University of South Florida

Digital Commons @ University of South Florida

Outstanding Honors Theses   Honors College

Spring 5-2-2012

Effect of Speaking Style on Variability of Vowel Production for Native Monolingual English Speakers

Amber Gordon

Follow this and additional works at: https://digitalcommons.usf.edu/honors_et

Part of the American Studies Commons

The Effect of Speaking Style on Variability of Vowel Production for Native Monolingual

English Speakers

Amber Gordon

University of South Florida Honors Thesis

Thesis Director: Dr. Catherine Rogers

Thesis Committee Member: Dr. Stefan Frisch

I. INTRODUCTION

When one listens to a vowel sound in a word, it may seem steady and unchanging. Despite the static sound one's ears may be hearing, vowel modifications and acoustic changes are rapidly occurring. These acoustic modifications include changes in fundamental frequency and formant frequencies, which affect the overall uniqueness and intelligibility of an individual's voice. While one may not realize the frequency changes that are occurring during the milliseconds it takes to pronounce a vowel sound, one may be more aware of how he or she changes speaking style when a room becomes noisy. This conscious or subconscious change in speaking style to make one's self more clear has an important connection with the acoustic and durational characteristics of vowels that one rarely thinks about. Changing one's conversational style from conversational to clear tends to signal a change in vowel's inherent characteristics.

The present study seeks to understand what changes in a vowel must occur to make oneself more intelligible (clear) by answering the question "for native monolingual English speakers what is the effect of speaking style on variability of vowel production?" The study's hypothesis is that when looking at the difference between clear and conversational speech, a decrease in standard deviation will be evident. To see whether the hypothesis is valid, six vowel sounds (the vowels in "bead", "bid", "bayed", "bed", "bad", and "bod") produced by four different speakers in both clear and conversational speech were be analyzed for fundamental frequency, first formant frequency, and second formant frequency at 20%, 50%, and 80% of vowel duration. The specific speakers were chosen from a group of speakers in a previous study (Rogers, DeMasi, & Krause, 2010). Two of the speakers were chosen due to the fact that they had a large increase of intelligibility when switching from clear to conversational speech. The other two were chosen because their intelligibility actually decreased slightly when speaking in

clear speech. Thus these talkers can be compared to each other to see the differences in fundamental frequency, first formant frequency, and second formant frequency between and within the groups. To fully understand the present study it is important for one to understand what clear and conversational speech is, past studies related to the subject, and the acoustic properties of speech

Conversational and clear speech are two speaking styles that a talker may utilize when conversing with a listener. As can be deduced by the name, conversational speech refers to the way an individual typically speaks in normal, everyday conversation with friends or family. On the other hand, clear speech is the speech that one uses when asked to speak more clearly or when one is asked to repeat what he or she has said (Rogers et. al, 2010). Clear speech is most often used when speaking in situations where background noise is overpowering to the speech signal, such as at a concert or a noisy restaurant. For individuals with hearing within the normal range, a 8% increase in the intelligibility of vowel sounds in noise has been shown (Ferguson, 2004). In addition to being used in a noisy setting, clear speech is often utilized by individuals when they are conversing with someone who has a hearing loss or someone who is not a native English speaker. Studies have shown that for both normal-hearing listeners and listeners with a hearing loss 17 to 20 percentage point increase in sentence intelligibility occurs when speaking to someone in a noisy environment (Ferguson, 2004). The term that refers to the concept that talker's increase their intelligibility when using clear speech compared to conversational is referred to by several researchers as the "clear speech benefit."

One study (Rogers, DeMasi and Krause, 2010) has researched the degree of clear speech benefit in noise for monolingual English speakers compared to early and later learners English learners of English as a second language. This study is particularly relevant to the present study

due to the fact that the present study used four of the same monolingual English speakers as

Rogers et al. (2010). In the Rogers et al. study, the three groups of speakers were recorded saying

the words "bead, bid, bayed, bed, bad," and "bod" in both clear and conversational speaking

styles in order to see how much the individuals in the three groups could increase their

intelligibility in noise. The speakers' intelligibility was then judged by a single group of 20

individuals, who listened to each word and chose which one they heard from a list of six

alternatives on each trial. For all three talker groups, the clear speech benefit was significant but

the size of the varied from person to person, and across the groups. In general, it was found that

as one might suspect, individuals who learned English later in life had the smallest clear speech

benefit at only 4%. Perhaps not as obvious, the greatest clear speech benefit was not from native

