the production condition (Figure 4.9).

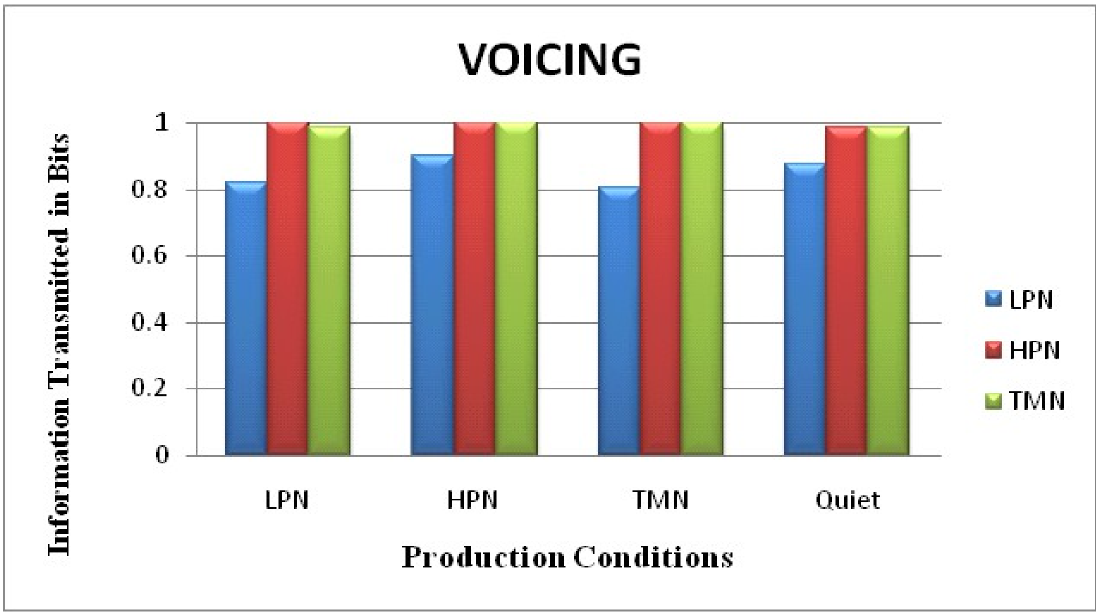


Figure 4.9. Information transmitted with respect to voicing feature across different conditions. Different perception conditions are indicated by different colors.

Total transmitted information. Figure 4.10 shows the summary of total information transmitted in different conditions. From Figure 4.10 it is clear that the total information was maximally transmitted when participants perceived the speech in LPN irrespective of the production conditions, followed by TMN and HPN conditions, respectively.

