

ACOUSTIC DIFFERENCES AMONG CASUAL, CONVERSATIONAL, AND READ
SPEECH

By

DeAnna Pinnow

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Communicative Sciences and Disorders-Master of Arts

2014

ABSTRACT

ACOUSTIC DIFFERENCES AMONG CASUAL, CONVERSATIONAL, AND READ SPEECH

By

DeAnna Pinnow

Speech is a complex behavior that allows speakers to use many variations to satisfy the demands connected with multiple speaking environments. Speech research typically obtains speech samples in a controlled laboratory setting using read material, yet anecdotal observations of such speech, particularly from talkers with a speech and language impairment, have identified a “performance” effect in the produced speech which masks the characteristics of impaired speech outside of the lab (Goberman, Recker, & Parveen, 2010). The aim of the current study was to investigate acoustic differences among laboratory read, laboratory conversational, and casual speech through well-defined speech tasks in the laboratory and in talkers’ natural environments. Eleven healthy research participants performed lab recording tasks (19 read sentences and a dialogue about their life) and collected natural-environment recordings of themselves over 3-day periods using portable recorders. Segments were analyzed for articulatory, voice, and prosodic acoustic characteristics using computer software and hand counting. The current study results indicate that lab-read speech was significantly different from casual speech: greater articulation range, improved voice quality measures, lower speech rate, and lower mean pitch. One implication of the results is that different laboratory techniques may be beneficial in obtaining speech samples that are more like casual speech, thus making it easier to correctly analyze abnormal speech characteristics with fewer errors.

ACKNOWLEDGMENTS

First, I would like to thank my thesis advisor Dr. Rahul Shrivastav for the support and encouragement throughout this project. He has been a great resource and motivating mentor throughout the past two year of this learning experience.

A special thanks goes to my thesis committee member Dr. Mark Skowronski who helped me work through some of the most important aspects of this study. He spent many hours with me explaining and helping me with formulating the results. Dr. Skowronski even spent Saturdays with me and my fellow colleague to help us with our studies and helped problem solve what we were having trouble with.

I thank Dr. Peter LaPine for his excitement about this study and being a part of my thesis committee. I appreciate his passion for the field of communication sciences and disorders and his willingness to help his students.

I acknowledge the time and effort each of my participants invested in this project and their willingness to help research in any way they can.

Most of all, I thank my family for their constant support and faith in me, especially in my moments of self-doubt. I am humbled and grateful for the great examples you all have shown me of how far hard work and passion for what I do can take me.

Last, but not the least I want to thank the challenging moments I had over the past two years throughout this project. I learned the most from what challenged me and I am thankful for the opportunity to learn and grown not only as a clinician but as a researcher as well.

TABLE OF CONTENTS

LIST OF TABLES.....	v
LIST OF FIGURES.....	vi
INTRODUCTION.....	1
LITERATURE REVIEW.....	2
<i>Clear Speech</i>	3
<i>Conversational Speech</i>	3
<i>Citation Speech</i>	3
<i>Hawthorne Effect</i>	3
<i>Casual Speech</i>	4
<i>Digitizing Sound for Acoustic Analysis</i>	4
<i>Spectral Analysis</i>	6
<i>Articulation Measures</i>	7
<i>Pitch</i>	8
<i>Rate of Speech</i>	9
<i>Current Study</i>	10
METHODS.....	12
<i>Participant Inclusion Criteria</i>	12
<i>Participant Exclusion Criteria</i>	12
<i>Recorder</i>	14
<i>Stimuli</i>	15
<i>Laboratory Read Recording</i>	15
<i>Laboratory Conversational Recording</i>	15
<i>Casual Speech Recording</i>	15
<i>Signal Editing and Acoustic Analysis</i>	16
<i>File Selection</i>	16
<i>Acoustic Analysis</i>	17
<i>Statistical Analysis</i>	18
RESULTS.....	19
<i>P0 (mean and std, ST)</i>	19
<i>CC SDS and DCC SDS</i>	24
<i>Analysis of Variance</i>	31
DISCUSSION AND CONCLUSIONS.....	40
APPENDIX.....	44
REFERENCES.....	46

LIST OF TABLES

Table 1: Participants, gender, age, number of clips for lab read, lab conversational, and casual speech.....	13
Table 2: Results from one way ANOVA comparing Rate of Speech for each recording environment.....	32
Table 3: Results from two way ANOVA comparing CC SDS, for each recording environment.....	33
Table 4: Results from two way ANOVA comparing DCC SDS for each recording environment	34
Table 5: Results from two way ANOVA comparing P0 (mean, ST) for each recording environment	36
Table 6: Results from two way ANOVA comparing P0 (std, ST) for each recording environment	37
Table 7: Results from two way ANOVA comparing PS (mean) for each recording environment	38
Table 8: Laboratory Read Sentences (16 spin sentences and 3 additional sentences).....	45

LIST OF FIGURES

Figure 1: Frequency in Semitones compared to Frequency in Hz	9
Figure 2: Female talker wearing the tie clip microphone plugged into the digital recorder.....	14
Figure 3: P0 (Mean, ST), for each environment.(Talkers 1-11).....	20
Figure 4: P0 (std, ST), for each environment (Talkers 1-11).....	21
Figure 5: PS (mean), for each environment (Talkers 1-11).....	22
Figure 6: CC and DCC, for each environment (Talkers 1-11)	24
Figure 7: Rate of Speech for each environment(Talkers 1-11).....	30
Figure 8: ANOVA Post Hoc Test Results for Rate of Speech in three recording environments.....	32
Figure 9: ANOVA Post Hoc Test Results for CC SDS in three recording environments.....	33
Figure 10: ANOVA Post Hoc Test Results for DCC SDS in three recording environments	34
Figure 11: ANOVA Post Hoc Test Results for P0 (mean, ST) in three recording environments	36
Figure 12: ANOVA Post Hoc Test Results for P0 (std, ST) in three recording environments.....	37
Figure 13: ANOVA Post Hoc Test Results for PS (mean) in three recording environments.....	38

INTRODUCTION

Different modes of speech have been well documented in research to further analyze various acoustic characteristics of normal and disordered talkers. Articulation occurs along a continuum of effort according to situational demands and informational load, ranging from low (hypo/conversational) to high (hyper/clear) articulation. There are two typical modes of speech documented in research literature (conversational and clear) that vary with different communicative and situational demands and which vary in internal effort (Lindblom, 1990). It is well known that speakers adjust their speech patterns to both production oriented and listener-oriented factors as demanded by the specific communicative situation (Moon and Lindblom, 1994). The probability of any one word being hypo-articulated (conversational) or hyper-articulated (clear) is thought to be related to its informational load (Lindblom, 1996); the less important the word is to a particular utterance's meaning and the more predictable it is from context, the more hypo-articulated it typically is. This dynamic process of determining which mode of speech to use in particular communicative situations as well as the acoustical differences between these modes is the topic of this research.

LITERATURE REVIEW

Clear Speech

Clear speech, also referred to as hyper-articulated speech, is commonly used when producing slow and exaggerated oral movements. The acoustical characteristics of clear speech are comprised of: longer production of short vowels and unvoiced fricatives (Picheny, 1986; Uchanski et al., 1996), higher F2 for front vowels, while a lower F2 for back vowels, significantly greater F1 range (Ferguson et al. 2007), decreased rate of speech, increased duration of phonemes, fuller differentiation between phonemes, and easily differentiated individual phonemes (Picheny 1986). Clear speech is adaptively used by speakers to enhance recognition and comprehension by the listener given the listening environment at the time (Lindblom, 1990; Moon and Lindblom, 1994; Lindblom, 1996), and can also be employed when the listener is unable to understand or has a hearing impairment. Women are usually more intelligible than men and clear speech is also directed at someone who the talker thinks has a hearing impairment (Bradlow, Toretta, & Pisoni 1996; Bradlow, Kraus, & Hayes, 2003; Goberman & Elmer, 2005; Ferguson & Kewley-Port, 2002; Pichney, Durlach, & Bradia, 1985; Schum, 1996). Interestingly clear speech is frequently used in human-computer interactions when automatic speech recognizers produce errors: speakers will shift to hyper-articulated speech which is helpful in human speech communication, but will often hinder the success of human-computer interaction given that automatic speech recognizers are not commonly trained in this style of speech (Hirschberg, Litman, & Swerts, 2004). Many studies have been done on clear speech, with an emphasis to understand what speech aspects enable talkers to be more intelligible than others.

Conversational Speech

Conversational speech or hypo-articulated speech is used in everyday conversations. It is a complex behavior that allows speakers to use many options to satisfy the social demands connected with spoken language. It is produced with lower effort by the talker and is generally characterized by lower intelligibility than clear speech. Most studies elicit conversational speech in a laboratory with highly structured speech tasks such as narratives, read speech, picture description, and asking specific questions.

Citation Speech

Another type of speech that has been used in previous experiments is citation speech, this is mainly described as read speech, and it is normally elicited in a reading task in a laboratory setting. Hyper-articulated or clear speech has been distinguished from citation speech in terms of greater vowel space expansion, decrease in speaking rate, increase in energy at the mid-frequency range, and a decrease in disfluencies (Harnsberger, Wright, & Pisoni, 2008).

Hawthorne Effect

Many studies that have analyzed clear and conversational speech elicited speech samples from participants within a laboratory setting. The limitation in obtaining conversational speech within a laboratory setting is that there are no controls for the Hawthorn Effect (Lansberger, 1959). This occurs when participants modify their performance simply by knowing they are being studied. This “performance” effect raises the question whether laboratory recording tasks elicit speech that is closely related to casual speech? The laboratory is not a natural setting for

the participant, thus recording speech outside of the laboratory may result in different speech characteristics than “conversational speech” produce in the laboratory. In other words, recording “conversation” speech within the laboratory may not control for the Hawthorne Effect, therefore, possibly biasing the results.

Since most experiments on clear and conversational speech have obtained speech samples inside of a laboratory one may raises questions such as: (a) Is laboratory speech the closest and best method of obtaining a speakers’ natural speech characteristics? (b) How accurate are speech diagnostic tools if they based off of laboratory recorded speech?

Casual Speech

The terms “conversational” and “clear” speech are well established in literature to reflect the “hypo-“ and “hyper-articulated” forms of speech respectively. Typically, both these forms of speech are elicited in a controlled laboratory condition. For the purpose of this research, a new term was required to describe speech obtained in a participant’s natural environment. For this study, the term “casual speech” is used to describe speech that is obtained outside a laboratory and during natural conversation. Casual speech is defined as informal speech that is used in everyday conversations.

Digitizing Sound for Acoustic Analysis

In order to understand acoustic measures and analysis, a brief overview of digitizing sound is given for readers not familiar with acoustic analysis. First sound is generated from a talker in the form of mechanical energy, is received by an acoustical receiver and is then converted to electrical energy through a transducer. In order to be stored digitally the sound

waveform is sampled and converted into a string of numbers during the conversion to electrical energy. There are several different formats for compressing and storing digital sound files. A commonly used format is the WAV format. WAV files are typically an uncompressed, discrete time series that represent the pressure wave of the sound(s) that were recorded. There are several parameters that characterize the WAV files. These are: sampling rate, Nyquist frequency, compression, channels, and amplitude resolution. Sampling rate is a term used frequently in acoustic analysis, it is defined as the number of samples per second taken from a continuous signal to make a discrete signal. The samples are uniformly spaced in time at a sampling period of $1/\text{frequency of sampling}$. The sampling rate sets the Nyquist rate, which is the highest frequency in the continuous-time signal retained in the discrete time signal after conversion to discrete time. Higher frequencies in the continuous-time signal are filtered out by the anti-aliasing filter built in to the analog-to-digital converter.

WAV files are sometimes compressed for data storage (less disk space) and transmission (less bandwidth). The two types of compression are data and amplitude compression. Data compression can either be lossy or lossless. Lossy compression is done by deleting information in the signal that is not perceived to be as important as other parts of the signal (i.e. MP3 files). For acoustic analysis lossless compression is used because files are stored in a format, which returns the exact same information as before compression. Amplitude compression is done to avoid peak clipping (hard clipping) which heavily distorts the signal, producing audible clipping artifacts that also interfere with conventional speech signal processing algorithms.

The number of channels for a WAV file refers to the number of microphones used for the recording. A two-channel WAV file means two microphones were used for recording and a one

channel WAV file means only one microphone was used. The number of microphones used for recording is determined by the acoustic application being used.

Amplitude resolution is the accuracy in representing the amplitude of the continuous-time signal with a finite number of bits in a digital signal. Amplitude resolution for WAV files is measured in bits, usually 8 bits or 16 bits. The more bits used for a recording, the higher is the amplitude resolution. If resolution is too low it can lead to audible "quantization" noise, which is not present in the continuous-time signal. In order to minimize quantization noise 16 bits are used for most audio applications. Once all these components are set the signal can then be stored and recreated for further analysis.

Spectral Analysis

Analyzing speech in a continuous way in order to study speech over time is called frame-based analysis. After a signal is stored digitally it is ready for spectral analysis. Intervals selected for analysis of a signal are called frames and the length (duration) is typically 20-30ms (Kent & Read, 2002). A frame interval is the amount of overlap between frames. Another function used with frame based analysis of a signal is waveform windows. Windows are a function applied to a waveform so that its amplitude is shaped in a particular way to enable the edges of a window to be minimized to analyze a representative part of the signal (Kent & Read, 2002). There are four commonly used windows in speech analysis: rectangular, Hamming, Hanning, and Blackman. Discrete Fourier Transform is used to translate the time domain values of a waveform in digital data into frequency domain values of a spectrum (Kent & Read, 2002). Fast Fourier Transform is a type of DFT that is used to create a Fourier spectrum and is the basis for many types of speech analysis (Kent & Read, 2002).