English speakers but rather from the early learners who had a 7-8% average benefit, which may

be due to increased closure durations. On average, the clear speech benefit for the monolingual

English speakers was about 5-7%, which is still a significant improvement. The benefit is

consistent with 8.5% benefit that was recorded by Ferguson's (2004). While the averages of the

clear speech benefits are important, it is also crucial to note that the differences between talkers

within a group was actually greater than the differences between talks in different groups

(Rogers et al., 2010). This idea that the differences within a group are greater than between

groups will be utilized in the present study. From the native monolingual speaker group of the

Rogers, DeMasi, Krause study, the largest clear speech benefit was 22% and the smallest was

-11% (a significant decrease in performance), showing variability in performance within the

group is quite large. This degree of variability across talkers is typical of clear speech studies;

some talkers are much better than others at improving intelligibility when asked to speak clearly.

The four talkers selected for the present study include two who showed a large and statistically

significant clear speech benefit of 15% and 16%, and two who showed no benefit (-2% and -3%, not statistically significant).   Understanding differences between the acoustic properties of speech produced by talkers who do, and do not, show a clear speech benefit is the cornerstone of the study, and therefore and understanding of different acoustic properties (fundament frequencies and formant frequencies) that will be analyzed is vital. Furthermore, a previous study (Bianchi, 2007) showed acoustic variability across talkers, but only two tokens of each word were analyzed in each speaking style, so within talker variability could not be examine. A comparison of within talker variability between talkers showing a large vs. a small clear speech benefit will help to better understand talker strategies in producing the two speech styles.

For the present study, the acoustic properties that were analyzed to determine the difference in variability between big benefit and no benefit speakers include the talker's fundamental frequency, the first formant frequency, and the second formant frequency. The fundamental frequency (F0) is defined as the lowest harmonic of a talker's voice, and formants are described as "vocal tract resonances" (Raphael, Borden, & Harris, 2003, p. 111). When a speaker changes the movement of his or her articulators for different sounds, the formant frequencies also change. While each speech sound is composed of more than two formant frequencies, the F1 and F2 are the two that were analyzed for the present study and are most important for vowel identification. A person's individual F1 is determined most by how open the mouth is when making a sound or how high or low the tongue is in the mouth. When the tongue is positioned high in the mouth, near the roof of the mouth, the vowel has a low first formant frequency. On the other hand, if the tongue is located away from the roof of the mouth and low in the mouth, the sound will have a high first formant frequency. The second formant frequency depends on how large the oral cavity is and if the tongue is located in the front or back

of the mouth (Raphael et al., 2003). Front vowels and consonants have a high F2, while vowels

that are made with the tongue towards the back of the mouth have a low F2 (Raphael, et. al,

2003). These relationships between articulator position and first and second formant frequencies

have been demonstrated in many different acoustic studies.

Ferguson and Kewley-Port (2007) examined acoustic differences between vowels

produced in conversational and clear speech. Similar to the present study, the vowels of six

individual speakers who produced a large clear speech benefit and six who did not produce any

clear speech benefit were analyzed. From the analysis, the researchers noted that vowels spoken

in clear speech were longer in duration. The individual speaking at the slowest rate had the

greatest clear speech benefit, which is consistent with another finding by Bond and Moore

(2004). Despite the consistency between the two studies, three other studies (Picheny et al.,

1989; Uchanski et al., 1996; and Nejime and Moore, 1998) did not find that decreasing speech

rate is of utmost important to intelligibility. The fact that individuals who use clear speech

increase the size of the vowel space also was noted to be important. The study showed that

increasing vowel space alone is not enough, because both Big Benefit (BB) talkers and No

Benefit (NB) talkers utilized increasing vowel space in the Ferguson and Kewley-Port (2007)

study. The difference was that BB talkers had a 9% increase while NB had only a 5% increase in

vowel space. Big benefit talkers also differed from no benefit talkers in the fact that BB

speakers' F1 values had a greater deal of difference between high versus low vowels. The

changes observed in the F2 for the study, such as high frequency for front vowels, did not seem

to play a part in clear speech intelligibility. In fact, according to the study, the increase in F2 for

front vowels actually can make one's voice less intelligible to hearing-impaired listeners

(Ferguson & Kewley-Port, 2007). Perhaps, more important than the static F1 and F2 measurements are the findings on the dynamic measures of vowels.