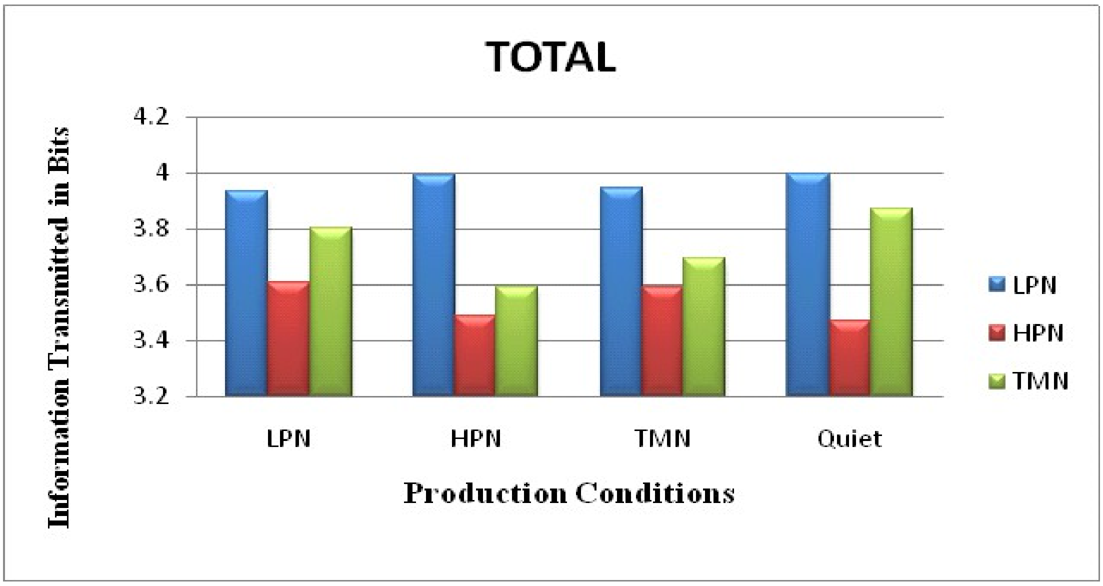


Figure 4.10. Total information transmitted across different conditions. Different perception conditions are indicated by different colors.

Sentences. Percent word correct scores was obtained for each condition averaged across 10 subjects. Figure 4.11 shows the descriptive statistics (mean and standard deviation) of sentences for 12 conditions.

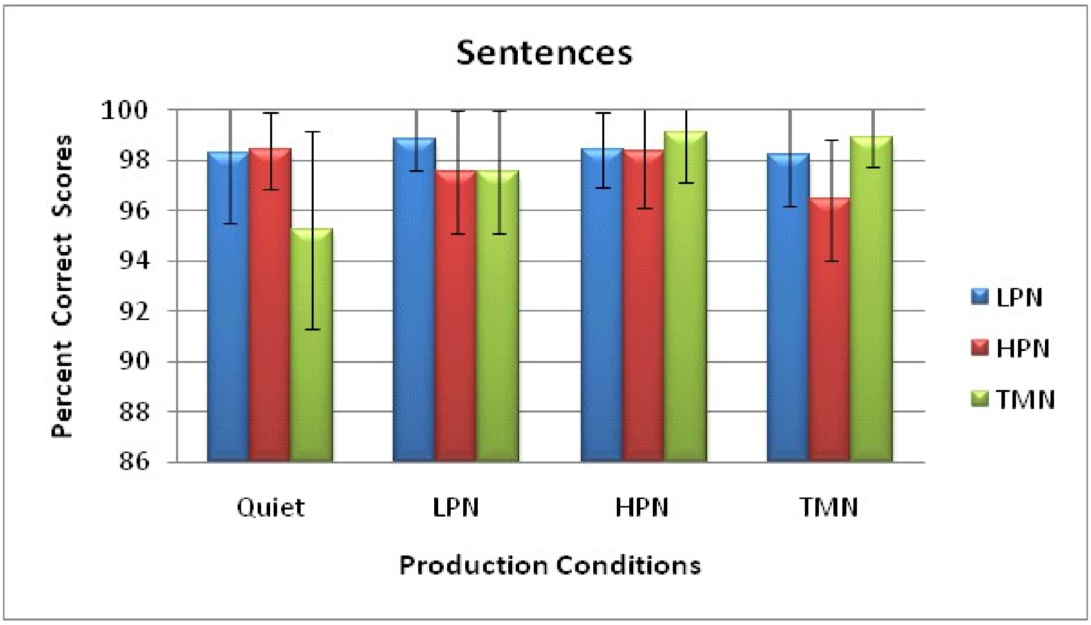


Figure 4.11. Percentage word correct scores for sentences across conditions. Different perception conditions are indicated by different colors.

As seen in Figure 4.11, the mean speech identification scores were similar across different listening conditions. These percentage correct scores were converted to rationalized arcsine transformed values for the purpose of statistical analysis. Two-way repeated measures ANOVA was done with production conditions and perception conditions as within subject factors, which showed there was no significant main effect of production conditions [F (3,27) = 1.921, $p > 0.05$] and no significant main effect of perception conditions [F (2,18) = 2.434, $p > 0.05$]. Hence it can be concluded that there was no significant difference between any of the conditions.