Articulation Measures

Articulation is a key characteristic of speech and can be a distinguishing factor of what type of speech (clear or conversational) is being produced. Acoustic measures of articulation evaluate the production of vowels and consonants that are manipulated by the oral structures of the vocal tract. In order to avoid subjectivity that can occur during perceptual experiments the current study used an articulation measurements computed from acoustic correlates of vocal tract configuration. Human factor cepstral coefficients (HFCC) and their differences were used for each speech sample of the current study (Skowronski & Harris, 2004). HFCC are acoustic correlates of vocal tract configuration and are originated from Mel frequency cepstral coefficients (MFCC) that are measured from a sound cepstrum. MFCCs have been used in many research studies and have been used in automatic speech recognition, speech synthesis, and speech coding due to their effective representation of the perceptual aspects of the vocal tract (Davis & Mermelstein, 1980). HFCCs have been shown to have improved noise robustness in automatic speech recognition over MFCCs and have also shown 92.8% accuracy in discriminating Parkinson's Disease speech from of healthy talkers (Skowronski, Shrivavstav, Harnsberger, Anand, & Rosenbek, 2012).

The HFCC standard deviation sum (CC SDS) is an acoustic correlate that represents the range of articulatory movement in a given speech sample. The delta HFCC standard deviation sum (DCC SDS) is the first-order temporal derivative (slope) of HFCC and may be described as the velocity coefficients. This measure reflects the rate at which the vocal tract changes from one frame to the next.

Pitch

Pitch is another speech characteristic that can be instrumental in determining what type of speech is being used. For the present experiment, pitch was determined through the Auditory-Sawtooth Waveform inspired Pitch Estimator (Auditory-SWIPE'). Auditory-SWIPE' measures the average distance between valleys and peaks at harmonic of pitch (Calderon, Alvarado, & Camacho, 2011). To estimate pitch, it measures the correlation of a signal to a sawtooth reference signal after accounting for human auditory system characteristics. Pitch is estimated in units of Hertz (Hz). Auditory-SWIPE' computes the mean and standard deviation (std) of the pitch of a signal, along with its pitch strength (PS).

The pitch measured in Hertz may be transformed to semitones (ST) units. The distribution of pitch in semitones is more normal compared to that in Hz, and the use of semitones allows for a more fair comparison of pitch between male and female talkers. This is because the transformation to ST reduces the upward skew of pitch measures in Hz. This makes it easier to compare the pitch of female and male speakers despite the fact that female voices are generally higher in pitch than males voices. The average pitch for males was 120Hz or 33 ST and average pitch for females was 210 Hz or 44 ST. The relationship between these two units (Hertz and Semitones) is shown in Figure 1. The calculation shown here is based on Hollien, Hollien, & Jong (1997).

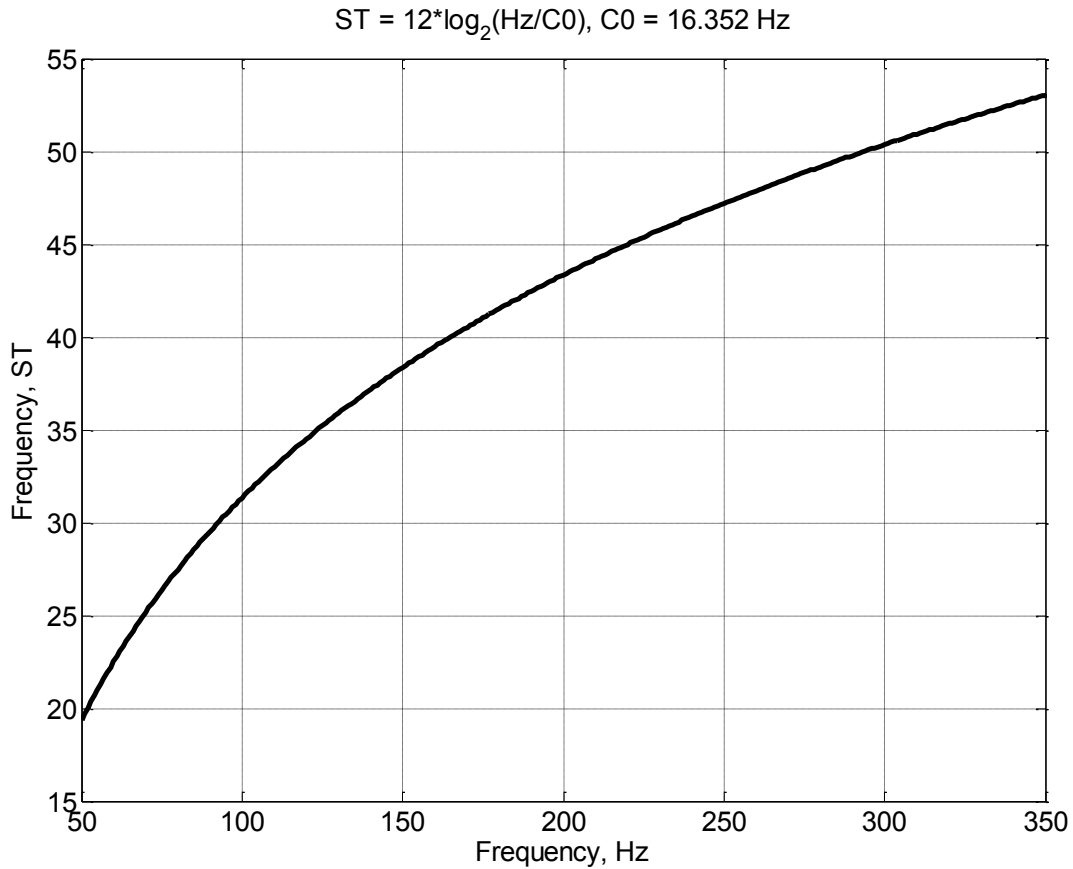


Figure 1: Frequency in Semitones compared to Frequency in Hz

Rate of Speech

Talkers change their rate of speech depending on what communicative task they are in. Rate of speech is a characteristic that has well documented variation between clear versus (slower) conversational speech (faster) (Picheny et al., 1989; Krause and Braida, 1995; Uchanski et al., 1996). Rate of speech is typically calculated as syllables per second or words per minute. These methods are usually done by manually counting each syllable in a spoken utterance. Due to the manual nature of this measurement, most studies report measures of inter- and intra-rater reliability to ensure accuracy and repeatability of this measurement.

Current Study

The goal of this experiment is to determine how speech varies across a continuum of speaking tasks that range from a highly controlled laboratory task to completely natural speech. Three different forms of speech are evaluated in this experiment. These include, the read speech (read sentences), laboratory conversational speech (dialogue), and casual speech (home recordings).

For this comparative study it is hypothesized that speech recorded outside the laboratory, when speakers are engaged in routine communicative tasks, will result in a better representation of casual (hypo-speech), than the conversational speech obtained in a speech laboratory and reported in most experiments. It is also hypothesized that laboratory speech includes a performance effect and a significant difference is evident in acoustic measures between speaking styles. The differences between such casual recordings and conversational speech elicited in the laboratory were quantified using select acoustic measures, including voicing measures [P0 (mean, ST and std, ST), and PS (mean)], articulatory measures [Human Factor Cepstral Coefficients (HFCC sum and standard deviation)], and temporal measures [Rate of Speech (syllables per second)].

Based on previous research that has reported clear (effortful) speech to have a slower rate and increased emphasis on words and sounds, it is predicted that casual speech will have a faster speaking rate, and lower PS relative to laboratory conversation speech and read speech. The following research questions were addressed in the current study:

- (a) Does casual speech and laboratory conversational speech differ in terms of their rate of speech, pitch, pitch strength, CC SDS, and DCC SDS?

(b) Does read speech and laboratory conversational speech in rate of speech, pitch, pitch strength, CC SDS, and DCC SDS?

The findings of this study could help develop better laboratory techniques (speech material, talking tasks, and acoustic measures) that are invariant between laboratory and casual speech. By investigating differences in the various forms of speech elicited in the laboratory and in natural environments, the current study could further aid researchers in obtaining speech that is the closest to participant's casual speech outside of the laboratory. Obtaining casual speech from a participant or patient may be beneficial to demonstrate speech characteristics that are the closest to their actual speech while minimizing differences resulting from a performance effect.

METHODS

Participant Inclusion Criteria

Two groups of speakers were recruited - young-adult group (18-30 years), and elderly group (61-92 years). All participants were native speakers of American English with normal hearing and no atypical speech characteristics (abnormal speech patterns were self-reported or screened in an informal interview). Participants were recruited through phone calls to a Parkinson's disease support group (spouse, partners and caretakers of patients of Parkinson's disease) and talking with colleagues. Participants were compensated \$75 for participation. The initial goal for the current study was to recruit an equal number of males and females for each participant group. However, time constraints prevented completion of the recruiting as planned and only a limited number speakers in the elderly group could be recruited for this study. Table 1 summarizes the participant data.

Participant Exclusion Criteria

Participants who were professional voice users (e.g. radio hosts, news anchors, actors, singers) and have atypical speech characteristics were not recruited. Since hearing loss can alter a participant's conversational speech patterns, participants were screened for hearing loss at 20 dB for young adults and 30-40 dB for older adults at frequencies of 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz.

Participant	Gender	Age	Group	Number of Audio Clips Lab Read Speech	Number of Audio Clips Lab Conversational Speech	Number of Audio Clips Casual Speech
S01	Male	26	Young Adult	19	34	88
S02	Female	24	Young Adult	20	24	39
S03	Female	25	Young Adult	14	40	74
S04	Female	26	Young Adult	12	51	17
S05	Male	67	Old Adult	19	59	74
S06	Female	25	Young Adult	8	27	16
S07	Female	24	Young Adult	14	10	60
S08	Male	25	Young Adult	19	47	145
S09	Male	65	Old Adult	19	52	57
S10	Male	85	Old Adult	16	80	26
S11	Male	25	Young Adult	14	61	42

Table 1. Participants, gender, age, number of clips for lab read, lab conversational, and casual speech

Recorder

All recordings were obtained using a portable handheld recorder with a tie clip microphone (Figure 2). The Olympus (WOS-2) was selected for the current study due to its portability, available recording formats, and large memory. An Olympus ME-15 tie clip (omnidirectional) microphone was used as the transducer and was selected for its small size that would be comfortable for wearing over prolonged periods of time, and not be as noticeable to the talker during everyday activities. A small number of samples were obtained and analyzed for overall quality, noise floor and ease-of-use in a pilot study. Recordings were saved in the PCM format (44.1kKz, sample resolution is 16bits).



Figure 2: Female talker wearing the tie clip microphone plugged into the digital recorder.

Stimuli

Three kinds of speech recording were obtained from the participants: Casual speech, Laboratory Read Speech, and Laboratory Conversational Speech.

Laboratory Read Speech

Laboratory read speech was obtained in the first test session for all speakers. Following informed consent and hearing screening, the investigator instructed the participant on the tasks and answered any questions the participant had. Then the participants were asked to read 16 low-predictability sentences from the Speech-Perception In Noise (SPIN) test (Kalikow et al., 1977). They also read three additional sentences (see Appendix 1). These recordings were called Laboratory Read Speech. Speakers were allowed to take any number of short breaks needed to prevent fatigue. The total recording time took a maximum of 5 minutes.

Laboratory Conversational Speech

Recording of laboratory conversational speech were done in a single test session with an appropriate number of breaks necessary to avoid fatigue. For the Dialogue task, participants were asked to speak on topics such as their hobbies, vacations, pets, etc. The investigator interacted with the participant during the dialogue task in order to elicit a conversational form of speech instead of a narrative. The total recording time for this task did not exceed 15 minutes.

Casual Speech Recording

Finally, participants were asked to wear a personal recorder with a microphone during daily activities for approximately 3 days. The participants were trained on how to use the

personal recorder and external microphone clip. The recorder training consisted of teaching how to turn the recorder on/off, saving the recordings to different folders within the recorder, and changing the batteries out of the recorder. Participants were allowed to turn the recorder off when they wished to have a private conversation, so as to decrease the chance of personal information being shared with the investigator.

Signal Editing and Acoustic Analyses

Acoustic analyses for laboratory recordings and home recordings was done using *Adobe Audition* and *MATLAB* (The Mathworks, Natick, MA, R2013b). Audition was used to snip the audio files to use for analysis. Recorded audio required channel separation before analysis could be completed. This was done using MATLAB. The files were named after the participants assigned number and recording condition (laboratory read, laboratory conversational, or casual speech) and the duration of the clipped segment.

File Selection

Only speech recordings from day 2 of the causal speech recordings were selected for this study. There were first screened using Adobe Audition to remove any segments that involved any conversational from were non-participants (conversational partners). This was done to ensure that non-participants were not analyzed without their consent or knowledge. To ensure optimal acoustical analysis, casual speech recordings were also screened for loud extraneous noises (e.g., music, appliances, animals, etc.). All sound clips that had clear recording quality and low amount of peak clipping were used for the study, irrespective of their content, perceived speaking style or conversational partner. Audio clips which demonstrated two or more instances

of peak clipping were not used for analysis. For laboratory read, laboratory conversational, and casual speech any speech clips that were less than one second were discarded from analysis. Once all suitable recordings were identified, the primary investigator ran initial analysis on each talker.