Every vowel and consonant has a range of frequencies that help to characterize it. For instance, for a typical male speaker, /i/ (as in the vowel sound in "bead") and /I/ (the vowel sound in "bid") would have a low F1 corresponding to between 250-350 Hzand a high F2 around 2100-2400 Hz (Small, 2005). While the previous numbers represent the average frequencies of the formants, the frequencies themselves may actually change throughout the entire duration of the vowel (Hillenbrand et al., 2005). Therefore, if one compares the vowel's resonant frequencies at 150 milliseconds it will be most likely be different than the frequency observed at 200 milliseconds. This means it is not enough to just report the F1 value for a specific speaker is 2100 Hz, because this is not true for the entire duration of the vowel. The fact that formant frequencies change throughout the duration of a phoneme is a relatively new idea being researched in the field of speech acoustics. In past studies, the vowel's formant frequencies would be computed at the midpoint (50%) of the vowel duration. However, in a study conducted by Hillenbrand et al. in 2005, it was shown that "monophthongal" English vowels, such as /i/ and /I/, show changes in the dynamic property of vowels, which refers to the "degree and direction of change of formant frequencies during vowel production (Bianchi, 2007, p. 3)".

In relation to the two different speaking styles, two research studies, one by Mon and Lindbolm (1994) and another by Kewley-Port (2002), showed that formant movement increased during clear speech, specifically in the vowel nucleus (Ferguson & Kewley-Port, 2007). According to Ferguson and Kewley-Port, it also seems that the dynamic properties of vowels during clear speech are related to increases in duration. Ferguson and Kewley-Port reason that "it may be that talkers who make their vowels more dynamic in clear speech deliberately slow

down to avoid overshooting the formant frequency targets for individual vowels" (Ferguson &

Kewley-Port, 2007, p. 1253). Another idea is that a talker might start to speak more slowly to

become clearer and realize that the change in duration also allows for them to make the vowels

more dynamic. Due to the importance of vowel change over duration of time, the present study

examined fundamental and formant frequencies at 20%, 50%, and 80% of the vowel duration for

both clear and conversational speaking samples.

II. METHOD

Talkers.

The talkers for the study were considered monolingual native English speakers due to the

fact that they reported that they were not fluent in a second language. The candidates also did not

grow up in families where an additional language was utilized in the home. Besides the

monolingual criteria, the talkers also did not have previous speech or hearing difficulties or a

strong region-specific accent.

In total, there were four female talkers who were identified as talker 6, talker 14, talker

15, and talker 20. Since all the talkers were female, it may be important to note that in a previous

study it was discovered that females tend to have a great vowel intelligibility than men when

talking in clear speech, so the differences between gender will not be analyzed or noted in the

present study (Ferguson, 2004).

The four talkers' ages varied from 20 to 27 years old. Talker 6 was 20 years old, talker 14

was 27 years old, talker 20 was 22 year old, and talker 15 was 20 years old. These specific

individuals were picked because they either showed a strong clear speech benefit or

no clear speech benefit in a previous study. The benefit is described as percent change for

speech presented in noise (Clear speech- Conversational). Talkers 6 had a 15.8% clear speech

benefit and talker 14 had a 16.2% clear speech benefit, thus making them the good benefit

talkers. On the other hand, talker 15 showed a -3.3% benefit and talker 20 showed a -2.1%

benefit, thus making them the two individuals who did not show any clear speech benefit (see

Table 1).

Table 1. Percent correct word recognition in noise by native monolingual listeners for the four talkers, for target words spoken in conversational speech style, clear speech style and the difference between the two. Two talkers had a strong and statistically significant clear speech benefit, while the other two had a negative benefit (not statistically significant).

Talker	Conversational	Clear	Difference
MO06	40	55	+15
MO14	30	46	+16
MO15	60	57	-3
MO20	41	39	-2

Materials

Before the study began, the following recording equipment was chosen and set up: Audio-Technica: AT4033 condenser microphone, an Applied Research and Technology microphone preamplifier with 48V phantom power supply, a Roland VS890 Digital Studio Workstation recorder, and Sennheiser HD265 headphones.