CHAPTER 5-DISCUSSION

Acoustic Analysis of Speech Produced in Noise

The 17 monosyllables produced in quiet and three noise conditions (low pass noise, high pass noise and temporally modulated noise) were subjected to acoustic analysis for six different parameters. The parameters were duration of monosyllables, intensity of monosyllables, fundamental frequency (Fo), first formant frequency (F1), second formant frequency (F2) and third formant frequency (F3). The results showed that there were significant differences in duration, intensity and fundamental frequency (Fo) for speech produced in noise with respect to quiet. For these three parameters, the speech produced in low pass noise (LPN) had maximum values and speech produced in quiet had minimum values. High pass noise (HPN) and temporally modulated noise (TMN) had values between the other two conditions and were not significantly different from each other. The formant frequencies were not significantly different across conditions. The findings are supported by the study conducted by Lu and Cooke (2009a) which reported that the Fo and intensity for speech produced in LPN and HPN were significantly more compared to speech produced in quiet. Furthermore, Fo and intensity for speech produced in LPN condition was more than the HPN condition, which is replicated in the current study. However, increase in F1 was also reported by Lu and Cooke (2009a) which is not reflected in the current findings. The results are also in accordance with the earlier studies which have compared Fo, intensity and duration with respect to speech produced in white noise and quiet (Dreher & O'Neill, 1957; Summers et al., 1988; Junqua, 1993; Tartter, Gomes, and Litwin, 1993). Tartter, Gomes, and Litwin (1993) found a decrease in F2 for noise induced speech with respect to quiet, which is not replicated in the present study.

Perceptual Analysis of Speech Produced in Noise

Perception of monosyllables. The recorded monosyllables and sentences were tested for perception by mixing with same and different maskers and also in quiet. In perceptual analysis of monosyllables produced in noise, it was found that irrespective of the production conditions, listeners always performed better in LPN perception condition, followed by TMN and HPN perception conditions, respectively.

Place and manner features were maximally transmitted in LPN. Voicing feature was maximally transmitted (proportion of 1) for both HPN and TMN conditions. Voicing consists of mainly low frequency information which gets effectively masked in the presence of low frequency noise (Miller & Nicely, 1955), which explains the poor perception of voicing feature in LPN perception condition, irrespective of the production conditions.

If we keep apart the LPN perception condition and then see the effect of production condition, then TMN reflected better perception in same noise condition for all features, where as HPN reflected better perception in same noise condition only for the place feature and the total transmitted information was poorest in HPN perception condition irrespective of the production conditions. Since all the vital information required for consonant perception is located in the higher frequencies, the high frequency noise effectively masks the place and manner information (Miller & Nicely, 1955). This might have resulted in poor results in HPN perception condition.

Temporally modulated noise facilitates 'dip listening' in the unmodulated gaps of noise (Füllgrabe, Frederic, & Lorenzi, 2006; Buss, Whittle, Grose, & Hall, 2009), which might have contributed to the increased perception compared to the steady state high pass noise condition.

Few of the earlier studies investigating acoustic and perceptual effects of speech produced in noise have found that speech produced in noise is more intelligible when compared to the speech produced in quiet (Dreher & O'Neill, 1957; Summers et al., 1988; Pittman & Wiley, 2001; Lu & Cooke, 2008). But in the present study, the spectro- temporal modifications found in noise induced speech did not aid in better perception as compared to speech produced in quiet condition. Possible reasons for this finding may be:-

1. Studies have found spectral shift of energy to higher frequencies for speech produced in noise with respect to quiet, which might have resulted in better perception of noise induced speech. Spectral tilt was not investigated in the current study. But, formant frequencies showed, no significant differences between different production conditions. This result could be interpreted as an indication of no much spectral change in the high frequencies which are vital for speech perception. Lu and Cooke (2009a) hypothesized that speakers actively shift their spectral energy distribution to regions least affected by the noise and measured the spectral parameters of speech produced in the presence of low and high-pass filtered noise. Parameters like Fo, F1 and spectral center of gravity effectively shifted towards the higher frequencies in low pass masking condition but did not shift towards the lower side, and increased instead in case of high pass masking condition. This might be one of the possible explanations for the poor perception in HPN condition even for speech produced in high pass noise.
2. Pittman and Wiley (2001) found intelligibility benefits in noise induced speech

but correlation analysis did not show any correlation between the spectro-temporal modifications of noise induced speech (spectral composition and duration) with the recognition score. The only parameter which showed the correlation was intensity. The study concluded that "increases in vocal level and spectral composition do not completely account for the observed increases in recognition". Hence it is possible that the observed spectro- temporal modifications of noise induced speech need not always result in perceptual benefits. In addition, in the present study, the stimuli for perception were normalized, thereby neutralizing the differences in intensity across different noise conditions.