Acoustic Analysis

The clipped files were then used for the analysis for Pitch and Pitch strength (using Auditory-SWIPE'), CC SDS and DCC SDS, and rate of speech. All of these measures were computed using custom scripts written in MATLAB. The following equations were used to determine HFCC measures (CC SDS, and DCC SDS):

$$\text{CC SDS} = \sum_{k=1}^K \sigma_{\text{CC}}(k)$$

$$\text{DCC SDS} = \sum_{k=1}^K \sigma_{\text{DCC}}(k)$$

Where \sum is the sum, k is cepstral index, K is total number cepstral coefficients, and σ is the standard deviation of CC SDS or DCC SDS.

Pitch mean and standard deviation in semitones and pitch strength mean were computed using Auditory SWIPE'. Rate of speech (syllables per second) was calculated manually by counting the syllables and dividing them by the duration of the audio clip. Analysis was done using audio clips between 30 seconds to 60 seconds for each talker in each environment. Two raters estimated the number of syllables in a sampling of laboratory read, laboratory

conversational, and casual speech snippets. Files used for rate of speech consisted of the original large files before they were snipped for additional analysis.

Statistical Analysis

In order to determine if laboratory read and conversational speech was different from casual speech, a two-way ANOVA was performed for each acoustic measure (CC SDS, DCC SDS, P0 (mean, ST), P0 (std, ST), PS (mean)). The two dependent variables for the ANOVA were speaking environment (laboratory read speech, laboratory conversational speech, and casual speech) and talker. The acoustic measures being studied were the independent variables. The null hypothesis (H0) was: the means of the variables are the same for all speaking environments and talkers. The three alternative hypothesis (H1) were: (i) There is no difference in mean (for all acoustic measures) among recording environments, there is no difference in mean (for all acoustic measures) among talkers, and there is no difference among interactions of environment and talker. A p-value of 0.05 was used as a threshold of significance for the current study.

A second one-way ANOVA was calculated for rate of speech (syllables per second) to determine if there was a significant difference between the three recording environments (laboratory read, laboratory conversational, and casual speech). The null hypothesis (H0) was: the means of rate of speech are the same among different recording environments. The alternative hypothesis (H1): the means of rate of speech are different among recording environments. Speech rate was calculated from a subset of speech material for each talker (due to the labor-intensive nature of the measure), and as such data from each talker was not in sufficient quantity to treat talker as a factor in the ANOVA.

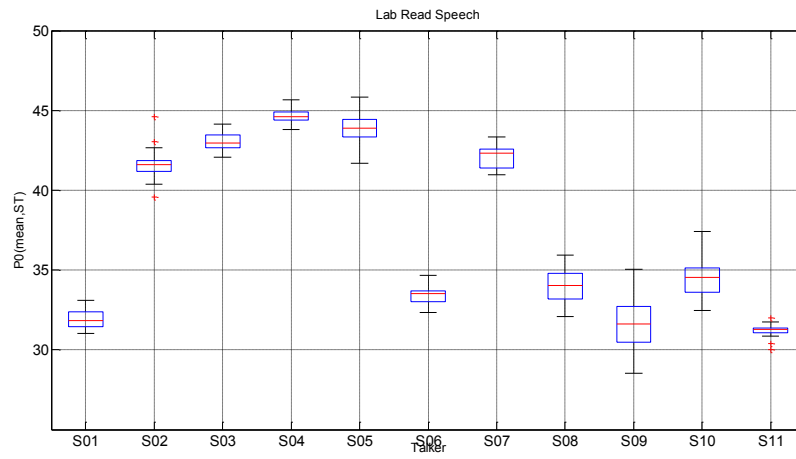
RESULTS

A total of 11 speakers were tested over a period of 4 months. Data obtained was edited and resulted in a total of 1,169 usable audio clips. The following figures describe the findings for the acoustic measures P0 (mean and std, ST), rate of speech (Syllables per second), CC SDS, and DCC SDS. It should be noted that talker 11 did not follow the instructions for home recording given by the primary investigator. Instead, Talker 11 recorded narratives for the recording. However, this data was not excluded from any of the analyses reported below.

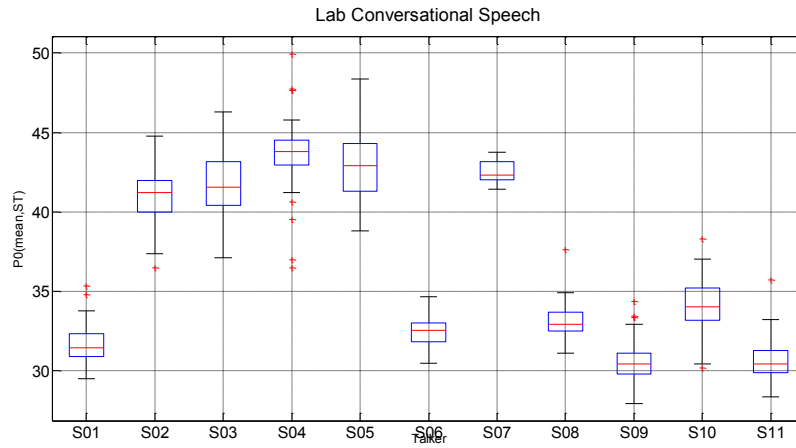
P0 (mean and std, ST)

Figures 3, 4, and 5 show box and whisker plots for P0 (mean and std) and PS (mean) for each talker in each environment, respectively. The box and whisker plots represent the distribution of data for each talker. The box in the plot is comprised of the median (represented by a horizontal red line), the 25th percentile (the bottom blue line), and the 75th percentile (the upper blue line of the box) of the data from each speaker. The two lines that extend from it represents the highest and lowest numbers without being an outlier (maximum and minimum observations). The box plots were used to show the distribution more evenly for each talker. Outliers, data points that are located outside of the whiskers of the plot (1.5 times the interquartile range above the upper quartile and below the lower quartile) are represented by a red cross.

(a.)



(b.)



(c.)

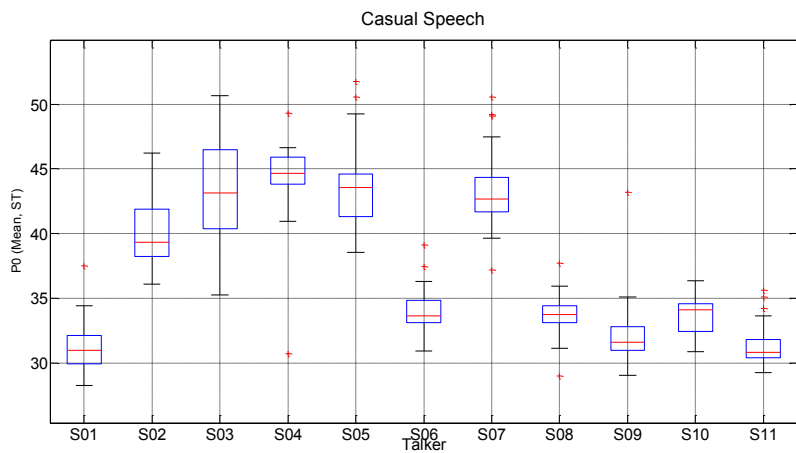
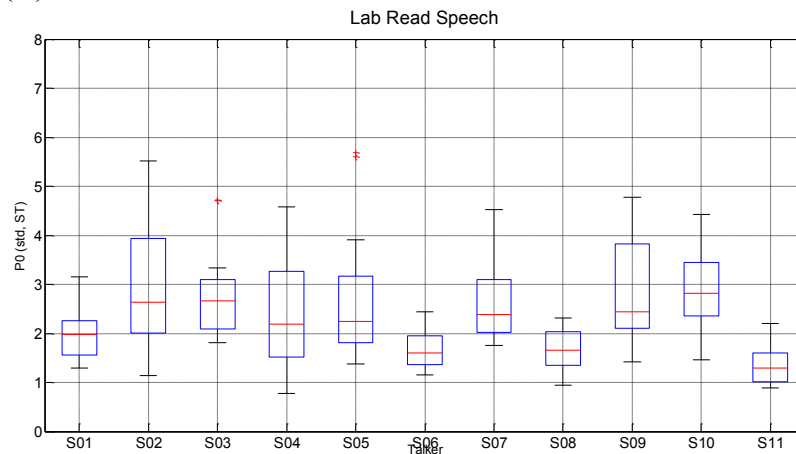


Figure 3: P0 (Mean, ST), for each environment. (Talkers 1-11).

In Figure 3, P0 (mean, ST) had greater variation in distribution for casual speech when compared to laboratory read and laboratory conversational speech. Laboratory read speech did not show as much variation as laboratory conversational or casual speech for P0 (std, ST). Overall P0 (mean, ST) was lower for both laboratory read and laboratory conversational than casual speech. Casual speech showed more outliers than laboratory read and laboratory conversational speech, thus further demonstrating the wider variation in P0 (mean, ST) in natural recording environments.

(a.)



(b.)

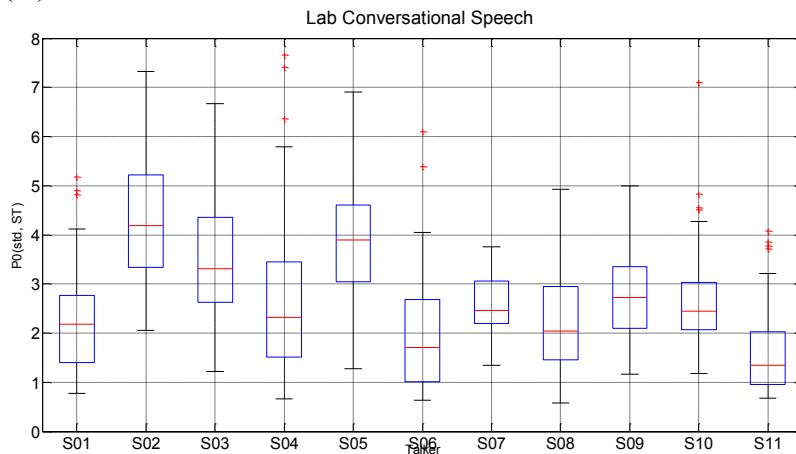


Figure 4: P0 (std, ST), for each environment. (Talkers 1-11). (Figure 4 continued before figure on page 22)

(Figure 4 cont'd)
(c.)

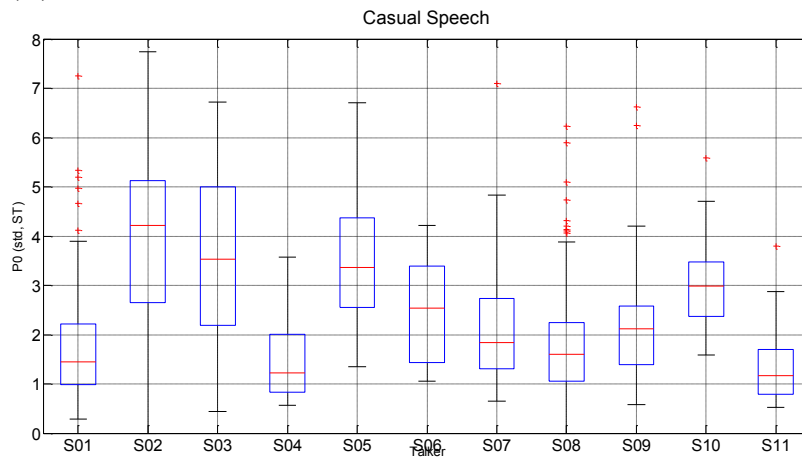


Figure 4 shows box-and-whisker plots for P0 STD in ST for each talker. Laboratory read speech had the lowest ranges (median, 75th percentiles, and maximum observations) of P0 (std, ST) for each talker. Laboratory conversational speech for each talker showed a higher range of median, 75th percentile, and maximum observations for P0 (std, ST). Casual speech showed the highest values compared to laboratory recorded speech for median, 75th percentile, maximum observations, and outliers for each talker.

(a.)

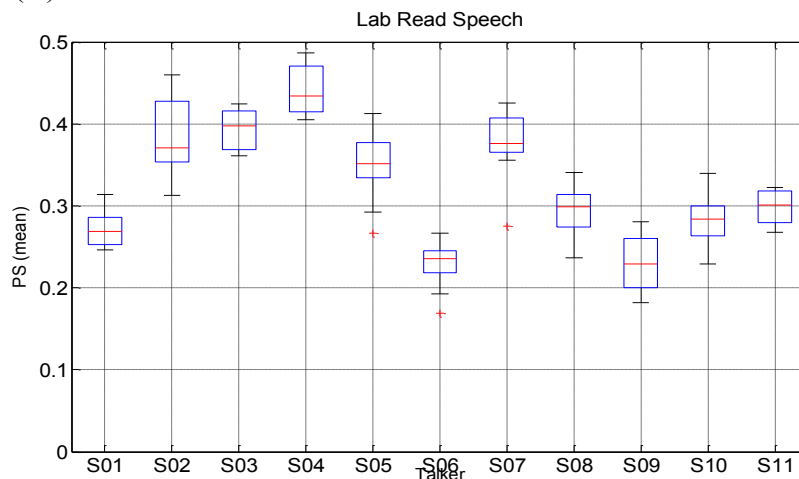
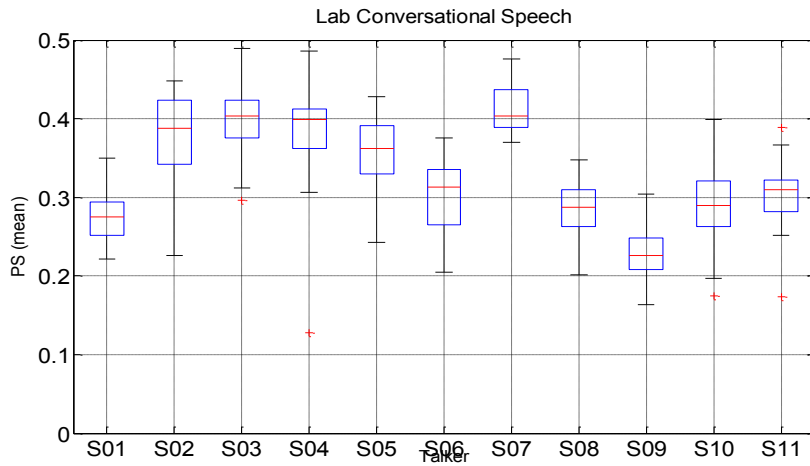


Figure 5: PS (mean), for each environment. (Talkers 1-11) (Figure 5 continued before figure on page 24)

(Figure 5 cont'd)
(b.)



