USEFULNESS OF COGNITIVE CUEING IN ELICITING VOCAL VARIABILITY AND VOCAL NATURALNESS

Tiffy (George Roy)
Register No.: 08SLP032

A Dissertation Submitted in Part Fulfillment for the

Degree of Master of Science (Speech-Language Pathology),

University of Mysore, Mysore

ALL INDIA INSTITUTE OF SPEECH AND HEARING

MANASAGANGOTHRI

MYSORE-570 006

May 2010

Dedicated to my family and guide

Certificate

This is to certify that the dissertation entitled **"Usefulness of cognitive cueing in eliciting vocal variability and vocal naturalness"** is a bonafide work in part fulfillment for the degree of Master of Science (Speech Language Pathology) of the student Registration No. 08SLP032. 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 Diploma or Degree.

Mysore
May, 2010

Dr. Vijayalakshmi Basavaraj
Director
All India Institute of Speech and Hearing
Manasagangothri, Mysore- 570 006

Certificate

This is to certify that the dissertation entitled **"Usefulness of cognitive cueing in eliciting vocal variability and vocal naturalness"** is a bonafide work in part fulfillment for the degree of Master of Science (Speech Language Pathology) of the student Registration No.08SLP032. This has been carried out under my guidance and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Yeshoda K.

Mysore

May, 2010

Ms. K. Yeshoda

Lecturer

Department of Speech–Language Sciences

All India Institute of Speech and Hearing

Manasagangothri, Mysore- 570 006

Declaration

This dissertation entitled **"Usefulness of cognitive cueing in eliciting vocal variability and vocal naturalness"** is the result of my own study under the guidance of Ms. K. Yeshoda, Lecturer, Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Mysore

Registration No.08SLP032

May, 2010

ACKNOWLEDGEMENTS

'The Lord is my shepherd; I shall not want- Psalms23:1'

'This Shepherd of mine knows each trail, each snare,

And at just the right moment, my Lord will be there,

On His shoulders to carry each burden for me,

Yes, The Lord is my Shepherd and shall always be'

Thank you, God Almighty, for helping and guiding me to complete this study. Thank you for answering every prayer not in the way I wanted but in the way that seemed best for me. Unfailing is Your matchless love to me, so kind, so true and so pure...*My soul boasts in the Lord...*

Thank you, director ma'am, Dr. Vijayalakshmi Basavaraj, for allowing me to carry out this study.

'Everything is possible for him who believes'

Thank you, Yeshoda ma'am, for guiding me throughout with your valuable suggestions. Though I have stood countless number of times confused before you and confused even you, still you were patient to make my thought clear. Thank you ma'am for being the person I could approach to and for the faith you had in me. I thank my Lord for helping me to know a person like you in this walk of life. You are a very understanding guide.

'I know the plans I have for you, declares the Lord, plans to prosper you and not to harm you, plans to give you hope and a future - Jeremiah 29:11'

Daddy and mummy, ever so supportive. Your calls every Friday helped me keep goals the whole week. The words you both keep saying to me 'Look to the Lord and His strength, Seek His face always......' still ring in my ears as I write this. Blessed am I to have such parents.

My little siblings, Geoffrey and Christy, ever so naughty, but yet with words to keep me smiling. Thank you my dear ones.

Dearest Shamp, without your ears to hear my non-stop brooding over small matters, I do not think I could have made this possible. Your encouragement and prayers kept me on the run....Thank you for bearing with my temper tantrums all these 5 years...

'Together forever, never apart. Maybe in distance, but never in heart'

All my classmates thank you for being there for me. Pages are not enough to pick each of your names. I'll miss you all greatly. Hope our friendship never departs.....

Never to miss those who participated in this study, without which, this book would never exist. My sincere gratitude for listening and participating in the study. I know how much you enjoyed imagining things...

My seniors and juniors.......my heart fills with gratitude. It would be unfair if I miss Salome. Thank you for being here at the nick of time to encourage me....

Yet there are many more who helped me in small and big ways....pages aren't enough to mention them all..........My heartfelt thanks to all of them.....

"Praise be to the name of God forever and ever;

Wisdom and power is His.

He changes times and seasons;

He sets up kings and deposes them.

He gives wisdom to the wise and knowledge to the discerning.

He reveals deep and hidden things;

He knows what lies in darkness, and light dwells with him."

Daniel 2:20-22

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

"There can be no knowledge without emotion. We may be aware of a truth, yet until we have felt its force, it is not ours. To the cognition of the brain must be added the experience of the soul" - Arnold Bennett (1931).

INTRODUCTION

Everyday experience suggests that voice carries clues to the underlying emotional state of the speaker and this plays an important aspect in the study of human communication. An emotion is a mental and physiological state associated with a wide variety of feelings, thoughts, and behavior. Emotions are subjective experiences, often associated with mood, temperament, personality, and disposition. Emotions color the language, and can make meaning more complex. The speech of every individual engaged in earnest and animated conversation has a degree of variety and emphasis viewing it to be an overlaid function. The shades of meaning, attitudes and feelings of the speaker, which have intellectual and emotional connotations, are conveyed through the expressive quality of voice.

There are five features of emotions which include elicitors, receptors, states, expressions and experience. Emotional elicitors are events that trigger emotional receptors. Emotional receptors are relatively specific loci or pathways in the central nervous system that mediate changes in the physiological and /or cognitive states of the organism. The particular constellation of changes in the somatic and or neuronal activity that accompany the activation of emotional receptors is the emotional state and emotional expressions are the overt, observable surface features of changes in face, body, voice and activity level that accompany emotional states. Finally, emotional experiences are individuals' conscious or unconscious interpretations and evaluations of their perceived emotional states and expressions. In the organismic view, the connection between elicitors, receptors, expressions and states is biologically determined. Thus, elicitors are connected to receptors, which in turn produce certain states. States can be directly inferred from expressions. Emotional experiences (i.e. the interpretation of states) are often not considered aspects of emotion (Lewis, Haviland-Jones, Barrett, 1998). But in the light of James' (1884) view, 'emotion is a somatic change and the expression of that change'.

Generally materials from three categories are used in investigating emotional speech: spontaneous speech, acted speech and elicited speech. All three groups have both pros and cons, and none of the groups can be pointed out as generally optimal.

Spontaneous speech is often argued to contain the most direct and authentic emotions, but the difficulties in collecting this kind of speech are also extensive. The speakers' speech are recorded without their knowledge, so that they behave completely natural, but this kind of data collection rises difficulties, since such a routine is ethically problematic (Campbell, 2000, 2001).

Acted speech does not have the same ethical problems that are present in spontaneous speech, but the degree of naturalness is often questioned. Acted speech conforms to stereotypes of how people believe that emotions should be expressed in speech, not on how emotions actually are expressed. This indicates that acted speech is more stereotypical, and that the expression of emotions is more extreme than in spontaneous speech (Banse and Scherer, 1996; Bortz, 1966; Cosmides, 1983; Green and Cliff, 1975; Kaiser, 1962).

In elicited speech, certain emotions are induced through procedures like watching a film, which evokes specific emotions. Here the idea is that the speech shall be coloured by the emotion induced. It is also possible to put a subject into a situation meant to evoke a specific emotion, and then record her/his speech. However, this method suffers from ethical problems, i.e., it is not fully ethical to scare someone, and then record her/his speech (perhaps it is even more unethical to do this, than to just record someone who is already scared). The induction method has the positive feature that it gives control over the stimulus; however, different subjects may react differently on the same stimulus. The validity of such elicited or induced speech depends to a large extent on how successful the induction process was (Stibbard, 2001).

In the past, there have been several attempts to classify the emotive state by a speaker on basis of prosody and voice quality. This implies the assumption that voice alone really carries full information about emotive state by the speaker. Few studies using spontaneous speech are reported in literature. The work in this area has made use of materials that was recorded during naturally occurring emotional states like dangerous fight situations, journalists reporting emotion-eliciting events, talk and game shows on TV (Eldred & Prince, 1958; Hargreaves, Starkweather & Blacker, 1965; Huttar, 1968; Williams & Stevens, 1969, 1972; Niwa, 1971; Cowie & Douglad-Cowie, 1996; Frolov, Milovanava, Lazarev & Mekhedova, 1999; Johannes, Petrovitsh Salnitski, Gunga & Kirsch, 2000). But Stibbard (2001) found nearly no correspondences at all between emotive categories and acoustic correlates. Though the use of naturally occurring voice changes in emotionally charged situations seems to be the ideal research paradigm owing to ecological validity, there are serious methodological problems. The voice samples obtained in natural situations, for a small number of speakers, were generally very brief and suffer from bad recording quality. There are also problems in determining the precise nature of the underlying emotion and the effect of regulation. However, in studies using acted (Kaiser, 1962; Green & Cliff, 1975; Cosmides, 1983;

Banse & Scherer, 1996) or elicited speech (Skinner, 1935; Roessler & Lester, 1976; Havrdova & Moravek, 1979; Hicks, 1979; Scherer & Tolkmitt, 1986; Bachorowski & Owren, 1995) the results are more encouraging and researchers report quite good correspondences.