3. Junqua (1993) found that, at equivalent SNRs, confusable monosyllables were less intelligible when produced in noise than in quiet. It can be recalled that maximum confusion in the current study occurred with confusable monosyllables like /d/- /ɖ/, /t/- /ʈ/ and /ʤ/- /j/ which highly affected the overall results, especially in HPN perception condition.
4. In the presence of noise, the vowels tend to be lengthened and consonants tend to be shortened (Junqua, 1993). Speakers tend to purposefully increase the vowel duration more in noise since vowels are audible at great distances and at high levels of noise. Also, as an articulatory response to the background noise, speakers tend to use a relatively more open articulatory postures which results in increased vowel duration (Junqua, 1993). These differential changes in duration of consonants and vowels might have contributed to reduced intelligibility for speech produced in noise compared to speech produced in quiet for certain perception conditions.

5. Some sounds are more susceptible to masking noise (stops, affricates, dentals, retroflex) than others (nasals, liquids, bilabials). Consequently, the intelligibility scores of these sounds tend to be poorer, largely biasing the overall results.
6. Lane and Tranel (1971) reported that speakers tend to modify their speech more when there is a communicative intent from their side. In the current study, there was no communicative intent present since the stimuli consisted of monosyllables and read sentences. This might have contributed to the reduced benefit of speech produced in noise.
7. The normalization of the stimuli might have artificially altered other acoustic-phonetic modifications of speech, other than intensity.

Perception of sentences. For sentence perception, there was no significant effect of both production and perception conditions. The contextual cues are more in sentences and they are highly predictable compared to other stimuli like words and monosyllables. This may be one of the possible reasons for obtaining such results. Other reason may be the signal-to-noise ratio (SNR) at which the perception task is carried out. As the sentence perception scores were at ceiling, it is possible that a SNR poorer than -5 dB SNR (which is used in the present study), is required to evidence the perceptual benefits of speech produced in noise.

CHAPTER 6-SUMMARY AND CONCLUSIONS

The aim of the present study was to measure the speech production modifications and perceptual benefits of speech produced in high pass, low pass filtered white noise and temporally modulated noise. The specific objectives of the study were

1) To compare the mean fundamental frequency, first three formant frequencies, duration and mean intensity of the speech produced in quiet, high pass and low pass filtered white noise and temporally modulated noise at syllable level.

2) To measure the speech intelligibility of speech produced in quiet, high pass and low pass filtered white noise, and temporally modulated noise in the presence of same and different maskers.

The method adopted in the study consisted of two phases. The first phase involved the acoustic analysis of speech produced in the presence of quiet and three types of noise and the second phase consisted of perceptual analysis of speech produced in noise. The three noise conditions were low pass noise (LPN), high pass noise (HPN) and temporally modulated noise (TMN). In first phase, recording of 17 monosyllables was done from 10 female participants. Sentences were recorded from one of the speakers. The monosyllables recorded in quiet and three noise conditions was acoustically analyzed for six parameters - duration, intensity, fundamental frequency (Fo), first formant frequency (F1) and second formant frequency (F2). In the second phase, perceptual benefits of the speech produced in noise was evaluated on 10 other listeners. For this purpose monosyllables and sentences produced in noise were mixed with LPN, HPN and TMN at -5 dB signal to noise ratio. Monosyllables were tested in a closed set while sentences were tested in an open set paradigm.

Monosyllables were analyzed by constructing the confusion matrices and by calculating the information transmitted for the features of voice, place and manner. Sentences were analyzed by calculating the total number of words repeated correctly.