(c.)

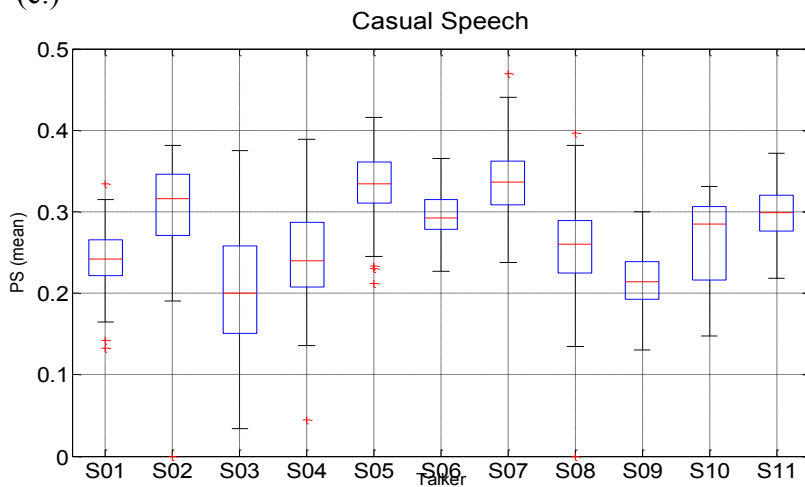


Figure 5 shows box-and-whisker plots for PS mean for each talker. PS mean was calculated only for analysis frames with PS above the voiced-speech detection threshold of 0.1. Overall PS (mean) values were higher for laboratory recorded speech than casual speech. For both laboratory read and laboratory conversational speech the median, 75th percentile, 25th percentile, minimal and maximum observation values showed a smaller range compared to casual speech. Casual speech PS(mean) values were overall lower than laboratory recorded speech.

CC SDS and DCC SDS

Scatter plots were created for each talker in each environment to demonstrate the change in CC SDS and DCC SDS among all three recording environments and to show any trends. The number of samples per talker varied, so scatter plots were done for each individual talker and not the entire group of participants in order to show a clear representation of the changes that occurred among recording environments.

(a.)

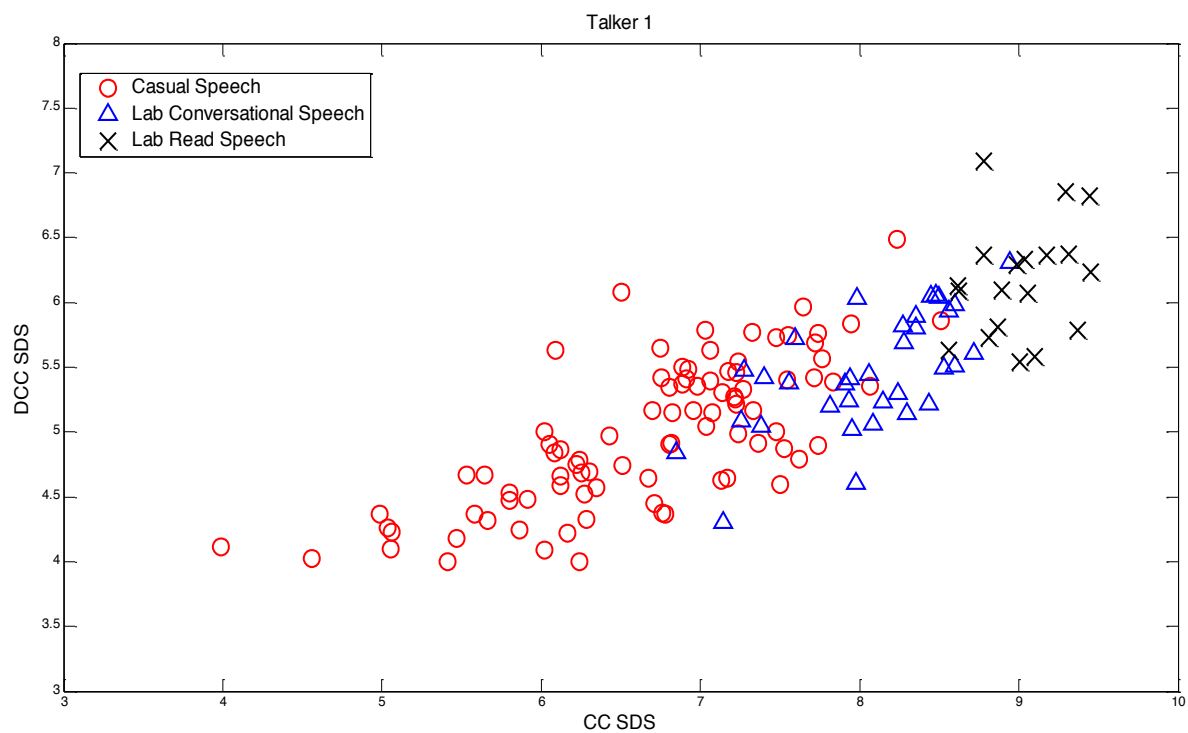
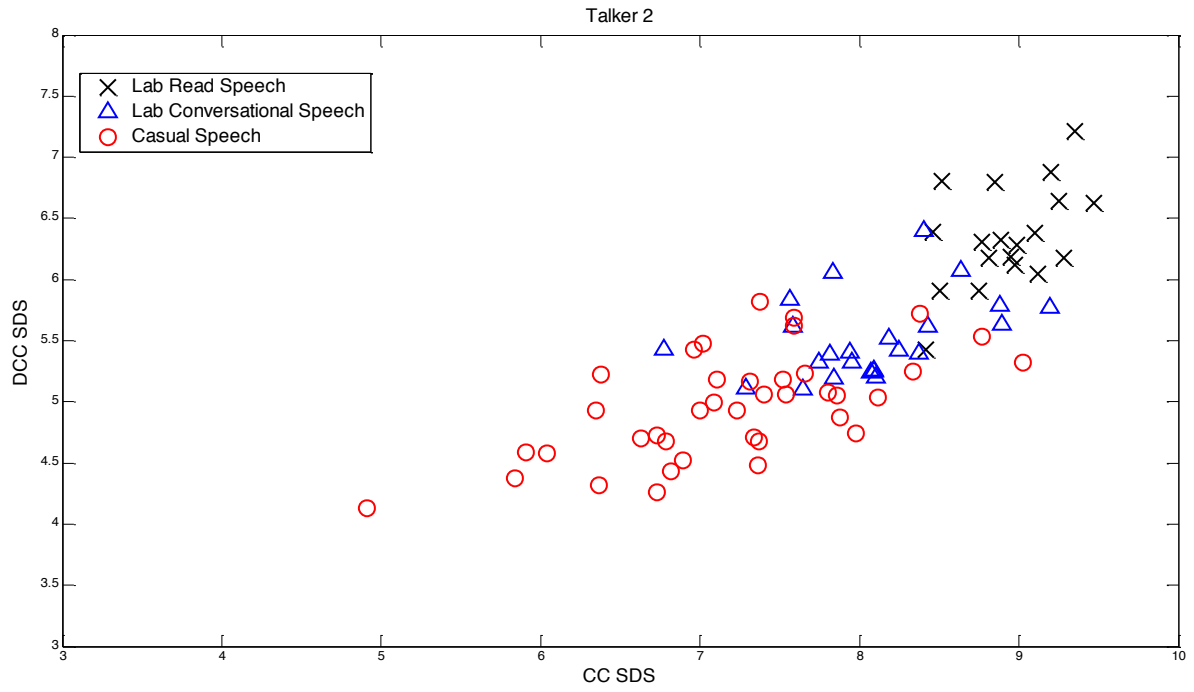


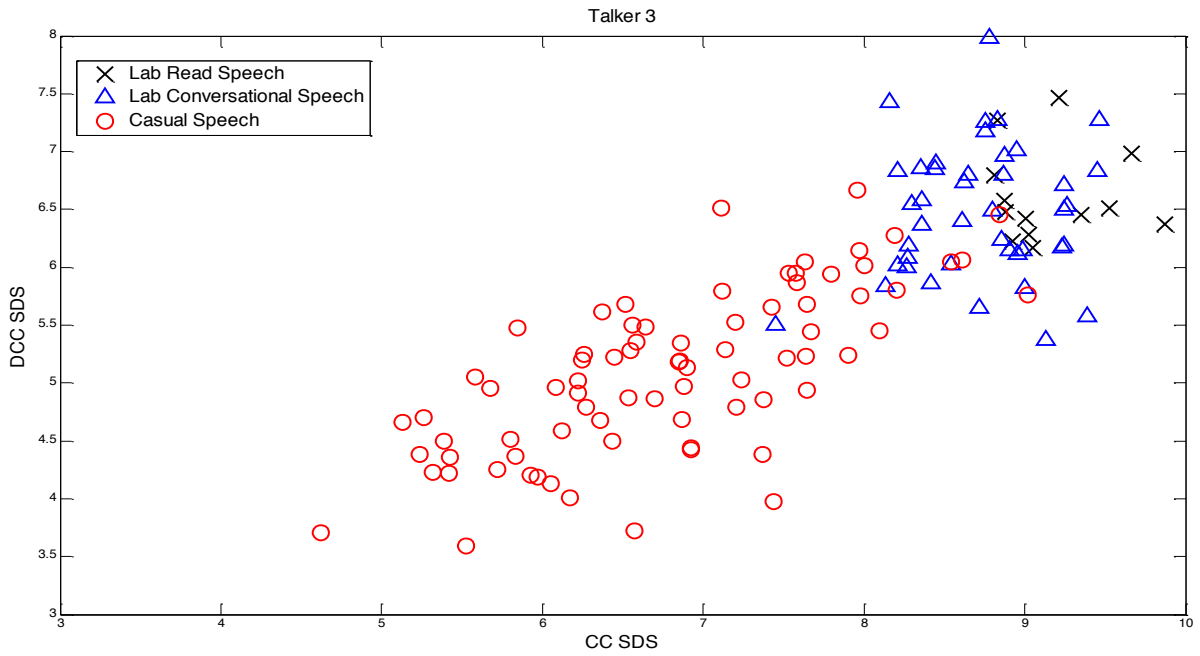
Figure 6 (a-k): CC and DCC for each environment. Talkers 1-11. (Figure 6 continued until figure on page 30)

(Figure 6 cont'd)

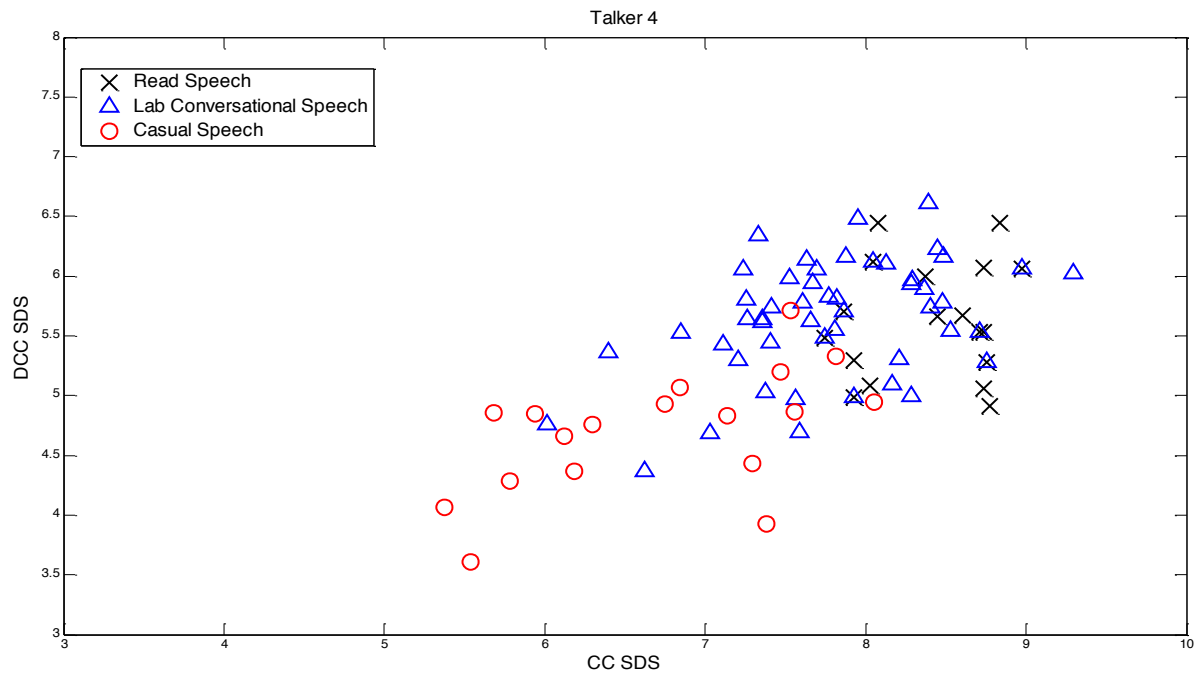
(b.)