Scherer (1986) proposed a theoretical model of vocal affect expression which says in vocalization process the physiological arousal due to an emotion is an involuntary force. Since voice carries information about the emotional state of the individual, speech emotion analysis can be done to profile the acoustic patterns of different emotions. Speech emotion analysis refers to the use of various methods to analyze vocal behavior as a marker of affect (e.g., emotions, moods, and stress), focusing on the nonverbal aspects of speech. The basic assumption is that there is a set of objectively measurable voice parameters that reflect the affective state a person is currently experiencing. Speech emotion analysis is complicated by the fact that vocal expression is an evolutionarily old nonverbal affect signaling system coded in an iconic and continuous fashion, which carries emotion and meshes with verbal messages that are coded in an arbitrary and categorical fashion.

Emotions have a physiological effect on the larynx which is reflected in the acoustic parameters of speech. Findings reported for the emotion of happiness were quite consistent between researchers. Risberg (1986) reported that happiness gave an increase in pitch and pitch range while Öster and Risberg (1986) noted a slow tempo. Fónagy & Magdics (1963) described it as "lively". However, findings described sadness as exhibiting normal or lower pitch, narrow pitch range and slow tempo (Skinner, 1935; Davitz, 1964). Comparing both these emotions, neutral emotion has a lower pitch, with pitch tending to be normally distributed about the average pitch level (Cowan, 1936).

The efficacy of cognitive therapy in the modification of certain cognitive processes resulting in the amelioration of symptoms is well documented in psychological literature. The assumption underlying cognitive therapy models have direct relevance to the area of vocal performance. Mental images and cues have traditionally been employed by voice teachers and voice clinicians to elicit and shape vocal behavior.

Cognitive cueing is an approach to voice treatment that stimulates thought patterns changing the speakers' voice. It focuses the attention of the speakers on the actions and images that are suggested by the information that is given (in the form of written text) and also on the speakers thought and feelings regarding that text. Cues of this type have been traditionally used by drama and singing teachers to elicit vocal behaviors (Andrews, Shrivastav, & Yamaguchi, 2000; Bohnenkamp, Andrews, Shrivastav, & Summers, 2002).

In the present study, cognitive cueing was used as a task elicitation method to understand vocal variability. To understand the role of cognitive cueing and the variability that may be present in voice, three basic emotions were considered as they are believed to have distinct acoustic patterns. The purpose of the present study was to understand if vocal variability was evident after the presentation of cognitive cues and does this variability occur in any direction of the emotional pattern presented. This will indirectly throw light on the relationship of mental imagery and vocal variability. This study will, thus, examine task elicitation with cognitive cues for and its effects on the acoustic parameters of voice.

CHAPTER II

REVIEW OF LITERATURE

There are two requirements for the production of sound of any kind: source of energy and vibrating structure. The prime source of energy for voice production is a smooth flow or air from the lungs which the vibrating vocal folds convert into sound. Energy in the form of air from the lungs passes into the trachea and the larynx. The larynx is the principle structure for producing a vibrating airstream and the vocal folds which are a part of the larynx make up the vibrating elements. Voice may be defined as the laryngeal tone which can be heard or measured. The normal voice should possess certain characteristics of pitch, loudness and quality, which make the meaning clear, arouse the proper emotional response or ensure a pleasant tonal effect upon the hearer.

Human voice and speech patterns are distinctive. Voice 'fingerprints' are composed of a number of speech-voice behaviors that act in combination as well as separately: the rate of speech, the number of words said in one breath, the rhythm of speech, ease in breathing, the pitch and loudness of voice, the relative relaxation or tension of voice, the individuals mood state, clarity of speech articulation and the resonance of voice. All of these speech-voice behaviors blend together and what the final end product is a distinct, individual's voice.

The notion that vocal acoustics are rich with cues to a vocalizer's emotional state has a long history, with Cicero (55 B.C.E/2001) suggesting that each emotion is associated with a distinctive tone of voice. The first comprehensive account of vocal emotion was provided by Darwin (1872), whose detailed comparative descriptions led to the conclusion that affective expressions, including those provided via the vocal channel, are veridical. This was followed by Fairbanks and Pronovost (1939) who reported that fundamental frequency of speech varied with simulated expressions of contempt, anger, fear, grief and indifference. Scherer (1986, 1989) and Banse & Scherer (1996) posited that discreet affective states experienced by the vocalizer are reflected in specific patterns of the acoustic cues in the speech being produced.

Emotion is not a simple phenomenon and many factors contribute to it. Izard (1977) stated that a complete definition of emotion must take into account the experience or the conscious feeling of emotion, the processes that occur in the brain and nervous system and the observable expressive patterns of emotions. Scherer (1981) described emotion as 'the organism's interface to the outside world'.

The expression of emotions plays a key role in human communication. Besides language, we also use non verbal cues such as facial expression and vocal expressions to convey emotions. The voice is important as an emotional medium as it is the carrier for spoken language. Additionally, listeners can perceive vocal expressions across long distances and in situations where facial or gestural information is not available as in the conversation using the telephone. Everyday experience shows that auditory stimuli routinely induce emotion-related reactions in listeners. These stimuli can vary from sounds such as a dramatic booming thunderstorm to a subtle dripping faucet. From one perspective, it is said that vocalizers encode affect related meaning in their signals and listeners subsequently decode that information (Laukka, 2005). Emotional expression is thus treated as a kind of code that vocalizers use to represent their emotional states in almost linguistic-like fashion.

From an evolutionary perspective, the key to understanding emotions is to look at the functions they serve (Keltner & Gross 1999). The two research traditions that have most strongly influenced research on vocal expression are discreet and dimensional emotional theories. According to discreet emotional theory, each person represents a unique person environment interaction with ones own adaptation significant for the individual. Each discreet emotion is thought to have its own unique pattern of cognitive appraisal, physiological activity, action tendency and expression (Darwin, 1872; Ekman, 1998). According to several discreet emotional theories, there exist a limited number of basic emotions such as fear, anger, happiness and sadness. The dimensional approach to emotion concentrates on the subjective feeling state and focuses on identifying emotions based on their placement on a small number of underlying dimensions. Wundt (1924) suggested three dimensions (pleasure-displeasure, strain-relaxation, and excitement-calmness) that account for all differences among emotional states.

The commonly measured acoustic correlates of emotions are pitch (fundamental frequency, both average and range), duration, intensity and voice quality (Murray & Arnott, 1993). Scherer (1981) noted various vocal indicators of emotional expression principally pitch, intensity and vocal quality with pitch which is often used to differentiate between emotions.

Studies on the effects of emotions on the acoustic characteristics have shown that average values and ranges of F0 differ from one emotion to another. There are several reasons why changes in F0 with time are potentially capable of providing information concerning the emotional state of the speaker. First, considerable latitude of change is possible in the variation of F0 since only certain aspects of F0 contour carry information with regard to the linguistic content of a message. The principle linguistic functions of F0 changes are to indicate stress and to mark boundaries of different types of sentence length or phrase length units. Subject to these constrains, a speaker is relatively free to use changes in F0 to convey nonlinguistic information, such as, emotions. The fundamental frequency can undergo variations that may not be intended or be under overt control of the speaker and hence may provide an indication of the speaker's emotional state. Thus emotional states are said to influence vocal quality as a result of changes in the muscle tonus. These changes are primarily brought about by the functioning of the sympathetic division of the automonic nervous system.

Cowan (1936) stated that unemotional speech has a narrow pitch range compared to emotional speech, with pitch tending to be normally distributed about the average pitch level. Williams & Steven (1972) also reported that in neutral situations, the sentences were generated with shorter duration than for the emotional situations.

Cowan (1936) and Öster & Risberg (1986), reported that happiness gives an increase in pitch and pitch range. Öster & Risberg (1986) also noted a slow tempo, while Fónagy & Magdics (1963) described it as "lively". Davitz (1964) reported an increase in speech rate along with an increase in intensity.