The four speakers' recordings took place inside a single-wall sound attenuating booth (IAC). Inside of the booth was a 15 inch flat screen on which the /b*V*d/ syllables ("bead, bid," etc.) appeared via a Microsoft PowerPoint presentation. From outside the booth, the experimenter sat at a separate computer in order to control the PowerPoint presenting the sentences with just a mouse click.

Recording Procedures.

The syllables analyzed include the following /bVd/ words or syllables: "bead, bid, bayed, bed, bad, and bod." The talkers were shown the /bVd/ words via the screen in the carrier phrase, "Say _____ again," and the syllables were also spoken to the participants by the experimenter to ensure the talkers knew how to correctly pronounce the words. The talkers spoke the syllables into a microphone that was adjusted at 45 degree angle and approximately 6 inches away from the talker's mouth.

During recording of the conversational speech sample, distractor /CVC/ words, such as "cat," were used to ensure that talkers did not focus too much /bVd/ syllables. Too much emphasis on /bVd/ syllables could cause the conversational speech to begin to resemble clear speech instead of normal conversational speech. Also before the conversational task, the speakers were instructed to speak how they would in a normal conversation. While distractor words were not utilized for clear speech recording, the talkers were asked to speak in the way that they would in response to being told that they were not understandable. To verify that the talkers were comfortable with the tasks, they performed 12 practice trials by reading target syllables in carrier phrases. After the practice trials were finished, the actual study began. From the recording, five of the repetitions for each /bVd/ syllable for each speaking style were chosen to be used with the intention of the vowel portion of the word being analyzed.

Editing Procedure

An acoustic editing program was utilized by two trained experimenters to edit all the recorded words from the phrases and store them as single isolated words. Each separate word was isolated by the experimenter first by locating and selecting the release of the initial /b/, plus selecting the 20 ms of silence that occurred before /b/ release. Any information that was present on the waveform before the 20 ms was then ready to be deleted. In addition the 20 ms was cut

down to 10 ms after the initial deletions were made. For the release of the /d/ in the word-final position, the /d/ release was selected from the waveform along with the following 20 ms. Eventually, the final 10 ms of the 20 ms was cut. Now, a new waveform that consisted of only the syllable and the preceding and following 10 ms was available for vowel analysis. This new waveform was then saved to its own new word file.

Acoustic Measurements.

Acoustic analysis software, which displays an image of a waveform, a spectrogram and formants, was utilized for measurements of time and frequency. Before the analysis on the software began, specific parameters were set up. The formant settings included a dynamic range set to 30 with a window length of .025s. The maximum formant frequency that could be detected was at 5500 Hz. The number of formants that the program showed varied between 4-6 and depended upon which parameters the author believed would make the formants more accurate for any given word. The spectrum settings included a wind band spectrogram range from 0-5000 Hz, with a window length of .005s and a dynamic range of 50 dB.

The author was trained by the thesis supervisor in use of the acoustic analysis software and identification of software settings needed to make a good match between computer generated and visually observed values of the formant frequencies and fundamental frequency. Following training, three repetitions of the six vowels produced by the four different speakers were analyzed, in addition to the two that had already been analyzed for each, making a total of five repetitions per talker. First, the author decided where on the waveform a particular vowel started (vowel onset) and where it ended (offset). The time of the first large positive amplitude after the /b/ release was determined to be the time of vowel onset. The first negative pulse of the

last cycle of voicing that had a shape that looked like the rest of the vowel was determined to be the vowel offset. In addition, vowel onset and offset could be verified on the spectrogram by determining the first place a pitch pulse containing F2 was seen for onset and the last place a pitch pulse containing F2 could be seen for offset. Once the author felt confident that the onset and offset were correct, the corresponding times were put into a spreadsheet. The spreadsheet then used a pre-programmed formula to compute where 20%, 50%, and 80% of the vowel were located.

After the spreadsheet computed at what time 20%, 50%, and 80% of the vowel was located, the experimenter would place the cursor at those three different locations to find the F0, F1, F2, F3, and F4 for each point in time. The software then computed all five frequency measures at the given time points. In some cases, the number of formants to be located within the frequency range was adjusted up or down from the default of five, to provide a better match with the visually observed formants. When the software could not provide a good match with the formants determined the frequencies by hand. The F0 never had to be determined by hand, and the F3 and F4 both were not used for the present study.