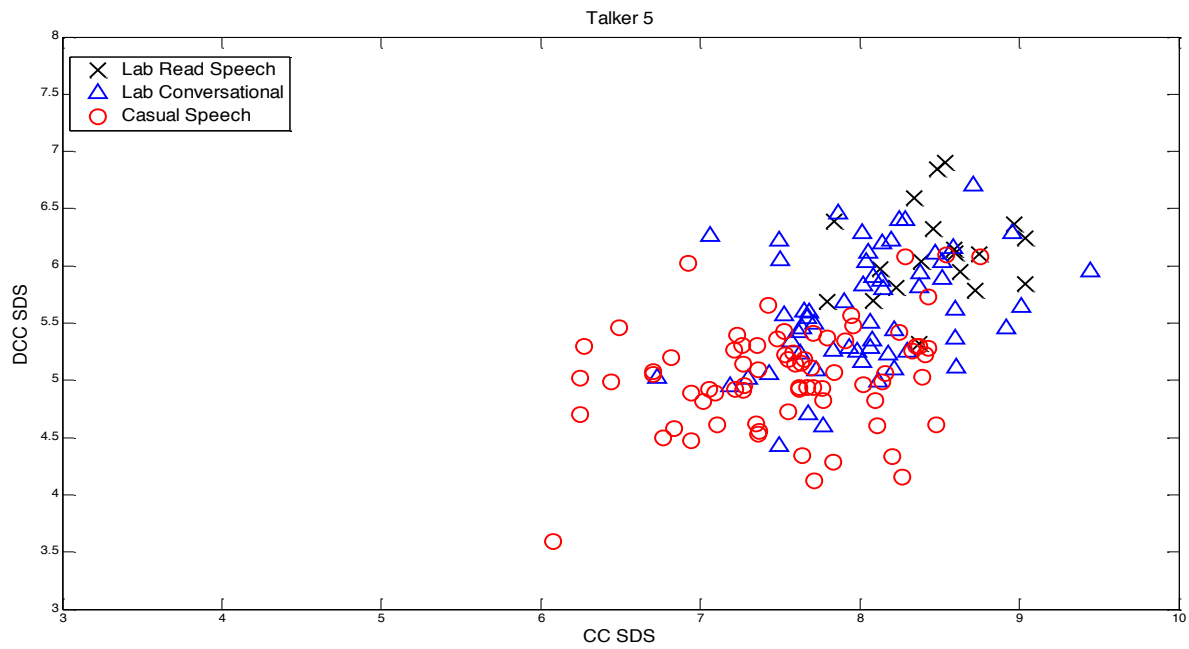
(c.)



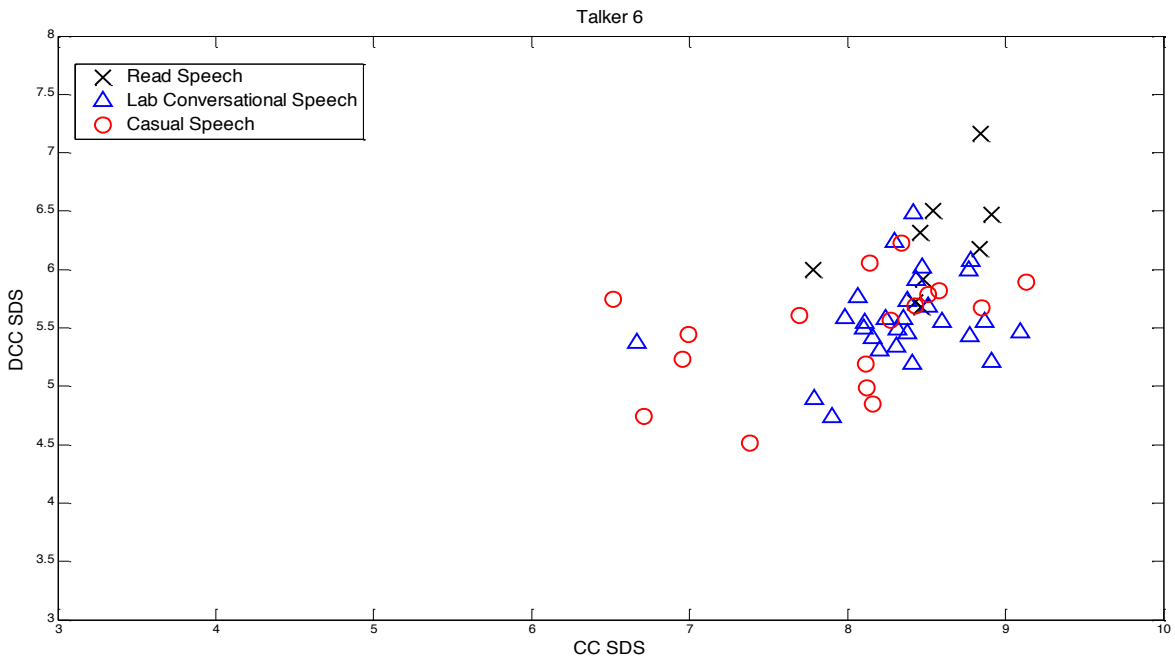
(Figure 6 cont'd)
(d.)



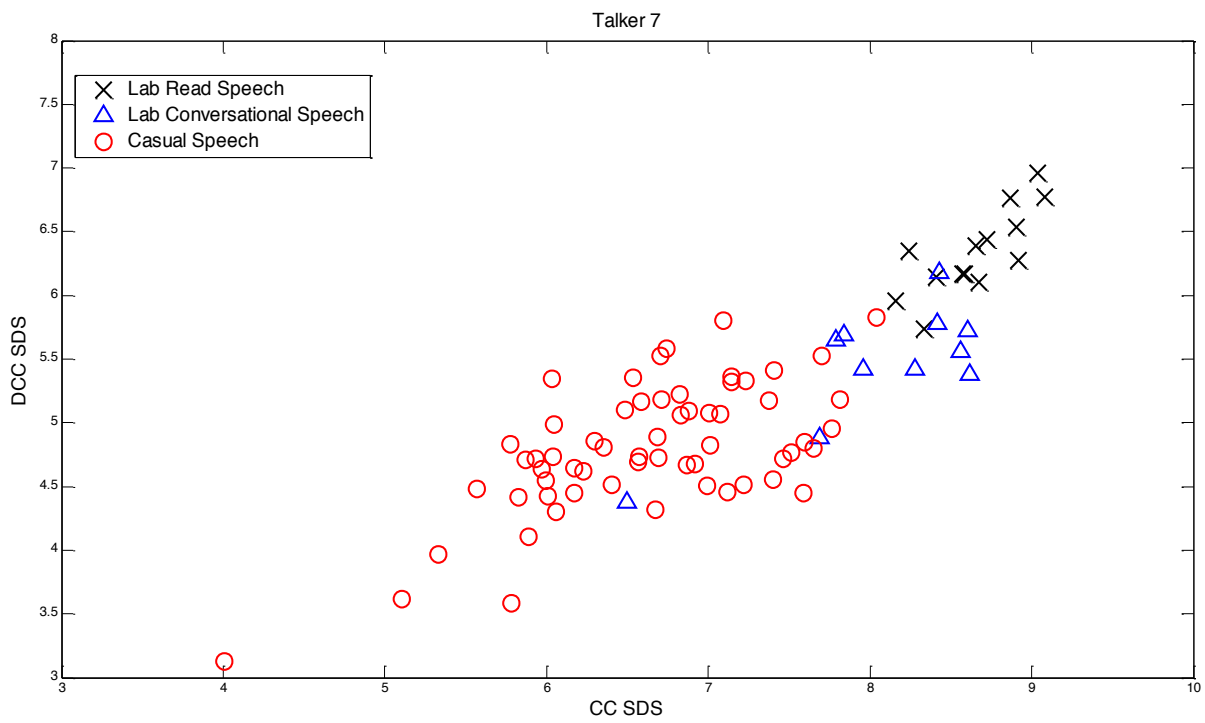
(e.)



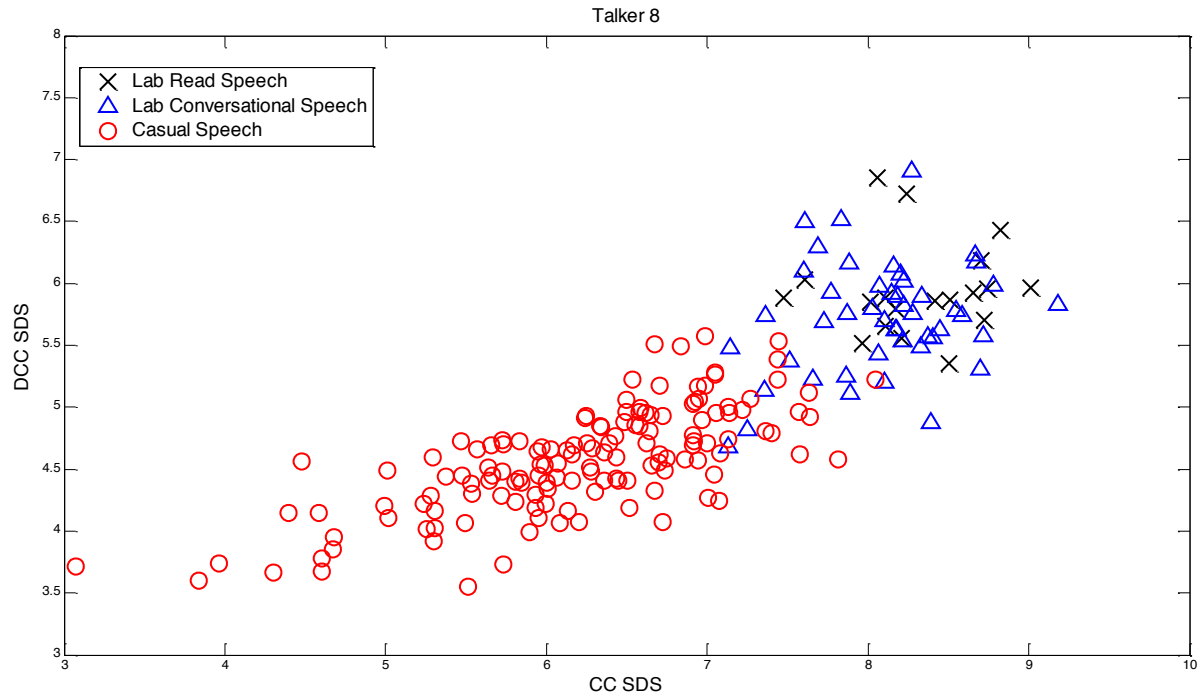
(Figure 6 cont'd)
(f.)



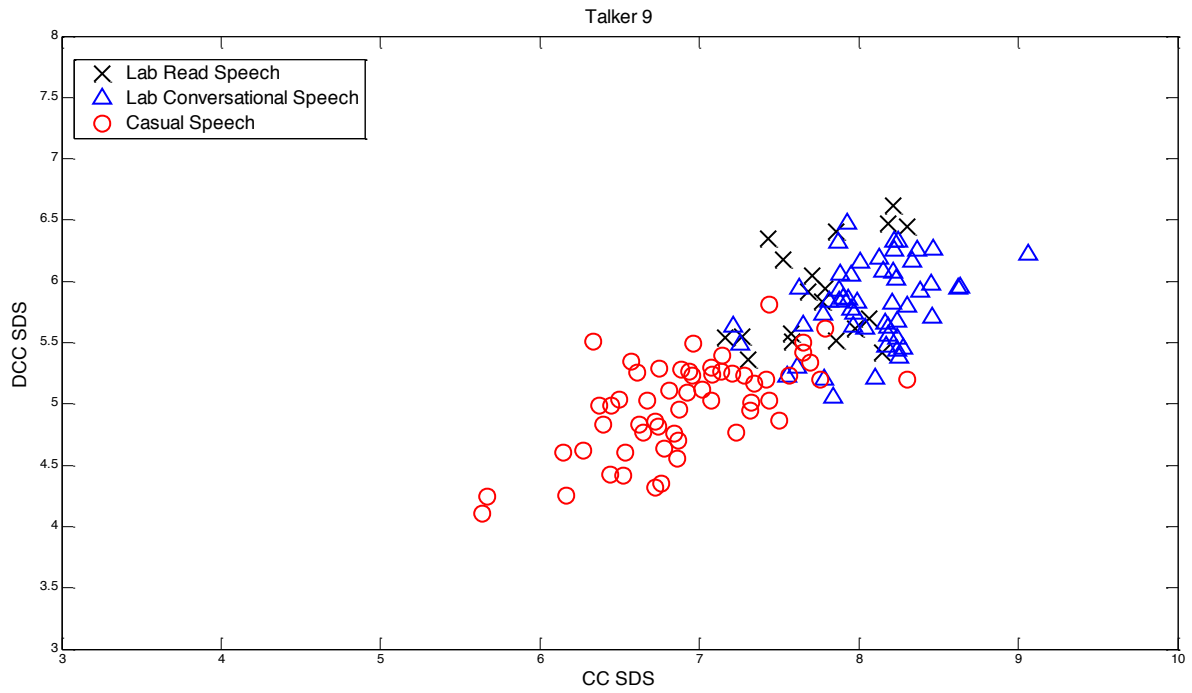
(g.)