Williams & Steven (1972) observed the average fundamental frequency for speaking in sorrow situations and found it to be considerably lower than that for neutral situations and the range of F0 was usually quite narrow. This change in F0 was accompanied by a marked decrease in the rate of articulation and an increase in the duration of an utterance. The increased duration resulted from longer vowels and consonants and from pauses that were often inserted in a sentence. In the review of Johnson and Scherer & Ceshi (2000), sadness was proved to decrease the mean fundamental frequency, F0 range and variability, speed and articulation rate and intensity. The patterns were contrary to the acoustic patterns in happiness in which a rise in the mentioned patterns was seen.

In comparison with the neutral emotion, Murray & Arnott (1993) classified happy and sad emotion on some acoustic characteristics. Happy emotion elicited a faster or slower rate, had a higher pitch and a wider pitch range, a higher intensity, with smooth upward inflections than neutral emotion. The voice quality seemed to be breathy or blaring and the articulation was classified to be normal. Comparing sad emotion with neutral emotion, it was reported that the average pitch and pitch range was slightly lower with reduced intensity and slower speech rate for sad emotion. The voice quality in this emotion was reported to be resonant and had a downward pitch inflection with articulation reporting to be slurred.

Scherer (1986) concluded that F0, energy and rate may be the most indicative of arousal. Arousal is defined as a subjective state of feeling activated or deactivated (Sanchez, Kirschning, Palacio, Ostrovskaya, 2005). Changes in F0 and rate of speech can be attributed to many factors such as general temperamental and personality characteristics of the individual. Individuals vary greatly in the rapidity of their motor activity such as walking, as well as with alacrity with which they think and come to a conclusion and even express themselves. Individuals who think slowly, move ponderously also speak with a deliberate, even tempo while for individuals who are active and quick, their speech is likely to be hurried and lively. Rate can also be a characteristic of the mental state. Slow rate can be a characteristic of mental state such as wonder, doubt, deep thought and sorrow while rapid tempo is associated with joy, excitement, humor and anger.

A mental image is an experience that, on most occasions, significantly resembles the experience of perceiving some object, event, or scene, but that occurs when the relevant object, event, or scene is not actually present to the senses (McKellar, 1957; Richardson, 1969; Finke, 1989; Thomas, 1999). Functional magnetic resonance imaging (fMRI) studies have shown that the corticomotor areas activated during mental imagery of speaking and singing overlap with those activated during actual speaking or singing (Shergill, Bullmore, Brammer, Williams, Murray, McGuire, 2001; Kleber, Birbaumer, Veit, Trevorrow, Lotze, 2007). The central processes involved during mental imagery of a movement are also largely identical to those required when performing that movement.

There are many methods for collecting vocal expression based on induction procedure (Skinner, 1935; Scherer & Tolkmitt, 1986; Havrdova & Moravek, 1979; Hicks, 1979; Roessler & Lester, 1979; Bachorowski & Owren, 1995). Materials can be displayed or selected for its emotional impact like photographs, movie clips or music excerpts and imagination techniques can be used. Majority of studies on vocal expression have used some variant of the standard content paradigm. In this paradigm, the person (e.g., an actor) is instructed to read some verbal material aloud, while simultaneously portraying particular emotions chosen by the investigator. Another method in eliciting vocal production is providing verbal cues, visual cues, and/or cognitive cues.

The use of cognitive cues as a task elicitation method stems from cognitive-behavioral therapy and focuses on prompting the individual to think about and feel a task prior to completing the task. Cognitive-behavioral therapy targets both cognition and behavior as primary change areas and focuses on three main principles: (a) cognitive activity affects behavior, (b) cognitive activity may be monitored and altered, and (c) desired behavior change may be affected through cognitive change (Dobson, 2001). Cognitive cues have been utilized when measuring voice productions and have been found to increase vocal variation in adults and children 9.0 to 11.11 years of age (Andrews, Shrivastav, & Yamaguchi, 2000; Bohnenkamp, Andrews, Shrivastav, & Summers, 2002). When utilizing cognitive cues as an elicitation method, the results indicated that all speakers increased the frequency range and altered the rate of speech.

Bohnenkamp et al. (2003) targeted vocal variation. The study utilized cognitive cues in an attempt to alter rate of speech. The cues prompted participants to make the listener feel that he was running faster in order to increase rate of speech. Thus studies found cognitive cues to have a significant effect on increasing vocal variation and to be a promising task elicitation method (Andrews et al., 2000; Bohnenkamp et al., 2002).

A key aspect of cognitive cueing is an individual's level of cognitive development. For cognitive cueing to be effective, an individual must be able to form a mental representation of objects to realize that objects exist even when they are not in sight (Piaget, 1952). Developmental specialists believe the ability to form mental representations develops from 2.0 to 8.11 years of age (Piaget, 1952; White, 1965). In addition to forming mental representations, the ability to reason is also critical when utilizing cognitive cueing. The ability to reason is thought to develop around 5.0 to 7.11 years of age, thus representing a minimum age range when cognitive therapy would have the potential to be effective (Piaget, 1952; White, 1965).

Reich, Mason, and Schlauch (1989) utilized aspects of cognitive training to elicit five stimulus-tone conditions of (a) discrete steps, (b) slow steps, (c) fast steps, (d) glissando, and (e) fast glissando when studying F0 range in children ages 8.0 to 11.11 years with healthy voice quality. Each task required the participants to think about their voice and produce a fast, slow, smooth, or divided voice production. Results indicated that all the stimulus conditions evoked significantly higher maximum frequencies than the discrete steps condition, which generated the smallest frequency ranges when compared to the other conditions.

Cognitive cueing as an approach to voice treatment stimulates voice patterns as a way of changing speakers' voices. Cognitive cues have been utilized when measuring voice productions (Andrews, Shrivastav, & Yamaguchi, 2000; Bohnenkamp, Andrews, Shrivastav, & Summers, 2002). Voerman, Langeveld & Van Rossum (2009) conducted a study on the short- and long-term results of two techniques (mental imagery and manual shaking of the larynx) in 116 patients with non-organic dysphonia or aphonia. It was found that the group of patients who retrieved their voice was subjected to mental imagery tasks but relapses were more common in this group. It was thus concluded that the cure rate for mental imagery was much higher than that for laryngeal shaking.

It has been reported that expression of emotions can be mediated by personality characteristics (Barrett & Niedenthal, 2004; Gross, John & Richards, 2000). This entails that some people are more expressive than others and may be more emotional than others. Gender differences have also been reported stating that women are more expressive than men.

Anderson (1961) formulated the quantification for characterizing an effective or natural speaking voice. He suggested the following

1. Adequate loudness: It is essential to have appropriate loudness when speaking. Individuals whose voices are thin and light are distinctly handicapped as it is difficult for them to create an impression of confidence, assurance and self possession.

2. Clearness and purity of tone: A good voice is characterized by clear voice and good quality. It must be free from disturbing, unpleasant elements in the tone.

3. A pleasing effective pitch level: For every voice, there is a pitch at which the voice sounds pleasant and effective.

4. Ease and flexibility: The normal voice is responsive and characterized by a degree of variety and melody contributing to what is called expressiveness. Absence of this quality makes the voice sound monotonous and dull.

5. Clearness and ease of diction: The good voice is easily intelligible without being conspicuously so. The speaker can be clearly understood because his diction is clear. He does not mutilate his speech by omitting or swallowing sounds and syllables nor does it appear that it is an effort for him to speak clearly.

The dictionary defines naturalness as ease and ease as freedom. The adjective 'natural' derived from the Latin word 'naturalis' means of nature meaning true to nature and freedom from artificiality, affectation or constraint. Till today, there is no clear definition of naturalness. The development of speech naturalness scales started with the work of Ingham and Packman (1978), who developed a binary scale based on certain features of speech behavior (prosody, rate, fluency and naturalness).

There are numerous studies that speak of the measurement of naturalness. But these studies pertain to its evaluation in stuttering especially to assess naturalness of speech in recovery from stuttering. There are limited studies on the naturalness of speech with respect to normals and its influence on emotions.

As stated by Scherer (1986) vocal parameters may be the most indicative of arousal and findings shed light to expanding knowledge of the extent of variations that can be present in the subjects subjected to them. Hence, in the present study an attempt was made to understand the usefulness of cognitive cueing in eliciting vocal variability employing different emotions and also assess the naturalness of the same.

Aims of the study

- To quantitatively analyze the vocal variability of speech in uncued and cued conditions across emotions and compare the same.
- To assess the naturalness of speech qualitatively in uncued and cued conditions across emotions.
- To correlate the obtained quantitative and qualitative measures.