Results showed that speech produced in noise had longer durations, higher F0 and higher intensities. There was no significant difference between the formant frequencies. Perceptual analysis showed that there was no significant benefit of speech produced in noise. In case of monosyllables, place and manner was best perceived when speech was mixed with LPN where as voicing was best perceived when speech was mixed with HPN and TMN. These results indicate that spectrum of the noise is more important during speech perception in noise rather than speaker intended acoustic-phonetic modifications.

REFERENCES

Amano-Kusumoto, A., & Hosom, J. P. (2011). A review of research on speech intelligibility and correlations with acoustic features. *Center for Spoken Language Understanding, Oregon Health and Science University.*

ANSI/ASA S3.1-1999 (R 2013) Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms. (n.d.), 1–17.

Boersma, P. and Weenink, D. (2014). Praat: doing phonetics by computer (Version 5.3.75) [Computer program]. Available from http://www.praat.org/

Bond, Z. S., Moore, T. J. & Gable, B. (1989). Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *Journal of the Acoustical Society of America, 85,* 907-912.

Boril, H. (2008). *Robust speech recognition: analysis and equalization of Lombard effect in Czech corpora.* Unpublished doctoral dissertation, Czech Technical University, Prague.

Buss, E., Whittle, L. N., Grose, J. H., & Hall, J. W. (2009). Masking release for words in amplitude-modulated noise as a function of modulation rate and task. *The Journal of the Acoustical Society of America*, *126*(1), 269.

Carhart, R., & Jerger, J. J., (1959). Preferred method for clinical determination of pure-tone thresholds. *Journal of Speech and Hearing Disorders*, 24, 330-345.

Charlip, W. S. & Burk, K. W. (1969). Effects of noise on selected speech parameters. *Journal of Communication Disorders, 2,* 212-219.

Dreher, J. J. & O'Neill, J. (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *Journal of the Acoustical Society of America, 29,* 1320-1323.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–124.

Füllgrabe, C., Berthommier, F., & Lorenzi, C. (2006). Masking release for consonant features in temporally fluctuating background noise. *Hearing Research*, *211*(1-2), 74–84.

Garnier, M., Bailly L., Dohen, M., Welby, P. & Loevenbruck H. (2006). An acoustic and articulatory study of Lombard speech: Global effects on the utterance. *International Conference on Acoustics Speech and Signal Processing,* 2246-2249.

Garnier, M. (2007). *Communication in noisy environments: from adaptation to vocal straining.* Unpublished doctoral dissertation, University of Paris, Paris, France.

Geetha, C. &and Pavan, M. & Kumar, S. (2012). *Development and Standardization of Sentence Test in Kannada Language for Adults.* Department of Audiology, All India Institute of Speech and Hearing, Mysore., Mysore.

Gnanateja, G. N. (2012). Speech in noise mixing, signal to noise ratio. http://www.mathworks.com/matlabcentral/fileexchange/37842-speech-in-noise-mixing-signal-to-noise-ratio, Matlab Central File Exchange. Retrieved on 03-11-13.

Gnanateja, G. N. (2014). Consonant confusion matrix. http://www.mathworks.com/matlabcentral/fileexchange/46461-consonant-confusion-matrix, Matlab Central File Exchange.

Hanley, T. D. & Steer, M. D. (1949). Effect of level of distracting noise upon speaking rate, duration, and intensity. *Journal of Speech and Hearing Disorders, 14,* 363-368.

Hansen, J. H. L. (1996). Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. *Speech Communication, 20,* 151-170.

Howell, P., Young, K. & Sackin, S. (1992). Acoustical changes to speech in noisy and echoey environments. *Proceedings of ISCA Tutorial and Research Workshop (ETRW) on Speech Processing in Adverse Conditions,* 223-226.

Johnson, M. (n.d.). FIX: Feature Information Xfer [computer software]. Available from http://www.phon.ucl.ac.uk/resource/software.html. Retrieved on 1/4/2014.

Junqua, J. C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America, 93,* 10-524.

Junqua, J. C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, *20*(1–2), 13–22.

Korn, T. S. (1954). Effect of psychological feedback on conversational noise reduction in rooms. *Journal of the Acoustical Society of America, 26,* 793-794.

Lane, H. L. & Tranel, B. (1971). The Lombard sign and the role of hearing in speech, *Journal of Speech, Language and Hearing Research, 14*, 677-709.