III. Results

After analysis for the speakers was completed, the next step was to find the average standard deviation for every talker at every time point for the following dependent measures: fundamental frequency, first formant, and second formant frequency (see Table 2). Standard deviations were computed separately for the conversational and clear speech styles, as shown in Table 2. In addition, the difference between the two speaking was also calculated for the dependent variables

Computing standard deviations was vital to the study, because the hypothesis for the present research was that when looking at the mathematical difference between clear and conversational speech, a decrease in the size of the standard deviation would be observed from the conversational to the clear speech style. It was hypothesized that the degree of improvement in intelligibility (talkers 6 and 14), the greater the decrease in standard deviation would be observed when computing the conversational and clear speech difference. In theory, the hypothesis seemed logical because if all the speakers have a specific idea of how to speak clearly already set in their mind and a clear articulatory target for each vowel, it would be shown as the participants reducing their variability of production to home in on this ideal prototype.

Once the standard deviations were computed, it was time to test the hypothesis. By taking a look at Table 2, it is obvious that talkers 6 and 14 (strong clear speech benefit) did not become less variable in standard deviation for their production of each target phoneme. They also did not become more variable though; no consistency either way is evident. Therefore, there was no apparent support for the original hypothesis, and a new approach to analyzing the data had to begin.

The new approach to understand the data was based off of the Formant-Vowel Ratio theory of vowel perception. This theory proposes that, "Vowel identity depends on the intervals between formants (i.e. formant ratios) rather than absolute formant values (Pickett, p. 155)." This means that for vowels that are considered "low, front vowels (/a/)" for which F1 and F2 have similar frequencies and a small interval between them, the ratio would be low and around 1.42. On the other hand, vowels such as /i/ that are high, front vowels and have a low F1 and a high F2 would have a larger ratio (approximately 8.71) (Pickett). These ratios are more important due to the fact that every person's frequency varies depending on their vocal tract. Children have a

14

Table 2. Standard deviation values for each talker at each of three time points measured, for the dependent measures of fundamental frequency, first formant and second formant frequency, for target words spoken in conversational speech style, clear speech style, as well as the difference between the two.

| | Fundamental frequency | | | | | | | | |
| | 20% | | | 50% | | | 80% | | |
Talker	Conv	Clear	Difference	Conv	Clear	Difference	Conv	Clear	Difference
MO06	7.73	10.59	2.87	8.57	11.94	3.37	7.31	9.30	1.99
MO14	4.35	3.28	-1.07	3.95	3.19	-0.76	3.63	3.96	0.33
						0.00			0.00
MO15	3.48	8.23	4.75	3.96	7.80	3.84	4.58	8.26	3.68
MO20	3.97	2.83	-1.14	2.96	4.06	1.11	3.43	6.68	3.25

| | First formant | | | | | | | | |
| | 20% | | | 50% | | | 80% | | |
Talker	Conv	Clear		Conv	Clear	Difference	Conv	Clear	Difference
MO06	23.37	34.99	11.62	30.70	35.06	4.35	22.03	30.96	8.92
MO14	24.08	27.47	3.38	24.61	22.47	-2.14	20.06	33.77	13.71
MO15	15.96	33.44	17.49	29.39	24.75	-4.64	33.87	21.21	-12.65
MO20	35.09	23.48	-11.60	28.11	23.65	-4.46	23.08	24.32	1.24

| | Second formant | | | | | | | | |
| | 20% | | | 50% | | | 80% | | |
Talker	Conv	Clear	Difference	Conv	Clear	Difference	Conv	Clear	Difference
MO06	57.23	81.72	24.49	45.78	48.93	3.14	39.16	72.97	33.82
MO14	45.76	47.25	1.49	51.75	59.35	7.60	63.67	41.63	-22.04
MO15	41.41	60.96	19.55	35.51	64.61	29.11	57.52	57.77	0.24
MO20	92.51	70.60	-21.91	48.88	57.70	8.82	61.57	48.56	-13.00