CHAPTER III

METHOD

Participants (Speakers)

Twenty participants (10 males and 10 females) in the age group of 20 to 25 years with mean age of 22.5 years were considered for this study. Participants who were fluent and competent in English language use were only chosen. They were selected if they had English as their medium of education from primary school and if they used English often as their language to communicate. The participants selected in this study had Kannada as their first language and English as their second language. Participants were excluded if they had a velopharyngeal disorder, or were perceived by a speech language pathologist to have a voice disorder, abnormal oral-peripheral structures, or hearing loss, neurological or psychological problems.

Participants (Judges for qualitative analysis)

Five qualified Speech Language Pathologists (three males and two females) with a minimum of two years of clinical experience formed the judges for the perceptual experiment in qualitative analysis

Procedure

Materials (Quantitative analysis)

Three primary emotions were considered in this study: happiness, sadness and neutral emotion. Sentences that evoked neutral, happy and sad emotion were considered. For each emotion two sentences (a total of 6) were constructed and written on A2 size cards. For every sentence cognitive cues were constructed and all cues were audio recorded. These cues were targeted to create specific mental images and to elicit the chosen emotions.

Instructions

All the participants were instructed as follows: ‘A few written sentences will be shown. You are required to read the sentences aloud. After reading all the sentences once, you will have to listen to a few audio recordings and then read the same sentences again. All your speech samples will be recorded’.

Recording

All the read samples were recorded in a quiet environment on to Olymus Digital Audio recorders (WS-100, Japan) with Verbatim Headset with Collapsible Microphone, (Recorder No: 41691) with the microphone-mouth distance of 5” – 6”.

Each participant was asked to read the sentence three times.

Trial 1: The cards with sentences were presented individually and instructed to read all the sentences aloud. This was the no cue condition.

Trial 2 and 3: The subjects were asked to read the sentences after listening to the recorded cognitive cues. This was the cognitive cued condition. The sentences and cues used are shown in the Appendix I.

Instrument

Quantative assessment was carried out using the software Real Time Pitch software (Model 5121) of Computerized Speech Lab 4500 (KAY Elemetrics).

Acoustic analysis involved two phases. The samples of trials 1 and 3 were considered for acoustic analysis.

Phase I - Quantitative analysis

The quantitative analysis involved extraction of the following parameters using the software Real Time Pitch.

1. Mean speaking fundamental frequency (MSF0) – It is the average pitch that is used during speaking and is expressed in Hertz (Hz).
2. Standard Deviation of Speaking Fundamental frequency (SDSF0) – The standard deviation reflects the frequency variability for a reasonably large time segment or passage.
3. Lowest speaking fundamental frequency - The lowest pitch the individual can produce during speaking and is expressed in Hertz (Hz).
4. Highest speaking frequency - The highest pitch the individual is able to produce during speaking and is expressed in Hertz (Hz).
5. Variation in Fundamental frequency (vF0) - It is the relative standard deviation of fundamental frequency which reflects the variation of F0 within the analyzed voice sample. It is expressed in terms of %.
6. Sentence Duration – It is the period of time during which the sentence is spoken, expressed in seconds.

Phase II: Qualitative analysis

Materials

The samples of subjects of the uncued trial (trial 1) and cued trail (trail 3) were transferred onto an audio CDS. For each subject six tracks were created and in total, 120 tracks formed the material for qualitative analysis.

Instructions

Each of the judges were instructed to listen to the audio samples individually and rate each of the sentences (uncued and cued) on a three point rating scale where 1 represented least variable and 3 represented most variable. They were given a free hand and instructed to listen to the samples as many times required and rate the sentences.

Qualitative parameters

Naturalness of speech is ambiguous and most difficult to quantify. Review of literature indicated no single attribute or any established attribute to signify naturalness. Parameters of speech as shown below were chosen for description of speech naturalness.

1. Voice - Pitch, loudness, and quality
2. Articulation - Pronunciation and intelligibility
3. Fluency - Rate, effort and continuity
4. Prosody - Intonation and stress

Scoring sheets incorporating these parameters were prepared and presented to the judges.

Statistical analysis

All the extracted vocal parameters of the quantitative analysis of the two trials (uncued and cued) were subjected to statistical analysis. Descriptive measures, repeated measure ANOVA and paired t test were obtained to check for significance, if any, across the trials, emotions and gender.

Qualitative analysis: individual scores of the judges were pooled emotion wise and overall percentages were calculated. The results are tabulated and presented.

CHAPTER IV

RESULTS AND DISCUSSION

A total of twenty subjects selected for the study. Among them, ten were females and ten were males. The task was to read sentences that were presented to them. In the first trial, the subjects had to read the sentences and the samples were recorded. In the cued trials, the subjects had to read the sentences after listening to the cognitive cues. The uncued and the cued samples were considered for analysis. The analyzed data were subjected to statistical analysis. The descriptive statistics, repeated measure ANOVA and paired t test was obtained.

The results will be discussed under the following headings

1. Quantitative analysis

    a) Mean Speaking Fundamental frequency (MSF0)

    b) Standard deviation of Speaking Fundamental Frequency (SDSF0)

    c) Minimum Speaking Fundamental frequency

    d) Maximum Speaking Fundamental frequency

    e) Variation of Speaking Fundamental Frequency (VSF0)

    f) Sentence duration

2. Qualitative analysis

Quantitative analysis

a) Mean Speaking Fundamental Frequency (MSF0)

Table 1: *Mean Speaking Fundamental frequency (MSF0) in uncued and cued condition across gender and emotions.*

Emotions	Gender	Mean(Hz)	SD	Emotions	Gender	Mean(Hz)	SD
UCN	Female	228.56	21.54	CN	Female	229.51	31.48
	Male	130.61	11.93		Male	140.84	26.85
UCH	Female	222.61	20.51	CH	Female	247.49*	26.41
	Male	125.83	8.54		Male	169.12*	41.18
UCS	Female	219.33	20.95	CS	Female	220.19	22.64
	Male	125.99	7.99		Male	125.12	14.23
UC Total	Female	223.50	21.00	C Total	Female	232.40	28.97
	Male	127.48	9.74		Male	145.03	34.36

UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

Table 2: *F value and significance of the overall effect of uncued and cued conditions across gender and emotions for MSF0.*

MSF0	F	Sig
Change across emotion	21.205	0.000*
Change across gender	3.567	0.061
Overall change between uncued and cued conditions	33.303	0.000*

*Level of significance (*p<0.05)*

Table 3: *Paired t test and significance for MSF0 between cued and uncued conditions across emotions and gender.*

MSF0 between emotions and trials	Gender	t (19)	Significance (2 tailed)
UCN – CN	Females	0.163	0.873
	Males	1.993	0.061
UCH-CH	Females	5.173	0.000*
	Males	5.096	0.000*
UCS-CS	Females	0.167	0.896
	Males	0.334	0.742

Level of significance (*$p<0.05$); UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

The mean speaking fundamental frequency (MSF0) for the uncued trail and cued trial is shown in Table-1. In the uncued neutral emotion, the female group obtained a mean value of 228.56 Hz (SD=21.54) which was not significantly different (Table 3) from the cued neutral emotion with a mean value of 229.51 Hz (SD =31.48) Also, there was no significant difference between the mean scores obtained amongst the uncued and cued condition in neutral emotion for males with values of 139.61 Hz (SD = 11.93) and 140.84 Hz (SD = 26.85) respectively. The emotion of happiness yielded a significant difference between the trials and across gender. The mean scores secured by females for uncued trail was 222.61 Hz (SD= 20.51) which rose to 247.49 Hz (SD=26.41) for the cued condition. Similarly, males obtained a value of 125.83 Hz (SD = 8.54) in the uncued condition and 169.12 Hz (SD=41.18) in the cued condition. However, there was no difference in the means obtained across the trials for sad emotion for males and females. The females obtained a score of is 219.33 Hz (SD = 20.95) and 220.19 Hz

(SD = 22.64) in the uncued and cued condition respectively and males secured a mean value of 125.99 Hz (SD = 7.99) and 125.99 Hz (SD = 14.23) for the uncued and cued condition respectively.

The overall level of significance was calculated using the repeated measure ANOVA and is displayed in Table 2. There was a significant difference observed across the three emotions [$F(2, 114) = 21.20$, $p<0.05$] and no difference across gender [$F(1,114)= 3.567$. $p>0.05$]. An overall change between the uncued and cued trials was also seen [$F(1,114) = 33.30$, $p<0.05$].