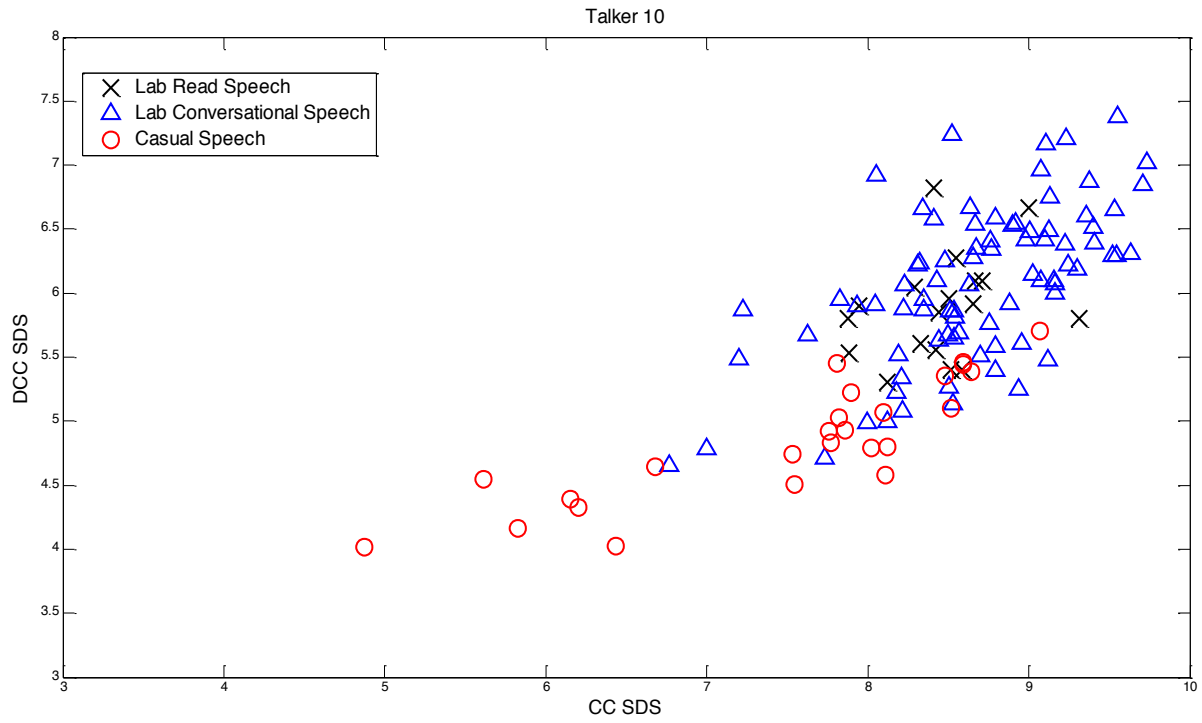
(Figure 6 cont'd)
(h.)



(i.)



(Figure 6 cont'd)
(j.)



(k.)

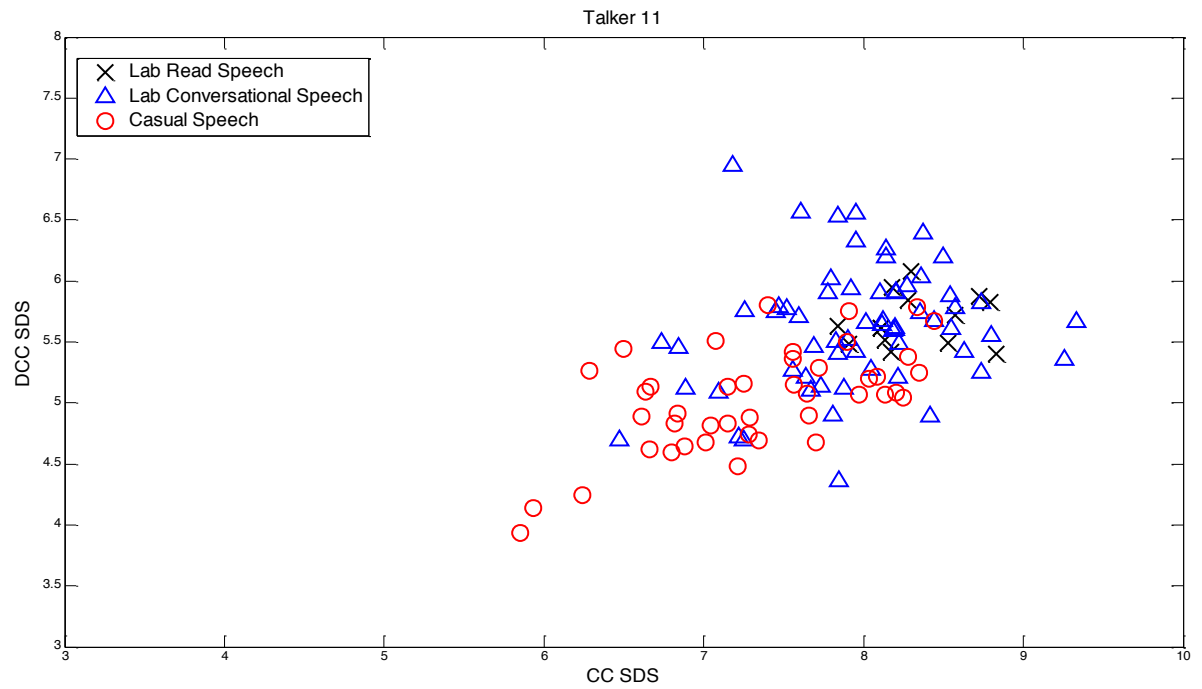


Figure 6 shows a scatter plot of CC SDS vs. DCC SDS for each talker in the three speech environments. For all talkers, the following trends are observed: articulation range (CC SDS) was highest for laboratory read speech and lowest for casual speech. The range of articulation range (CC SDS) was lowest for laboratory read speech and highest for casual speech. Overall DCC SDS was between 15%-20% smaller for laboratory read speech compared to casual speech. A wide variation for DCC SDS was also observed across different talkers. A large articulation range means that speech articulators move a large distance from their rest positions during speech production. Laboratory read speech had the lowest range of CC SDS and DCC SDS and fell within the similar parameters for each talker. For all three speaking styles laboratory read speech showed the lowest range. Laboratory conversational speech for each talker also fell within similar parameter for CC SDS and DCC SDS.

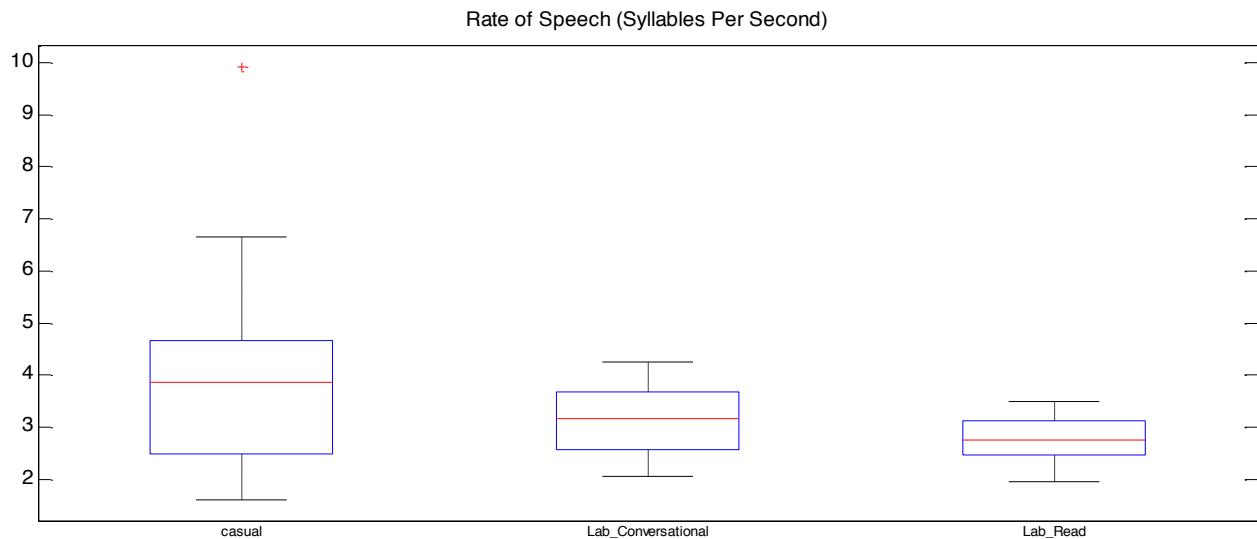


Figure 7: Rate of Speech for each environment. (Talkers1-11)

The rate of speech (in syllables/second) was measured independently by two judges. Correlation of speech rate between the two raters was high (Pearson $r = 0.99$, $p < 0.05$, $N = 59$). Figure 7 shows a box-and-whisker plot for speech rate for each environment. Casual speech

(mean = 3.8 syllable/sec) demonstrated an overall higher rate of speech than laboratory conversational (mean = 3.1 syllable/sec) and laboratory read (mean = 2.7 syllables/sec) speech.

Analysis of Variance

The acoustic variables were analyzed for statistical significant differences using a 1-way ANOVA or a 2-way ANOVA for the factors of Environment and Talker. Degrees of freedom, F statistic, and p-value are reported in Tables 2-7. The degrees of freedom represent the number of variables for each dependent variable minus one. F-statistic is a ratio of between-group variance to within-group variance (or, between-level variance / within-level variance). The between-level variance refers to difference in means among levels of a factor, while within-level variance refers to variance of each level of a factor. The p-value was set at a threshold of 0.05. A post-hoc test using Tukey-Kramer honestly significant difference was used to illustrate the differences among recording environments; confidence intervals were employed to compare means for each factor. Post hoc test were completed for each variable comparing all three recording environments. Post hoc tests were not completed for each individual talker as this was not the focus of the current study and time constraints prevented a detailed analysis of differences across different talkers or talker-environment interaction effects.

<i>Source</i>	<i>d.f.</i>	<i>F</i>	<i>Prob>F</i>
Environment	2	5.71	0.0046
Error	96		
Total	98		

Table 2: Results from one way ANOVA comparing Rate of Speech for each recording environment.

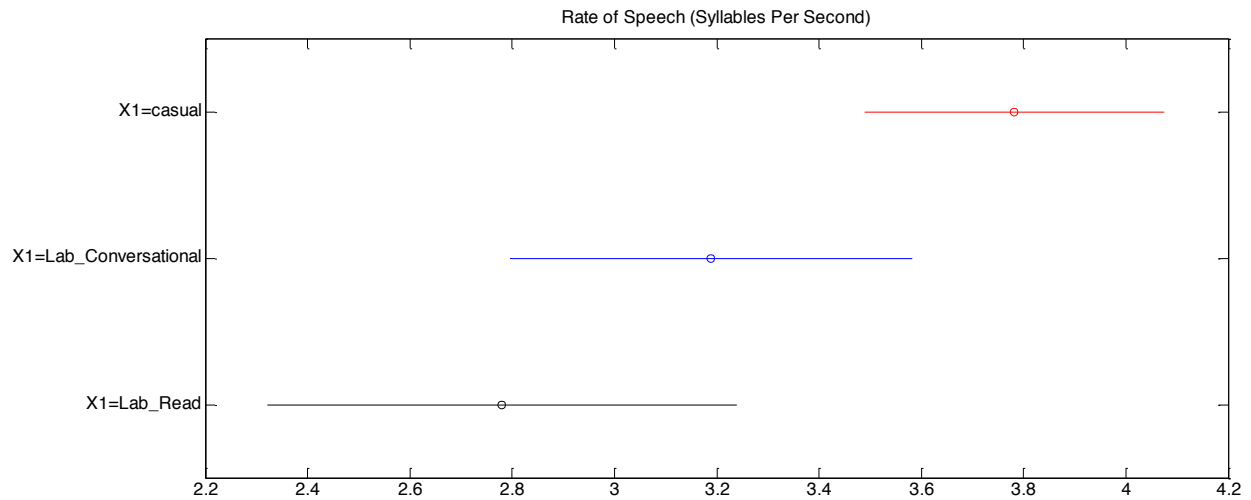


Figure 8: ANOVA Post hoc test results for rate of speech in three recording environments

Table 2 summarizes the 1-way ANOVA results for speech rate for the Environment factor. An unbalanced ANOVA was conducted due to the different number of samples in each environment. Speech rate was calculated manually from a subset of speech material for each talker (due to the labor-intensive nature of the measure), and resulted in a different sample size than the other measures which were all computed automatically. For this reason, speech rate data from each talker was not included as a factor in the 2-way ANOVA. Instead, a separate ANOVA was completed to determine whether speaking rate varied across the three recording environments.

The interval bars represent the Tukey Kramer honestly significant difference, and are the range in which the true mean exists. The circle represents the sample mean of the data. The

ANOVA (table 2) shows that rate of speech was significantly different among environments ($F=5.71$, $p = 0.0046$). According to the post hoc test (figure 8) the mean speech rate was highest for casual speech (3.8) followed by laboratory conversational (3.2) and laboratory read (2.7). The post-hoc test results reveal that speaking rate changes significantly, ($p<0.05$) 31% from laboratory read speech to casual speech, and ($p<0.05$) 17% from laboratory conversational speech to laboratory read speech. Speaking rate changed by ($p>0.05$) 7% from laboratory conversational to casual speech.

<i>Source</i>	<i>d.f.</i>	<i>F</i>	<i>Prob>F</i>
Environment	2	421.67	$p<0.0001$
Talker	10	12.58	$p<0.0001$
Interaction	20	11.35	$p<0.0001$

Table 3: Results from two way ANOVA comparing CC SDS for each recording environment.