Letowski, T., Frank, T. & Caravella, J. (1993). Acoustical properties of speech produced in noise presented through supra-aural earphones. *Ear and Hearing, 14,* 332-338.

Lombard, E. (1911). Le Signe de l'Elevation de la Voix (The sign of the rise in the voice). *Ann. Maladiers Oreille, Larynx, Nez, Pharynx (Annals of diseases of the ear, larynx, nose and pharynx), 37,* 101-119.

Lu, Y. and Cooke, M. P. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *Journal of the Acoustical Society of America, 124,* 3261-3275.

Lu, Y. and Cooke, M. P. (2009a). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *Journal of the Acoustical Society of America, 126,* 1495-1499.

Lu, Y. and Cooke, M. P. (2009b). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication, 51,* 1253-1262.

Lu, Y. (2010). *Production and perceptual analysis of speech produced in noise.* Unpublished doctoral dissertation, University of Sheffield, Sheffield, United Kingdom.

MacDonald, E., & Raufer, S. (n.d.). Intelligibility of speech produced in temporally modulated noise. In *International Conference on Acoustics* (pp. 1297–1300).

Miller, G. A., & Nicely, P. E. (1955). An Analysis of Perceptual Confusions among some English Consonants. *The Journal of the Acoustical Society of America, 27 (*2), 338- 352.

Mixdorff, H., Pech, U., Davis, C., & Kim, J. (2007). Map task dialogs in noise – a paradigm for examining Lombard speech. *Proceedings of International Congress of Phonetic Sciences,* 1329-1332.

Paradigm v2.2.0.197 (n.d.) [computer software]. Retrieved on 12-02-14 from http://www.paradigmexperiments.com/index.html.

Patel, R. & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech Language and Hearing Research, 51,* 209-220.

Pisoni, D. B., Bernacki, R. H., Nusbaum, H. C. & Yuchtman, M. (1985). Some acoustic phonetic correlates of speech produced in noise. *International Conference on Acoustics Speech and Signal Processing,* 1581–1584.

Pittman, A. L. & Wiley, T. L. (2001). Recognition of speech produced in noise. *Journal of Speech Language and Hearing Research, 44,* 487-496.

Schiffman, H. F. (1983). *A Reference Grammar of Spoken Kannada*. University of Washington Press.

Sridhar, S.N. (1990). *Modern Kannada Grammar.* New Delhi: Routledge

Stanton, B., Jamieson, L. & Allen, G. (1988). Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions. *International Conference on Acoustics Speech and Signal Processing,* 331-334.

Steeneken, H. J. M. & Hansen, J. H. L. (1999). Speech under stress conditions: overview of the effect on speech production and on system performance. *International Conference on Acoustics Speech and Signal Processing,* 2079-2082.

Stowe, L. M., & Golob, E. J. (2013). Evidence that the Lombard effect is frequency-specific in humans. *The Journal of the Acoustical Society of America*, *134*(1), 640–647.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I. & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analysis. *Journal of the Acoustical Society of America, 84,* 917- 928.

Tartter, V. C., Gomes, H., & Litwin, E. (1993). Some acoustic effects of listening to noise on speech production. *The Journal of the Acoustical Society of America*, *94*(4), 2437–2440.

Varadarajan, V. S. & Hansen, J. H. L. (2006). Analysis of Lombard effect under different types and levels of noise with application to In-set Speaker ID system. *International Conference on Acoustics Speech and Signal Processing,* 937-940.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: a study of perceptual features. *The Journal of the Acoustical Society of America*, *54*(5), 1248–1266.

Webster, J. C. & Klumpp, R. G. (1962). Effects of ambient noise and nearby talkers on a face to-face communication task. *Journal of the Acoustical Society of America, 34,* 936-941.

Womack, B. and Hansen, J. (1996). Classification of speech under stress using target driven features. Speech Communication, 20, 131-150.

Yathiraj, A. & Vijayalakshmi, C. S. (2005). *Phonemically Balanced Word List in Kannada.* Department of Audiology, All India Institute of Speech and Hearing, Mysore.