In general, the MSF0 varied across conditions and emotions in both males and females as expected. However, a significant difference was observed only for the emotion happy.

The results conform to the findings of the acoustic patterns of the emotions reported by various authors. Experiments have proved that in unemotional speech such as that of neutral emotion, there will not be much changed present in the mean F0 and its parameters and the above results are in line with the studies stated by Cowan (1936) and William, & Steven (1972). The results are in parallel to the study done by Murray & Arnott (1993) as happy emotion has shown an elevation in mean F0 while sad emotion had not changed considerably from the uncued and the cued trial. This difference was noticed irrespective of gender. It can be thus assumed that the presentation of cognitive cues elicited vocal variability in the direction of emotions. It is evident that the subjects were influenced by the cues given to them and this influence is manifested in the overt, vocal expression.

b) *Standard deviation of speaking fundamental frequency (SDSF0)*

Table 4 depicts the mean of the standard deviation of fundamental frequency (SDSF0). SDSF0 reflected the frequency variability across the two conditions [F(1,114)=22.29, p<0.05] and across emotions [F (2,114)=8.75, p<0.05] but there was no significance across gender [F (1,114)=0.82, p> 0.05] (table 5).

Table 4: *Standard deviation of speaking fundamental frequency (SDSF0) in uncued and cued condition across gender and emotion.*

Emotions	Gender	Mean	S D	Emotions	Gender	Mean	SD
UCN	Female	25.01	8.61	CN	Female	42.27*	12.65
	Male	19.21	12.29		Male	24.67	13.20
UCH	Female	32.57	13.95	CH	Female	45.37*	11.66
	Male	21.57	10.09		Male	33.89*	12.95
UCS	Female	33.90	16.69	CS	Female	30.16	10.28
	Male	19.46	10.88		Male	19.52	9.11
UC Total	Female	30.49	13.85	C Total	Female	39.26	13.16
	Male	20.08	10.99		Male	26.03	13.14

UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

It was seen that in neutral emotion, females were observed to have a significant variability in frequency from the uncued trail to the cued trial. Their values were 25.01 (SD=8.61) and 42.27 (SD=12.65) respectively. However, males did not have a significant difference in the SDSF0 in the neutral emotion. They secured a value of 19.21 (SD=12.29) and 24.67 (SD=13.20) respectively. Happy emotion, showed a significant difference between the trails and across gender. There was a significant rise from 32.57(SD=13.95) in uncued condition to 45.37 (SD=11.66) for females and a rise of values from 21.57 (SD=10.09) to 33.89 (SD= 12.95) in uncued and cued trials respectively in males. The values did not differ to a large extent for sad emotion in females and males. Females obtained a SDSF0 of 33.90 (SD=16.69) in uncued trial and 30.16 (SD=10.28) in the cued trial. Males also obtained a score of 19.46 (SD=10.88) in the uncued trial and it remained the same for the cued trial with a value of 19.52 (SD=9.11). this is more clearly evident in table 6.

Table 5: *F value and significance of the overall effect of uncued and cued conditions across gender and emotions for SDSF0.*

SDSF0	F	Sig
Change across emotion	8.75	0.000*
Change across gender	0.82	0.367
Overall change between uncued and cued conditions	22.29	0.000*

Level of significance (*p<0.05)

As seen in Table 5, there was a significant difference in the reading of the subjects before the cues were presented an after the cues were presented. There is an obvious difference across the emotions as the acoustics between the emotions vary. It can be observed that variation of F0 was more for females than for males in the neutral emotion. This is quite contrary to the characteristics pattern for an unemotional speech. This could be because of the cognitive cues used to elicit the sentence. For the neutral emotion, the sentence, 'this is a pen' was considered. The cue given for this sentence was to imagine the concept of a pen being taught to a child. It is well known that in sociolinguistics, a sentence when used with children will be modified drastically. In this study, although males had variations in F0, females were observed to have had a significant variation. This is consistent with the studies with respect to gender that females are more expressive compared to the male group. Happy emotion elicited a significant upward inflection after the cognitive trial which is in accordance to the studies by authors, Cowan (1936), Öster & Risberg (1986), Murray and Arnott (1993). But for the emotion of sadness, no significant change could be observed between the conditions. Johnson and Scherer (2000) found sadness to decrease the mean fundamental frequency and have a lesser variations. In the present study, the F0 did not vary significantly for the neutral emotion and the variations in F0 were minimum.

Table 6: *Paired t test and significance for SDSF0 between cued and uncued conditions across emotions and gender.*

SDSF0 across emotions and trials	Gender	t (19)	Significance (2 tailed)
UCN – CN	Females	4.779	0.000*
	Males	1.305	0.208
UCH-CH	Females	3.276	0.004*
	Males	3.563	0.002*
UCS-CS	Females	0.922	0.368
	Males	0.017	0.987

Level of significance (*p<0.05); UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

c) Lowest Speaking fundamental frequency

Table 7 shows the mean scores of lowest speaking fundamental frequency. Repeated measure ANOVA shows that there is no significant difference between the trials, emotions and gender [F (1,114) = 2.37, p>0.05], [F(2,114)= 2.02, p>0.05] and [F (1,114) = 1.83, p>0.05] (table 8). Analyzing table 7 and table 9, a significant difference is observed only for females in the neutral emotion. There is decrease in low F0 from 173.98 (31.48) to 144.66 (49.62) from uncued to cued condition. As for males, there is no statistical difference in the mean low F0 as the value obtained were 103.28 (13.65) in uncued condition and 97.09 (20.79) in cued condition. In happy emotion, there was a slight decrease in LSFO from 137.91 (39.06) to 133.27 (40.89) in females and in males, there was a slight increase from 95.83 (11.28) to 101.04 (27.28). As for sad emotion, there was only a difference of approximately 1Hz between the uncued and cued condition. Females scored a value of 145.91 (38.94) in uncued condition and it decreased to 144.65 (42.83) in cued condition. Males obtained a value of 95.36 (11.85) in uncued condition and it decreased to 94.05 (11.88) in cued condition.

Table 7: Lowest speaking fundamental frequency in uncued and cued condition across gender and emotions.

Emotions	Gender	Mean	SD	Emotions	Gender	Mean	SD
UCN	Female	173.98	31.48	CN	Female	144.66*	49.62
	Male	103.28	13.65		Male	97.09	20.79
UCH	Female	137.91	39.06	CH	Female	133.27	40.89
	Male	95.83	11.28		Male	101.04	27.28
UCS	Female	145.91	38.94	CS	Female	144.65	42.83
	Male	95.36	11.85		Male	94.05	11.88
UC Total	Female	152.60	39.27	C Total	Female	140.86	44.17
	Male	98.16	12.63		Male	97.39	20.80

UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

Table 8: *F value and significance of the overall effect of uncued and cued conditions across gender and emotions for lowest speaking fundamental frequency*

Lowest speaking fundamental frequency	F	Sig
Change across emotion	2.018	0.138
Change across gender	1.83	0.179
Overall change between uncued and cued conditions	2.37	0.126

Table 9: *Paired t test and significance for lowest speaking fundamental frequency between cued and uncued conditions across emotions and gender*

Lowest speaking fundamental frequency across emotions and trials	Gender	t (19)	Significance (2 tailed)
UCN – CN	Females	2.223	0.039*
	Males	1.272	0.219
UCH-CH	Females	0.402	0.692
	Males	0.882	0.389
UCS-CS	Females	0.087	0.932
	Males	0.341	0.737

Level of significance (*p<0.05); UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued.

d) Highest speaking fundamental frequency

The mean values of the change in highest speaking fundamental frequency in uncued and cued condition across gender and emotions are grouped in table-10.

Table 10: *Highest speaking fundamental frequency in uncued and cued condition across gender and emotions*

Emotions	Gender	Mean	SD	Emotions	Gender	Mean	SD
UCN	Female	287.56	34.52	CN	Female	324.14*	43.28
	Male	176.41	49.48		Male	189.03	53.69
UCH	Female	307.17	41.19	CH	Female	345.23*	57.66
	Male	208.40	55.81		Male	269.72*	75.82
UCS	Female	300.50	44.97	CS	Female	298.52	35.19
	Male	185.53	35.70		Male	180.61	37.63
UC Total	Female	298.41	40.60	C Total	Female	322.63	49.43
	Male	189.80	48.66		Male	212.16	69.32

UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

Table 11 portrays that there is statistically significant difference between the trials and emotions [$F(1,114)= 17.642, p<0.05$] and [$F(2,114)= 7.428, p<0.05$] respectively.