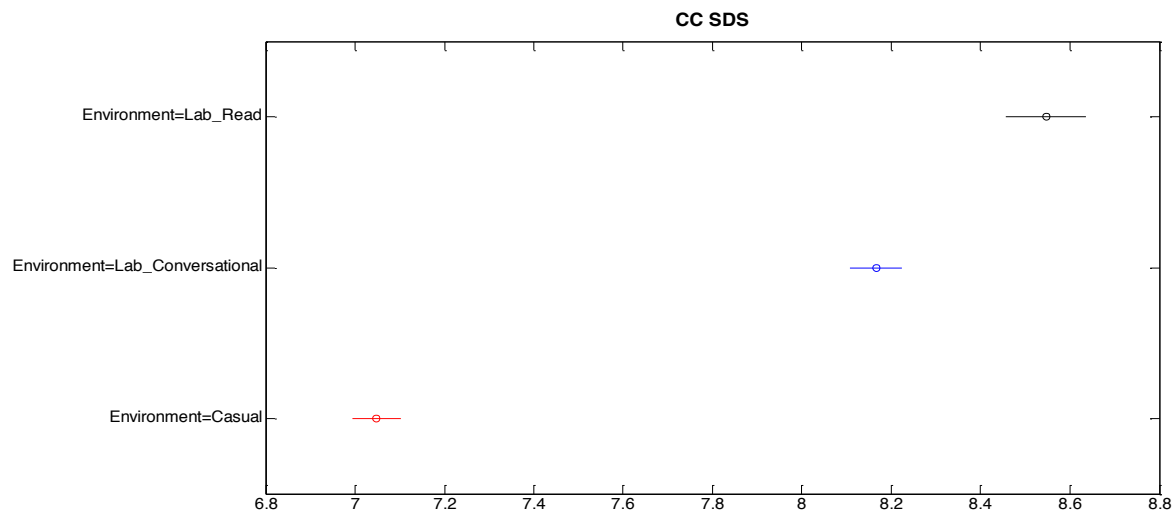


Figure 9: ANOVA Post Hoc Test Results for CC SDS in three recording environments

<i>Source</i>	<i>d.f.</i>	<i>F</i>	<i>Prob>F</i>
Environment	2	419.12	p<0.0001
Talker	10	13.02	p<0.0001
Interaction	20	8.56	p<0.0001

Table 4: Results from two way ANOVA comparing DCC SDS for each recording environment.

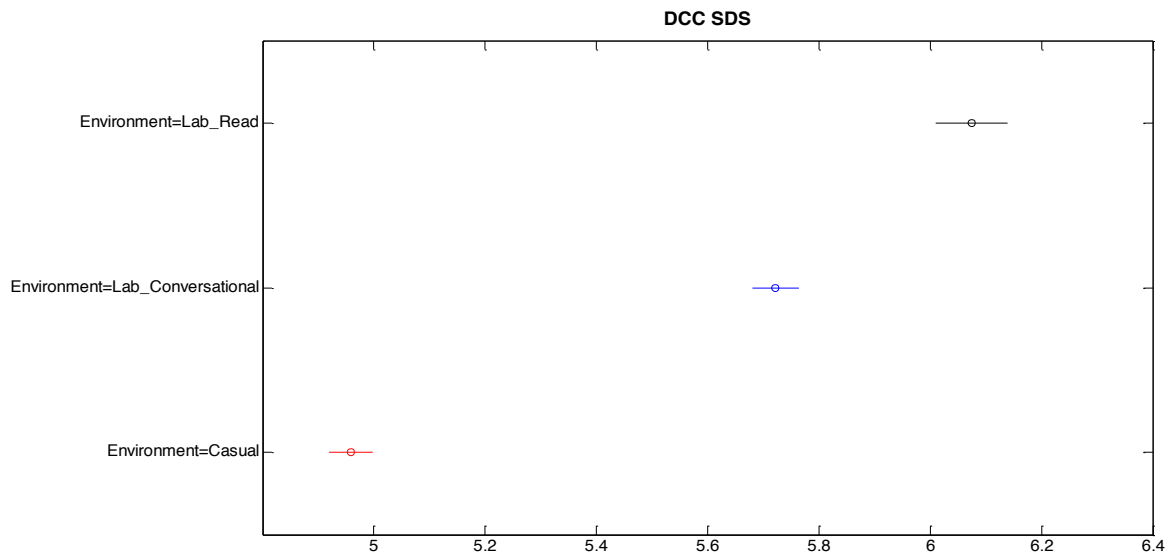


Figure 10: ANOVA Post Hoc Test Results for DCC SDS in three recording environments

Tables 3 and 4 summarize two-way ANOVAs using environment and talker as factors and interaction for both CC SDS and DCC SDS. Both tables show a p-value <0.001 for environment, talker, and interaction meaning there is a significant statistical difference in CC SDS and DCC SDS across the two dependent variables. The interaction effect for both tables (p<0.001) showed a statistical difference as well, meaning different talkers demonstrate different forms of speech in different environments. The ANOVA Table 3 shows that CC SDS was statistically significant among the different speaking environments (F= 421.67, p<0.0001). Figure 9 (ANOVA Post Hoc Test Results) and Table 3 illustrate a significant difference (p<0.0001) in CC SDS for each recording environment. The highest mean value for CC SDS was laboratory

read speech (8.5), followed by laboratory conversational speech (8.2), and the lowest mean value for CC SDS was casual speech (7.1). Results from the post-hoc test show that CC SDS changes significantly (21%) ($p < 0.0001$) from laboratory read speech to casual speech, and 14% ($p < 0.0001$) from laboratory conversational speech to casual speech. CC SDS changed by 4% ($p < 0.05$) from laboratory read speech to laboratory conversational speech. Similarly in figure 10 (ANOVA Post Hoc Test Results) and table 4 a significant difference is shown for DCC SDS ($F=419.12$, $p < 0.0001$) among recording environments. The highest mean value for DCC SDS was laboratory read speech (6.2), followed by laboratory conversational speech (5.7), and the lowest mean for DCC SDS was casual speech (4.8). Results from figure 10 reveal that CC SDS changes significantly (25%) ($p < 0.0001$) from laboratory read speech to casual speech, and by 17% ($p < 0.0001$) from laboratory conversational speech to casual speech. CC SDS changed the smallest amount (8%) ($p < 0.05$) from laboratory read speech to laboratory conversational speech.

<i>Source</i>	<i>df</i>	<i>F</i>	<i>Prob>F</i>
Environment	2	3.77	0.0233
Talker	10	12.58	p<0.0001
Interaction	20	11.35	0.0439

Table 5: Results from two way ANOVA comparing P0 (mean, ST) for each recording environment.

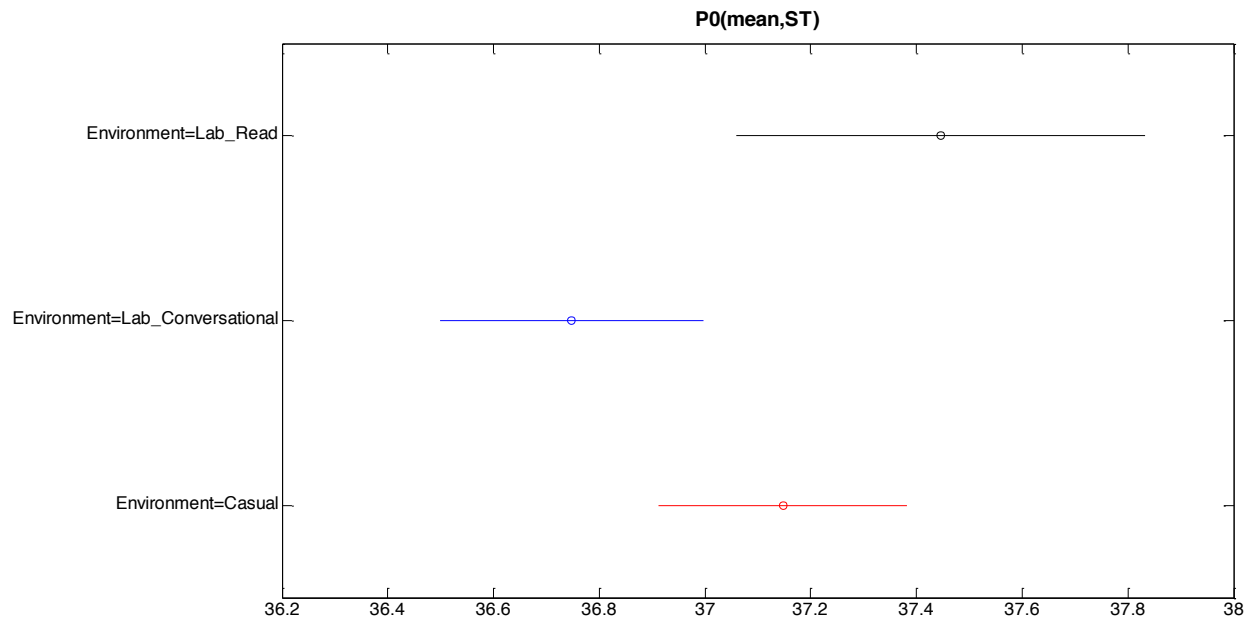


Figure 11: ANOVA Post Hoc Test Results for P0(mean, ST) in three recording environments

<i>Source</i>	<i>d.f</i>	<i>F</i>	<i>Prob>F</i>
Environment	2	7.24	0.0007
Talker	10	27.70	p<0.0001
Interaction	20	2.45	0.004

Table 6: Results of two way ANOVA comparing P0 (std, ST) for each recording environment.

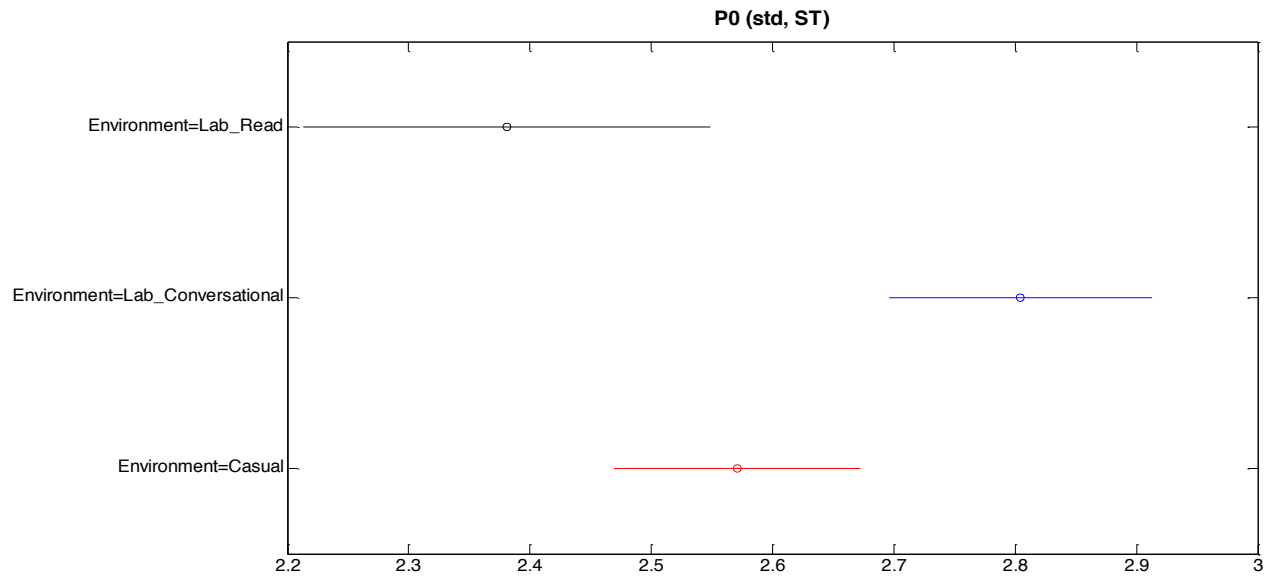


Figure 12: ANOVA Post Hoc Test Results for P0(std, ST) in three recording environments

Tables 5 and 6 summarize the ANOVA results for P0 mean and STD, respectively. Both tables show a p-value of <math><0.0001</math> for the talker factor. Both the environment and interaction p-values were below the threshold value of F = 3.77, p < 0.0233) among recording environments. Figure 11 (ANOVA Post Hoc Test Results) shows a significant difference for P0(mean, ST) between laboratory read and laboratory conversational ($p < 0.05$). The highest value for P0 (mean, ST) was laboratory read speech (37.5) followed by casual speech (37.1) and the lowest value for P0 (mean, ST) was laboratory conversational speech (36.7). Results from the post-hoc test reveal the P0 (mean, ST) changes by 2% ($p < 0.05$) from laboratory read speech to laboratory

conversational speech. P0 (mean, ST) changed by 1% ($p > 0.05$) from laboratory conversational speech to casual speech, and 1 % ($p > 0.05$) from laboratory read speech to casual speech. Figure 12 shows (ANOVA Post Hoc Test Results) a significant difference between laboratory conversational speech and casual speech. Table 6 shows P0(std, ST) a significant difference among recording environments ($F = 7.24$, $p < 0.0007$) Figure 12 also illustrates a significant difference between laboratory read speech and laboratory conversational speech. P0 (std, ST) for laboratory conversational speech had the highest mean (2.8) followed by casual speech (2.59) and laboratory read speech (2.39). Results from the post-hoc test demonstrate a 15% ($P < 0.05$) change in P0 (std, ST) from laboratory conversational speech to laboratory read speech. P0 (std, ST) changed by 8% ($p < 0.05$) both from casual speech to laboratory read speech and laboratory conversational speech to casual speech.