small vocal tract and thus higher formant frequencies. One can verify this by thinking of most children that they have met. On the other hand, men have a large vocal tract and thus lower formant frequencies (Raphael et. al, 2003). While the frequencies may vary from person to person, the ratios remain relatively constant.
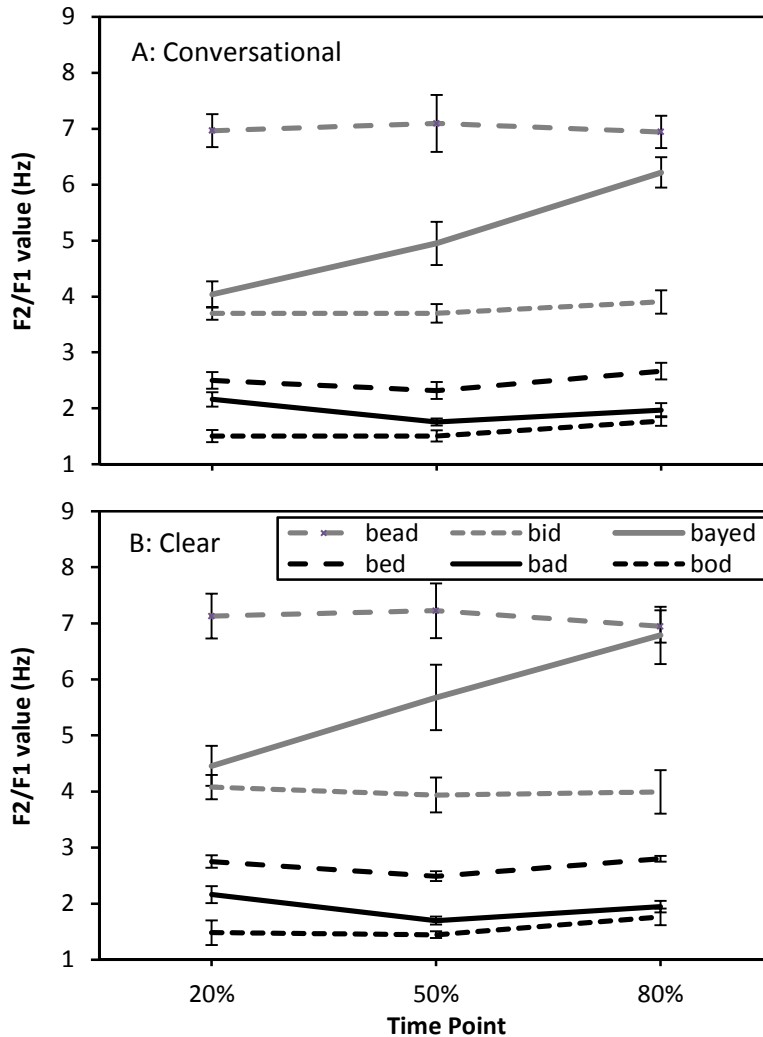
Figure 1. Average F2 to F1 ratio at the three time points measured and each target syllable for talker MO06, for conversational (panel A) and clear speech (panel B) styles. Error bars indicate one standard deviation from the mean.

Figure 1 shows the ratio between F2 and F1 (F2/F1) at each time point for the conversational (panel A) and clear speech (panel B) styles for one example talker (MO06). Error bars indicate one standard deviation from the mean, across the five tokens measured. If one takes a glance at Figure 1, variation in the size of the standard deviation across the target syllables for the ratio of F2 to F1 can be noted. This variation follows the Ratio theory. It can be seen that for both clear and conversational speech some words have rather large standard

deviations, such as "bead" and "bayed", while others have much small standard deviations, such as for "bed and "bod." For words such as "bed" and "bad," it can be seen from Figure 1 that they are very close to other syllables in the figure, which depicts the fact that the vowel space tends to be crowded for those specific vowels. This crowding seems to explain the decrease in variability. Since the vowels "bed" and "bad" are so near each other when being produced in the mouth, if there was an increase in variability, the two vowels' frequencies may actually overlap, which would decrease their intelligibility and cause them to be confused for one another. Therefore, it seems that the decrease in variability of articulation is helping each vowel to have a strong, obvious identity.