Table 11: *F value and significance of the overall effect of uncued and cued conditions across gender and emotions for highest speaking fundamental frequency*

Highest speaking fundamental frequency	F	Sig
Change across emotion	7.428	0.001*
Change across gender	0.011	0.915
Overall change between uncued and cued conditions	17.642	0.000*

Level of significance (*p<0.05)

Table 12: *Paired t test and significance for highest speaking fundamental frequency between cued and uncued conditions across emotions and gender*

Highest speaking fundamental frequency across emotions and trials	Gender	t (19)	Significance (2 tailed)
UCN – CN	Females	3.338	0.003*
	Males	0.890	0.384
UCH-CH	Females	3.354	0.003*
	Males	3.044	0.007*
UCS-CS	Females	0.227	0.823
	Males	0.345	0.734

Level of significance (*p<0.05); UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued.

In table 12, it can be viewed that happy emotion evoked the significant difference from uncued to cued condition in both females and males; a rise from 307.17 (SD=41.19) to 345.23 (SD=57.66) in females from uncued to cued condition respectively and 208.40(55.81) to 269.72(SD=75.82) in males. In the neutral emotion, only females showed a rise in highest speaking fundamental frequency from 287.56(SD=34.52) to 324.14(43.28). There was barely a change in males in neutral emotion across the trials ranging from 176.41(SD=49.48) in uncued condition to 189.03 (SD=53.69) in cued condition. In sad emotion, there was no significant difference between the uncued and cued trial across the genders. Females scored a value of 300.50 (SD=44.97) in uncued trial and it slightly decreased to 298.52 (SD=35.19) in cued trail which is not a significant change while males scored 185.53 (SD=35.70) in uncued trail which decreased to 180.61 (SD=37.63) in the cued trail.

As observed, the lowest speaking fundamental frequency did not change much across the emotions for the cognitive cued trial but the highest speaking fundamental frequency did have a significant change and this change was different with respect to emotions.

The difference between lowest and highest F0 (Range) was wider in females than males whose range was lesser. This shows the expressive nature of females. It also raises a significant issue of the cues that is used for eliciting the vocal changes. The sentence *'this is a pen'* with the situation of talking to a child elicited more pitch variations in the speech of the individuals while the following sentence *'the tests are in the cupboard'* with a situation of helping a friend find the test did not yield much variation. By manipulating the cognitive cues presented to the subjects, the vocal parameters could be varied to suit the variations desired by the experimenter. To further support this assumption, the study also proves a widened F0 range for happy emotion and a narrow range with no significant difference from the unemotional reading for sad emotion. The statement can be thus made that the subjects were influenced by the cognitive cues that were presented in the direction of the type of cues that were used.

e) Variation in fundamental frequency (vF0)

Table 13 presents the mean values of the variation in fundamental frequency (vF0) across the emotions.

Table 13: *Variation in fundamental frequency (vF0) in uncued and cued condition across gender and emotion.*

Emotions	Gender	Mean	S D	Emotions	Gender	Mean	S D
UCN	Female	0.11	0.037	CN	Female	0.18*	0.060
	Male	0.14	0.078		Male	0.16	0.052
UCH	Female	0.14	0.036	CH	Female	0.18*	0.045
	Male	0.17	0.076		Male	0.19	0.064
UCS	Female	0.15	0.080	CS	Female	0.14	0.054
	Male	0.15	0.078		Male	0.15	0.066
UC Total	Female	0.13	0.057	C Total	Female	0.17	0.055
	Male	0.15	0.079		Male	0.17	0.063

UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

Table 14: *F value and significance of the overall effect of uncued and cued conditions across gender and emotions for vF0*

vF0	F	Sig
Change across emotion	3.64	0.029*
Change across gender	1.00	0.319
Overall change between uncued and cued conditions	10.31	0.002*

Level of significance (*p<0.05)

As seen in table 14, there is an observable difference between the overall mean for uncued and cued condition but there is only a slight difference between the emotions and no difference in gender as per the repeated measure ANOVA

Table 15: *Paired t test and significance for vF0 between cued and uncued conditions across emotions and gender*

vF0 across emotions and trials	Gender	t (19)	Significance (2 tailed)
UCN – CN	Females	4.145	0.001*
	Males	1.042	0.310
UCH-CH	Females	4.797	0.000*
	Males	1.349	0.193
UCS-CS	Females	0.530	0.602
	Males	0.102	0.920

Level of significance (*p<0.05); UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

From table 15, it can be observed that there is only a significant difference in the values obtained for females in neutral and happy emotion and not in sad emotion. In neutral emotion, there was an increase in variability from 0.11 (SD=.037) to 0.18(SD=.060). In happy emotion, there was a slight raise from 0.14 (SD= .036) to 0.18 (SD=.045). But there is no significant difference in VF0 of males in all the three emotions.

Exploring this particular parameter, it can be noted that registers have played a role in the significant difference in variation that is observed in females as compared to males in the neutral emotion and sad emotion is known to have a very narrow range of change. However, there was a significant variation of F0 for females and a slight change in males for the emotion of happiness. This reason can be due to the individual differences in which the cognitive cues were perceived and to an extent, there is also an influence of the life experiences and the personality of the individual. The variability in voice was scattered in the study and when they are averaged, there was no significant difference across the trials especially in the case of males. Individual difference can be present on decoding of the cues that were presented.

f) Sentence duration

Table 16 depicts the mean sentence duration in uncued and cued condition across gender and emotions.

Table 16: *Sentence duration in uncued and cued condition across gender and emotions*

Emotions	Gender	Mean	SD	Emotions	Gender	Mean	SD
UCN	Female	1.50	0.444	CN	Female	1.46	0.321
	Male	1.35	0.495		Male	1.53	0.424
UCH	Female	2.30	0.230	CH	Female	2.26	0.260
	Male	2.09	0.255		Male	2.21	0.347
UCS	Female	1.87	0.359	CS	Female	2.10*	0.589
	Male	1.62	0.354		Male	1.89*	0.455
UC Total	Female	1.89	0.480	C Total	Female	1.94	0.539
	Male	1.69	0.485		Male	1.88	0.492

UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

Table 17 shows that there is a significant change in the overall scores across trials [$F(1, 114) = 10.85$, $p<0.05$] and across emotions [$F(2, 114) = 3.51$, $p<0.05$]. However, there is no significant change across gender [$F(1, 114) = 3.916$, $p>0.05$].

Table 17: *F value and significance of the overall effect of uncued and cued conditions across gender and emotions for sentence duration*

Sentence duration	F	Sig
Change across emotion	3.517	0.033
Change across gender	3.916	0.050
Overall change between uncued and cued conditions	10.85	0.001*

Level of significance (*p<0.05)

Table 18: *Paired t test and significance for sentence duration between cued and uncued conditions across emotions and gender*

Sentence duration across emotions and trials	Gender	t (19)	Significance (2 tailed)
UCN – CN	Females	0.642	0.529
	Males	1.364	0.188
UCH-CH	Females	0.814	0.426
	Males	1.254	0.225
UCS-CS	Females	3.015	0.007*
	Males	3.458	0.003*

Level of significance (*p<0.05); UCN= uncued neutral emotion; UCH=Uncued happy emotion; UCS=Uncued sad emotion; UC= Uncued; CN= Cued neutral emotion; CH= Cued happy emotion; CS= Cued sad emotion; C= Cued

Table 18 portrays that there is no significant difference between the sentence duration in neutral and happy condition for both males and females but there is a significant increase in duration for sad emotion for both females and males. Females had a mean duration of 1.87 (SD=0.359) in uncued condition which increased to 2.10 (SD=0.589) after the sentence was read after the presentation of the cognitive cue.

Males, at the same time, had a mean duration of 1.62 (SD=0.354) in the uncued trial and the duration of the sentence read increased to 1.89 (SD=0.0455) after the cognitive cue was presented.

As stated by literature, it can be noted in the present study that the duration did not vary significantly for neutral and happy emotion while there was a longer duration for the emotion of sadness. Murray and Arnott (1993) reported that the duration can be faster or even slower in happy emotion in comparison to the neutral emotion while it is slower for the sad emotion. Davitz (1964) and Fonagy (1981) reported a downward inflection, slurred enunciation and rhythm with irregular pauses which contributes to the longer time taken for this emotion. Williams & Steven (1972) also reported the increased duration resulted from longer vowels and consonants and from pauses that were often inserted within the sentence for sad emotion.