<i>Source</i>	<i>d.f</i>	<i>F</i>	<i>Prob>F</i>
Environment	2	167.39	p<0.0001
Talker	10	83.35	p<0.0001
Interaction	20	25.15	p<0.0001

Table 7: Results from two way ANOVA comparing PS (mean) for each recording environment.

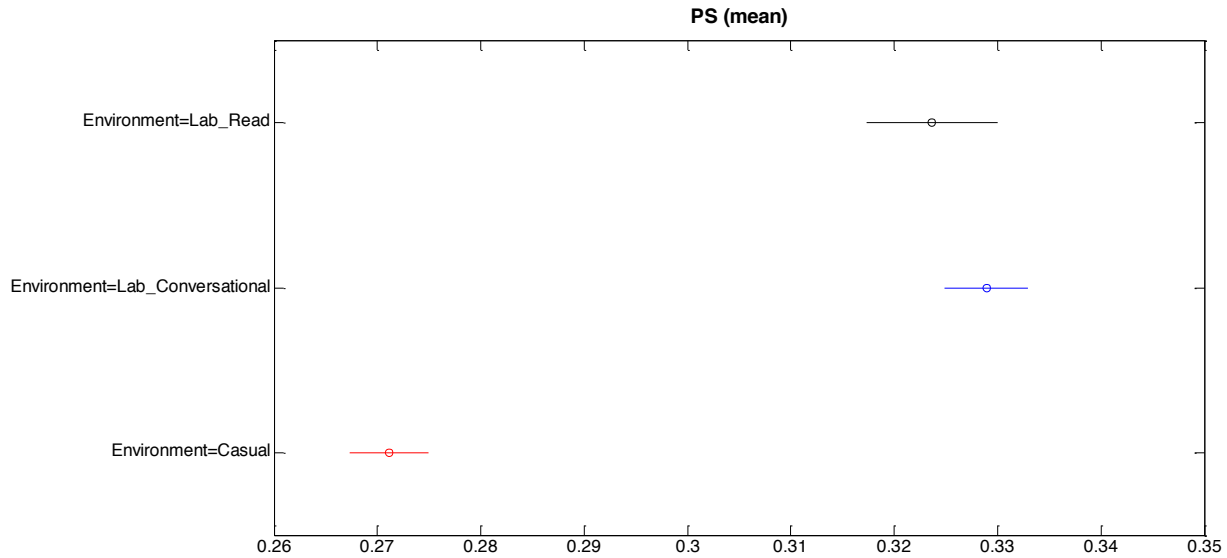


Figure 13: ANOVA Post Hoc Test Results for PS (mean) in three recording environments

Table 7 summarizes the ANOVA results for PS mean. Table 7 shows a significant difference in PS (mean) among recording environments ($F=167.39$, $p<0.0001$). The table shows the p-values for environment, talker, and interaction were all significantly different from each other ($p<0.0001$ for all comparisons). Figure 13 (ANOVA Post Hoc Test Results for PS (mean)) illustrates a significant difference between casual speech compared to both laboratory conversational and laboratory read speech. Laboratory conversational speech had the highest mean (0.33) followed by laboratory read speech (0.325) and the lowest was casual speech (0.27). Post-hoc test for PS (mean) reveal a significant difference, 20% ($p<0.05$) change from laboratory conversational speech to casual speech, and 19% ($p<0.05$) from laboratory read speech to casual speech. PS (mean) changed by 2% ($p<0.05$) from laboratory conversational speech to laboratory read speech.

DISCUSSION AND CONCLUSIONS

The purpose of the current study was to investigate differences in speech characteristics among different recording environments. A change in speech characteristics depending on environment show a performance effect (Goberman, et. al 2010). For the current study it was hypothesized that (a) speech recorded outside the laboratory, when speakers are engaged in routine communicative tasks, will result in a different form of speech (casual) than the conversational speech obtained in a speech laboratory and reported in most experiments,(b) laboratory read and conversational speech are different due to a performance effect and (c) a significant difference (means p value <0.05) will be evident in acoustic measures among different recording environments. Results from the current study confirm that several acoustic characteristics were significantly different between laboratory read, laboratory conversational, and casual speech, evidenced by p values being below 0.05 in analyses of variance tests.

From the results a pattern emerged from all acoustic measures: effortful speech (laboratory read speech) produced higher means of CC SDS, DCC SDS, P0 (mean, ST and std, ST), and lower rate of speech (syllables per second). As mentioned earlier CC SDS (cepstral coefficient standard deviation sum) is an acoustic correlate of that represents the range of articulation. CC SDS mean comparisons revealed that laboratory read speech showed the highest mean, followed by laboratory conversational, and the lowest mean was casual speech. CC SDS mean comparisons also showed that the laboratory conversational speech CC SDS mean value was closer to casual speech CC SDS mean than laboratory read speech. Mean comparisons for DCC SDS showed a similar pattern as CC SDS mean comparisons where laboratory read speech showed the highest mean, followed by laboratory conversational speech, and then casual speech showing the lowest mean value. The laboratory conversational DCC

SDS mean was also closer in value to the casual speech DCC SDS mean. CC SDS Post Hoc test revealed a significant (21%) ($p < 0.05$) change from laboratory read speech to casual speech. DCC SDS post-hoc test also showed a significant (25%) ($p < 0.05$) change from laboratory read speech to laboratory conversational speech.

P0 (mean, ST) mean for casual speech was lower for the laboratory read speech but higher than laboratory conversational speech. P0 (std, ST) mean for laboratory conversational speech was higher than both laboratory read and casual speech. Post-hoc test results showed no significant differences for P0 (mean, ST) in all three environments. For both pitch measures (P0 mean, ST and P0 std, ST) at least one laboratory speech (read or conversational) mean value was higher than casual speech. PS mean was higher for laboratory read speech and laboratory conversational speech. PS post-hoc test revealed a significant change (20%) from laboratory conversational speech to casual speech, and a 19% ($p < 0.05$) change from laboratory read speech to casual speech. Since PS has been observed to correlate with lower breathiness (Anand, et al, 2013), this may reflect more effort and better glottal closure during voice production in the laboratory then in their natural environment. P0 (std, ST) post-hoc test revealed significant differences (15%) ($p < 0.05$) from laboratory conversational speech to laboratory read speech.

Rate of speech (syllables per second) was slower for laboratory speech, which demonstrates the same effect as previous research studies stating that more effortful speech produces a slower rate of speech (Bradlow et. al, 1996; Bradlow, Kraus, & Hayes, 2003; Goberman & Elmer, 2005; Hargus Ferguson & Kewley-Port, 2002; Helfer, 1997; Pichney et al., 1985; Schum, 1996). These findings reflect a performance effect in laboratory recording when compared to casual speech. Statistical analysis done on the data of the current study revealed a

significant difference between casual speech and laboratory recording. Rate of speech Post Hoc test revealed a significant change (31%) ($p < 0.05$) from laboratory read speech to casual speech.

Recording environment and the type of laboratory speaking tasks played a role in unmasking the performance effect speakers naturally give when being studied. Each talker produced a different form of speech depending on recording environments as evidenced by difference in various acoustic measures. Thus, speakers change their form of speech depending on the situational task and environment they are in.

There were a few limitations to the current study. First, due to time constraints, an unequal number of young adult and old adult talkers could be recruited for the study. This prevented a direct comparison of age related changes in acoustic characteristics of speech in various environments. The intention of the current study was to have a larger number of talkers in different age groups, but the primary investigator was unable to complete the data collection as planned due to time constraints.

Second, the current study did not account for other differences in speakers. A multitude of comparisons could be done in future research such as analyzing if the performance effects on males vs. females, young adult vs. old adult, or healthy participants vs. atypical participants. Further research could also investigate what speech tasks elicit more casual speech characteristics and the duration of casual speech sample that is adequate to represent a speaker's typical behavior.

Results from the current study indicate that most laboratory conversational speech means were the closest to casual speech means obtained in a laboratory setting. This result could be beneficial for future research attempting to elicit a talker's actual speech characteristics in the laboratory. The results also could be beneficial in the development and implementation of speech

diagnostic tools and activities targeting talker's natural speech characteristics. These results show that a performance is evident in laboratory recorded speech and natural recordings reveal a better representation of talker's actual speech characteristics.

APPENDIX

APPENDIX

Laboratory Read Sentences (16 spin sentences and 3 additional sentences)	
1.	I made the phone call from a booth.
2.	The cut on his knee formed a scab.
3.	I gave her a kiss and a hug.
4.	The cop wore a bullet-proof vest.
5.	How long can you hold your breath?
6.	At breakfast he drank some juice.
7.	The soup was served in a bowl.
8.	The cookies were kept in a jar.
9.	The baby slept in his crib.
10.	I ate a piece of chocolate fudge.
11.	The judge is sitting on the bench.
12.	The boat sailed along the coast.
13.	His boss made him work like a slave.
14.	He caught the fish in his net.
15.	The beer drinkers raised their mugs.
16.	The pirates buried the treasure
17.	They're glad we heard about the track.
18.	On the beach we play in the sand.
19.	I should have considered the map.

Table 8: Laboratory Read Sentences (16 spin sentences and 3 additional sentences)

REFERENCES

REFERENCES

- Bradlow, A.R., Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112, 272–284.
- Bradlow, A. R., Kraus, N., and Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, 46, 80-97.
- Bradlow, A.R., Torretta, G.M., Pisoni, D.B. (1996). Intelligibility of normal speech i: global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*. 20. 255–272.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, 39–54.
- Calderon, S., Alvarado, G., & Camacho, A. (2011). AUD-SWIPE-P: A Parallelization of the AUD-SWIPE Pitch Estimation Algorithm Using Multiple Processes and Threads.
- Ferguson, S.H, Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*. 112, 259–271.
- Ferguson, S.H. (2004). Talker differences in clear and conversational speech :Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical society of America*. 116, 2365–2373.
- Ferguson, S.H, Kewley-Port,D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of speech, Language and Hearing Research*. 50, 1241–1255.
- Ferguson, S.H., Kerr, E.E. (2009). Subjective Ratings of Sentences in Clear and Conversational Speech. *Journal of the Academy of Rehabilitative Audiology*. 42, 51–66.
- Ferguson, S.H. (2012). Talker Differences in Clear and Conversational Speech: Vowel intelligibility for Older Adults with Hearing Loss. *Journal of Speech, Language, and Hearing*. 55, 779–790.
- Gagne, J.P., Masterson, V., Munhall, K.G., Bilida, N., Querengesser, C. (1994). Across Talker Variability in Auditory, Visual, and Audiovisual Speech Intelligibility for Conversational and Clear Speech. *Journal of the Academy of Rehabilitative Audiology*. 27, 135–158.
- Goberman, A.M., Elmer, L.W. (2005). Acoustic Analysis of clear versus conversational speech with individuals with Parkinson Disease. *Journal of Communication Disorders*. 38, 215–230.

- Goberman, A.M., Recker, B., Parveen, S. (2010). Performance Effect: Does the Presence of a Microphone Influence Parkinsonian Speech? *Journal of Medical Speech Language Pathology*. 18 (4), 40–45.
- Harnsberger, J. D., Wright, R., & Pisoni, D. B. (2008). A new method for eliciting three speaking styles in the laboratory. *Speech Communication*. 50,323-336.
- Hirschberg, J., Litman, D., Swerts, M. (2004). Prosodic and other cues to speech recognition failures. *Speech Communication*. 43, 155-175.
- Hollien, H., Hollien, P., Jong, G.D. (1997). Effects of three parameters on speaking fundamental frequency. *J.Acoust. Soc. A.*. 102, 2984.
- Kalikow, D.N., Stevens, K.N., Elliot, L.L. (1993). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, 61, 1337–1351.
- Kent, R.D., & Read, C.R. (2002). *Acoustic Analysis of Speech: Second Edition*. Albany, NY. Thomson Learning Inc.
- Krause, J.C., Braida, L.D. (1995). The effects of speaking rate on the intelligibility of speech for various speaking modes. *Journal of the Acoustical Society of America*. 98 (2), 2982.
- Krause, J., & Braida, L. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America*, 112, 2165–2172.
- Lansberger, H.A. (1959). Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry. *American Sociological Review*, 24, 277–278.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory, in: Hardcastle, W.J., and Marchal, A., Eds., *Speech Production and Speech Modeling*. Kluwer Academic Publishers: Dordrecht, 403–439.
- McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*. 44, 128–143.
- Moon, S.J., Lindblom, B. (1994). Interaction between duration, context, and speaking style in english stressed vowels. *The Journal of the Acoustical Society of America*. 96, 40–55.
- Picheny, M.A., Durlach, N.I., Braida, L.D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *J. Speech Hear. Res.*, 28, 96–103.

- Picheny, M.A., Durlach, N.I., Braida, L.D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *J. Speech Hear. Res.*, 29, 434–446.
- Picheny, M.A., Durlach, N.I., Braida, L.D. (1989). Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*. 32, 600–603.
- Schum, D.J. (1996). Intelligibility of Clear and Conversational speech of Young and Elderly Talkers. *Journal of the American Academy of Audiology*. 7, 212–218.
- Skowronski, M.D., & Harris, J.G. (2004). Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition. *Journal of the Acoustical Society of America*, 116(3), 1774–1780.
- Skowronski, M.D., Shrivastav, R., Harnsberger, J., Anand, S., and Rosenbek, J., (2012), "Acoustic discrimination of Parkinsonian speech using cepstral measures of articulation," *Journal of the Acoustical Society of America*, 132(3), 2089.
- Smiljanic, R., Bradlow, A.R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America*. 118, 1677–1688.
- Uchanski, R.M., Choi, S., Braida, L.D., Reed, C.M., Durlach, N.I., (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research*. 39, 494–509.