On the other hand, the vowels for "bead" and "bayed" are rather spread apart on Figure 1. In contrast to "bed" and "bad", these vowels are not crowded and would not overlap in the vowel space, so they do not need to be less variable in articulation. Even if an individual spoke the words "bead" and "bayed" somewhat sloppily and with high variability in articulation across repetitions, the specific vowel would still be clear and easily understood. Their frequencies would not overlap with those of other vowels, so no confusion in vowel identify would occur. Therefore, it appears that it can be said that whether the standard deviation increases or decreases depends upon if there is a large or small distance between two neighboring vowels, for this talker. This indicates a positive correlation.

A series of correlation analysis was performed to test the hypothesis of a relationship between distance between a vowel and its nearest (lower) neighbor in the F2/F1 space and the size of variability for the vowels, for each talker and speaking style. Because the distance to the next lower vowel was computed in each case, distances were only computed for five of the six vowels, so fifteen values were used in each correlation (five vowels times three time points). For

each speaking style and each talker, a Pearson's r statistic and p value was calculated from the 15

standard deviation values and the 15 distance scores. Table 3 shows the results of the correlation

analyses for each talker and speaking style.

The Pearson's p values were multiplied to ensure the outcome was not due to a family

wise error rate (see Table 3). Once the correlation analysis was computed, it shwed that only

talkers 6 and 14 (who had big benefit in clear speech) had an increase in the size of correlation

between distance to the nearest vowel and standard deviation from the analysis for

conversational speech to the analysis for clear speech. For talkers MO15 and MO20, who

showed no clear speech benefit, there was no increase in the strength of the correlation between

distance from neighbor vowel and size of standard deviations. These data suggest, therefore, that

at least a portion of the improvement in talkers' intelligibility from conversational to clear speech

can be explained by their ability to minimize variability in their productions for vowels that are

located close to other vowels in the vowel space. Such talkers must have good innate awareness

of the vowel space, the target trajectory of each vowel and the ability to precisely control their

articulations from one repetition to the next.

Table 3. Pearson's r and p values (adjusted in parentheses) for the correlation between standard
deviation and distance to the nearest neighbor vowel for each talker and speaking style.

Talker	Conversational		Clear	
	r	p (adj)	r	p (adj)
MO06	0.617	0.014 (0.112)	0.686	0.005 (0.040)
MO14	-0.090	0.75 (6.00)	0.908	0.000 (0.000)
MO15	0.481	0.07 (0.560)	0.427	0.11 (0.896)
MO20	0.569	0.03 (0.216)	0.531	0.04 (0.336)

Figure 2 shows scatter plots of the data for these analyses, with distance to the nearest (lower F2/F1) neighbor plotted on the y axis and size of the standard deviation for the vowel in question plotted on the x axis.Filled symbols indicate conversational speech vowels and open symbols indicate clear speech vowels.   From the figure, a trend towards a positive correlation is clearest for talker MO06. For talker MO14, on the other hand, a positive trend is evident more for the clear speech data than for the conversational speech data, as indicated by the correlational analyses.
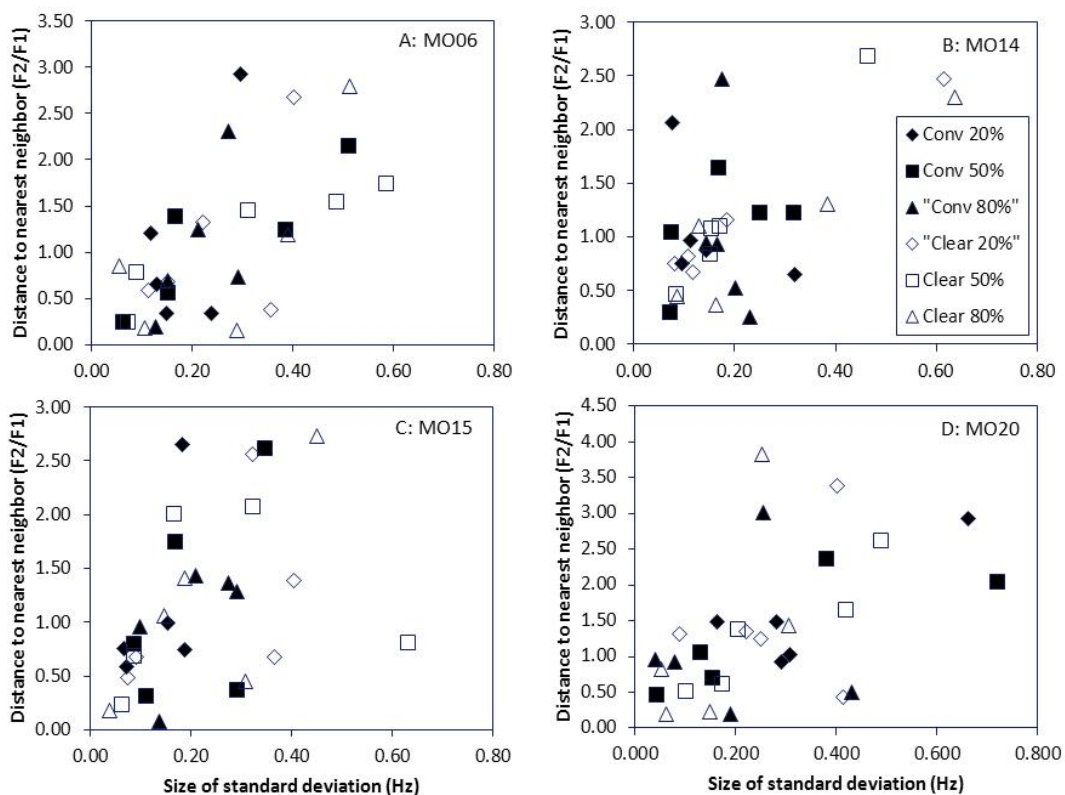