Phase II: Qualitative analysis

Five judges rated using a 3 point rating the read samples of the speakers on the parameters of speech, i.e., on parameters of pitch, loudness, quality, rate, effort, continuity, pronunciation, intelligibility, intonation and stress. Table 19 gives an overall view of the overall percentage given by the five judges for the samples on the 3 point rating scale.

Table 19: Percent scores of the judges' ratings for uncued and cued trials across emotions.

Emotion	Uncued condition	Cued condition
Neutral	1 = 8.10% 2 = 34.30% 3 = 57.60%	1 = 12.85% 2 = 31.95% 3 = 55.20%
Happy	1 = 13.65% 2 = 33.05% 3 = 53.25%	1 = 5.30% 2 = 26.00% 3 = 68.70%
Sad	1 = 7.40% 2 = 34.85% 3 =57.75%	1= 7.40% 2 = 31.8% 3 = 60.8%

It can be seen that for neutral emotion, the judges perceived the sentence read in the uncued trial a little more natural 57.6% of the times than the cued condition 55.2% of the times. This difference in the perception could be attributed to the cognitive cues that were provided. For the first sentence, '*This is a pen*', the cue given was that of '*how would you teach a child the concept of a pen*'. This cue elicited exaggerated pattern making the listeners perceive the sentence as unnatural. Happy emotion exhibited the greatest change in the perception from the uncued to the cued trial. 68.7% of the times the judges perceived the cues condition more natural than the uncued condition while only 53.25% of the times, the listeners perceived the uncued happy sentence as natural. There was also a rise in the percentage of perception of cued condition for sad emotion than in the uncued condition. The results obtained from the perceptual evaluation can be thus attributed to the fact that the provision of cognitive cues, indeed did improve the naturalness of the sentences that were read. The overall results indicated a variation across the conditions, in that, cued conditions were perceived to be most variable when compared to uncued conditions for the emotions happy and sad.

The results of the quantitative analysis revealed increased means values for most of the vocal parameters in cued conditions across all emotions. Further, the qualitative analysis also revealed that the judges perceived the sentences in cued conditions to be most variable and more natural than the uncued conditions. This indicates that variability in acoustic parameters was noticed both quantitatively as well as qualitatively across conditions (uncued and cued).

CHAPTER V

SUMMARY AND CONCLUSION

A mental image is an experience that, on most occasions, significantly resembles the experience of perceiving some object, event, or scene, but that occurs when the relevant object, event, or scene is not actually present to the senses. The use of cognitive cues as a task elicitation method for mental imagery stems from cognitive-behavioral therapy and focuses on prompting the individual to think about and feel a task prior to completing the task. The present study focused on expanding the knowledge of how useful cognitive cues are in eliciting vocal variability and further to assess the naturalness of the same.

Sentences were constructed for three emotions: neutral, happy and sad. Twenty participants (10 males and 10 females) in the age group of 20 to 25 years who were competent in English language were considered for this study. Each participant was asked to read the sentence three times. Trial 1 was the no cue condition. Trial 2 and 3 were the cognitive cued condition. All the read samples were recorded in a quiet environment. Only trials 1 and 3 were considered for acoustic analysis. The acoustic analyses involved quantitative analysis and the qualitative analysis. Six voice parameters were extracted using the software Real Time Pitch of Computerized Speech Lab 4500. Qualitative analysis involved rating the material constructed (appropriately juxtaposed uncued and cued samples) by five experienced judges on a three point rating scale.

The observed outcome of the study was that there was a definite effect of cognitive cueing on the vocal variability. Gender difference was also observed in the subjects owing to the known fact that females are more expressive than males. The quantitative analysis showed parallel changes with the emotions that were considered in the study. Happy emotion elicited a faster rate, had a higher F0 and a wider F0 range with upward inflections than neutral emotion. Comparing sad emotion with neutral emotion, the average F0 and F0 range was slightly lower with slower speech rate. It can thus be concluded that the subjects were influenced sufficiently by the cognitive cues to manifest their involvement in the emotions as as depicted in the overt, vocal expressions. As stated by Scherer (1986), vocal parameters are the most indicative of arousal. The results in general, revealed increased means values for most of the vocal parameters in cued conditions across all emotions. The qualitative analysis also revealed that the judges perceived the sentences in cued conditions to be most variable and more natural than the uncued conditions. This indicates that variability in acoustic parameters was noticed both quantitatively as well as qualitatively across conditions (uncued and cued). Hence, it could be concluded that cognitive cueing brought about changes in vocal attributes which could be quantified and perceived, strengthening the view that cognitive cues could help in achieving greater variations in voice and speech thereby, making the speech more natural.

The implications for these findings include utilizing cognitive cues

- To enhance performance in professional voice users by obtaining the desired change.
- For recovery of symptoms in clients with voice disorders and other speech disorders.
- The use of cognitive cues in place of a clinician model could be particularly beneficial for situations when the clinician and client are not the same age or gender. The use of cognitive cues as an assessment and therapy technique has the potential to enhance multiple disciplines and therefore should continue to be studied.

REFERENCES

Anderson, A. V. (1961). (2^{nd} Ed.). *Training the speaking voice.* New York: Oxford university press.

Andrews, M., Shrivastav, R., & Yamaguchi, H. (2000). The role of cognitive cueing in eliciting vocal variability. *Journal of Voice, 14*(4), 494-501.

Bennett, A. (1931), *Brainy Quote.* Retrieved May 27, 2010, from http://www.brainyquote.com/words/co/cognition145399.html

Bachorowski, J. A. & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychology Science, 6* (4), 219-224.

Bachorowski, J. A. & Owren, M. J. (2003). The sounds of emotion: Production and perception of affect-related vocal acoustics. *Annals of the New York Academy of Sciences, 1000,* 244–265.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614-636.

Barrett, L. F., & Niedenthal, P. M. (2004). Valence focus and the perception of facial affect. *Emotion, 4*(3), 266–274.

Bohnenkamp, T., Andrews, M., Shrivastav, R., & Summers, A. (2002). Changes in children's voices: the effect of cognitive cues. *Journal of Voice, 16* (4), 430-443.

Boone, D.R. (1991). *Is your voice telling on you? How to find and use your natural voice.* San Diego, California: Singular publishing Group Inc.

Campbell, N. (2000). *Databases of Emotional Speech.* In Cowie, R., Douglas-Cowie, E. & Schröder, M. (Eds.). Proceedings of the ISCA Workshop on Speech and Emotion. Belfast, Ireland.

Campbell, N (2001). Building a Corpus of Natural Speech – and Tools for the Processing of Expressive Speech – the JST CREST ESP Project. In Proceedings of Eurospeech 2001, Aalborg, Denmark.

Cicero, M. T. (1904). *De oratore* (2nd Ed.). Translated and with an introduction by E. N. P. Moore. London: Methaen.

Cosmides, L. (1983). Invariance in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 864-881.

Cowan, M. (1936). Pitch and Intensity characteristics of stage speech.. Arch. Speech, Suppl. To December issue.

Cowie, R., & Douglas-Cowie, E. (1996). *Automatic statistical analysis of the signal and prosodic signs of emotion in speech.* In: Proc. ICSLP'96, Philadelphia, pp. 1989-1992.

Darwin, C. (1872). *The expression of emotions in man and animals.* New York: Oxford University.

Davitz, J. R. (1964). *A review of research concerned with facial and vocal expressions of emotion.* The communication of Emotional Meaning, edited by J. R. Davitz. New York: McGraw – Hill. Pp 57 – 68.

Dobson, K. (2001). *Handbook of cognitive-behavioral therapies.* New York: The Guilford Press.

Dunbar-Wells, R. (1999). The relevance of metaphor to effective voice teaching strategies. *Australian Voice, 5,* pp 50-59,

Ekman, P. (1998). An argument for basic emotions. *Cognition and Emotion, 6,* 169-200.

Eldred, S. H., & Prince, D. B. (1958). A linguistic evaluation of feeling states in psychotherapy. *Psychiatry, 21,* 115-121.

Fairbanks, G., & Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the Expression of emotions. *Speech Monograph, 6,* 87 – 194.

Finke, R.A. (1989). *Principles of Mental Imagery.* Cambridge, MA: MIT Press.

Fónagy, I., & Magdics, K. (1963). Emotional patterns in Intonation and Music. Z. *Phonet. Sprachwiss. Kommunikationsforsch, 16,* 293-326.

Frolov, M. V., Milovañava, G. B., Lazarev, N. V., & Mekhedova, A. Y. (1999). Speech as an indicator of the mental status of operators and depressed patients. *Human Physiology. 25* (1), 42-47.

Green, R. S., & Cliff, N. (1975). Multidimensional comparison of structures of vocally and facially expressed emotion. *Perception Psychophysiology. 17,* 429-438.

Gross, J. J., John, O. P., & Richards, J. M. (2000). The dissociation of emotion expression from emotion experience: A personality perspective. *Personality and Social Psychology Bulletin, 26,* 712– 726.

Hargreaves, W. A., Starkweather, J. A., & Blacker, K. H. (1965). Voice quality in depression. *Journal of Abnormal Psychology, 70,* 218 – 220.

Havrdova, Z., & Moravek, M. (1979). Changes of the voice expression during suggestively influenced states of experiencing. *Activitas Nervosa Superior, 21,* 33-35.

Hicks, J. W. (1979). An acoustical/temporal analysis of emotional stress in speech. Dissertation Abstract. Intern. *41* (4-A).

Huttar, G. L. (1960). Uber Beziehungen von Sprachmelodie und Lautstarke [On relationships between speech melody and intensity]. *Phonetica 5,* 129-159.

Ingham, R. J., & Packman, A. C. (1978). Perceptual assessment of normalcy of speech following stuttering therapy. *Journal of Speech and Hearing Research, 21*, 63-73.

Izard, C. E. (1977). *Human Emotions.* New York: Plenum

Izdebski, K. (2008). *Emotions in Human Voice*. Volume 1. Plural publishing , Inc, San Diego.

James, W. (1884). What is an emotion? *Mind, 9,* 188-205.

Johannes, B., Petrovitsh Salnitski, V., Gunga, H. C., & Kirsch, K. (2000). Voice stress monitoring in space – possibilities and limits. *Avait. Space Environ. Md,* 71 (9, Section2, Suppl.). 58 - 65.

Kaiser, L. (1962). Communication of affects by single vowels. *Synthese, 14,* 300 – 319.

Keltner, D., & Gross, J. J. (1999). Functional accounts of emotions. *Cognition and Emotion, 13,* 467 – 480.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., & Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage. 36*, 889–900

Laukka, P. (2005). Categorical perception of vocal emotion expression, *Emotion, 5*, 277–295.

Laver, J. (1994). *Principles of phonetics.* Cambridge, UK: Cambridge University Press.

Lewis, M., Haviland-Jone, J. M., Barrett, F. (1998). Handbook of emotions (2nd Ed.). New York: Guilford Press.

Martin, S., & Darnley, L. (2004). *The teaching voice.* (2nd Ed.) . London: Whurr Publishers.

McKellar, P. (1957). Imagination and Thinking. London: Cohen & West.

Murray, I. R., Arnott, J. L., (1993). Towards the simulation of emotion in synthetic speech: A review of literature on human vocal emotion. *Journal of Acoustical Society of America, 93* (2), 1097-1108.

Niwa, S. (1971). Changes of voice characteristics in urgent situations. Reports of the *Aeromedical Laboratory, Japan Air Self - Defence Force, 11*, 264 – 251.

Oatley, K. (1989). The importance of being Emotional. *New Science*, 123 (Pt. 1678), 33 – 36.

Öster, A., & Risberg, A. (1986). The identification of the mood of a speaker by hearing impaired listeners. *Speech transmission Lab, Quarterly Progress. Stat. Rep.* 4/1986, 79 – 90.

Piaget, J. (1952). *The child's conception of number*. New York: Basic Books.

Reich, A., Mason, J., Frederickson, R., & Schlauch, R. (1989). Factors influencing fundamental frequency range estimates in children. *The Journal of Speech and Hearing Disorders*, *54*, 429-438.

Richardson, A. (1969). *Mental Imagery.* London: Routledge & Kegan Paul.

Roessler, R., & Lester, J. W. (1976). Voice predicts affect during psychotherapy. *Journal of Nerv. Mental Disease, 163,* 166-176.

Sanchez, J. A., Kirschning, I., Palacio, J. C., & Ostrovskaya, Y. (2005) Toward mood-oriented interfaces for synchronous interaction. CLIHC'05, 23-26, Cuernavaca, Mexico.

Scherer, K. R. (1981). *Speech and Emotional States.* In Darby, J. K. (Ed.) Speech Evaluation in Psychiatry. New York: Grune and Stratton.

Scherer, K. R. (1986). Vocal affect expression: A review and model for future research. *Psychological Bulletin, 99,* 143 -165.

Scherer, K. R. (1989). Vocal measurement of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience Vol. 4, The measurement of emotions* (pp. 233-259). New York: Academic Press.

Scherer, K. & Ceshi, G. (2000). Studying affective communication in the airport: The case of lost baggage claims. *Personality and Social Psychology Bulletin*, *26* (3).

Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychol. Med. 31*, 241–253.

Skinner, E. R., (1935). A calibrated recording and analysis of pitch, force and quality of vocal tones expressing happiness and sadness. *Speech Monograph. 2*, 81-137.

Stibbard, R. M. (2001). Vocal Expression of Emotions in Non-laboratory Speech: An Investigation of the Reading/Leeds Emotion in Speech Project Annotation Data. Unpublished PhD thesis. University of Reading,. UK.

Thomas, N.J.T. (1999) Are theories of imagery theories of imagination? An active perception approach to conscious mental content. *Cognitive Science, 23,* 207-245.

Tolkmitt, F. J., & Scherer, K. R. (1986). Effects of experimentally induced stress on vocal parameters. *Journal of experimental psychology: Human perception Performance, 12,* 302-313.

Voerman, M. S., Langeveld, A. P. M., & Van Rossum M. A. (2009). *The Journal of Laryngology & Otology. 123,* 528–534.

Williams, C. E., & Steven, K. N. (1969). On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Md.40,* 1369 – 1372.

Williams, C. E., & Steven, K. N. (1970). Acoustical manifestation of emotional speech. *Journal of Acoustical society of America, 47,* 66(A).

Williams, C. E., & Steven, K. N. (1972). Emotions and speech. *Journal of Acoustical society of America, 52* (4), 1238-1250

White, S. (1965). *Evidence for a hierarchical arrangement of learning processes*. In L. Lipsitt & C. Spiker (Eds.), Advances in child development and behavior. New York: Academic Press.

Wundt, W. (1924). *An introduction to psychology*. London: Allen & Unwin. (Orginal work published in 1912)

Zraick, R., Nelson, J., Montague, J., & Monoson, P. (2000). The effect of task on determination of maximum phonational frequency range. *Journal of Voice, 14,* 154-160.

APPENDIX I

Sentence 1: This is a pen (Neutral emotion).

Cue: Imagine you are teaching a child diagnosed with mental retardation. You are teaching him the concept of things we write with. You begin to teach him the concept of the pen. How would you teach him?

Sentence 2: The tests are in the cupboard (Neutral emotion).

Cue: Imagine you performed a test on a case which is not very familiar and had still to finish. Your friend takes up the case the next day but she too does not know much about the test and asks you where they are kept. What would you tell her?

Sentence 3: I got the highest IA in class (Happy emotion).

Cue: Imagine you could not study for well for a difficult paper. But to your amazement you stand first in scores you obtained for IA. You never expected it. How would you feel?

Sentence 4: The director announced a holiday today (Happy emotion).

Cue: Imagine you are very tired after the annual day as you have been working very hard. But you also know that you haven't studied for a test the next day. But to your surprise you wake up to hear that classes are suspended. How would you feel?

Sentence 5: He cannot walk anymore (Sad emotion).

Cue: Imagine you have a cousin you like so much. One fine day, you receive a call saying he met with an accident. You are unable to meet him but a few months later you see him in your institute unable to walk and run as he used to. How would you feel?

Sentence 6: The results are out and I failed (Sad emotion)

Cue: Imagine you worked very hard for your exams. But despite the effort you lose a paper. All your friends passed and are joyful. How would you feel?

APPENDIX II

Instructions: You will be presented two sentences, which are audio recorded one following the other. You have to listen to the recorded material and judge the naturalness between the sentences presented below. You are instructed to rate your perception on a 3 point rating scale where 1 is least and 3 is most. Given below are the tables to mark your ratings.

	(N) S1		(N) S2		(H) S1		(H) S2		(S) S1		(S) S2	
	a	b	b	a	b	a	a	b	a	b	b	a
Pitch												
Loudness												
Quality												
Rate												
Effort												
Continuity												
Pronunciation												
Intelligibility												
Intonation												
Stress												