Figure 2. Scatter plots showing standard deviation (x-axis) and distance to nearest neighbor vowel (y-axis) for conversational (filled symbols) and clear speech style (open symbols). Data are plotted separately for each talker (panels A-D), with data for the two talkers who showed a significant improvement in intelligibility from conversational to clear speech in the upper two panels and data for the two talkers who showed no improvement plotted in the lower two panels.

IV. Discussion

In summary, the original hypothesis appears to be incorrect. In order to produce an increase in clear speech benefit, one does not need to have a decrease in standard deviation from conversational to clear speech for all vowels. Instead it appears that when an individual increases his or her intelligibility from clear to conversational speech there is a decrease in standard deviation only for vowels that are crowded together in the vowel space. When vowels are not crowded in the space, a decrease in standard deviation is not necessary due to the fact that the two vowels would be less likely to be confused for one another. These findings, along with previous studies' results on what specific strategies a talker utilizes to make himself or herself more clear, are important for a variety of different reasons.

First, information regarding how speakers make themselves clearer is both practical and important because if there is only limited variability in how individuals achieve clear speech, this information can be used to help with understanding not only production of speech, but also the perception of speech. If one knows what specific strategies individuals implement to make themselves clear, the information could be used to help hearing aid manufacturers' program hearing aids based on these specific strategies, which would improve the quality of hearing for hearing aid users. Also, once the same vowel analysis is conducted on clear and conversational speech of non-native English speakers, differences and similarities between native and non-native production and perception could be better understood. Greater understanding can lead to new methods of helping second language learners with their production and perception of the English Language. It is due to these reasons that continued research in the areas of speech production and perception are crucial.

References

Bianchi, M. (2007). Effects of clear speech and linguistic experience on acoustic characteristics of vowel production. University of South Florida, Tampa, Florida.

Ferguson, S. (2004). Talker differences in clear and conversational speech: vowel intelligibility for normal-hearing listeners. *Journal of the Acoustic Society of America*, 116, 2365-2373.

Ferguson, S., & 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.

Pickett, S. & Strange, W. (1999). Perception of vowels: Dynamic constancy. In A. Dragin (Ed.), *Speech decoding by human and machine: From sound stream to words* (153-165). Boston: Allyn & Bacon.

Small, L. (2003). Fundamentals of Phonetics: A practical guide for students. Boston: Pearson.

Raphael, L., Borden G., & Harris, K. (2011). *Speech science primer: Physiology, acoustics, and perception of speech*. Philadelphia: Lippincott Williams & Wilkins.

Rogers, C., DeMasi, T., & Krause, J. (2010). Clear and conversational intelligibility of /bVd/ syllables produced by native and non-native English speakers. *Journal of the Acoustic Society of America*, 128, 410-423.