Spring 2018

Conversational speech characteristics during entrainment

Pamela Molnar
James Madison University

Conversational Speech Characteristics

During Entrainment

An Honors College Project Presented to

the Faculty of the Undergraduate

College of Health and Behavioral Studies

James Madison University

by Pamela Florence Molnar

May 2018

Accepted by the faculty of the Department of Communication Sciences and Disorders, James Madison University, in partial fulfillment of the requirements for the Honors College.

FACULTY COMMITTEE:

Project Advisor: Christina Kuo, Ph.D.
Assistant Professor, Communication Sciences and Disorders

Reader: Susan Ingram, Ph.D., CCC-SLP
Assistant Professor, Communication Sciences and Disorders

Reader: Jaime Lee, Ph.D., CCC-SLP
Assistant Professor, Communication Sciences and Disorders

HONORS COLLEGE APPROVAL:

Bradley R. Newcomer, Ph.D.,
Dean, Honors College

PUBLIC PRESENTATION

This work is accepted for presentation, in part or in full, at The Spring 2018 Honors Symposium on April 18, 2018.

Table of Contents

List of Tables

List of Figures

Abstract

This study examined the acoustic characteristics of conversational speech associated with entrainment, which is the tendency for communicative behaviors of individuals engaged in a given communication context to become alike (Borrie & Liss, 2014). The study adopted a within-speaker approach to evaluate changes in speech production characteristics relative to the given individual, defined as the ***repeated speaker***. Across experiment sessions, the repeated speaker interacted with different communication partners, who were defined as the ***non-repeated speakers***. In each session, the ***repeated speaker*** and one ***non-repeated speaker*** engaged in a series of tasks in the following order: conversation, interactive picture description task, card game, interactive picture description task, and conversation. The two conversation tasks, one at the beginning of the session when the speakers began to interact and the other at the end of the session after a period of interactions, were examined. The placement of the conversation tasks was meant to allow for the evaluation of conversational speech characteristics of the ***repeated speaker*** when entrainment with the communication partner was hypothesized to be minimal and after an opportunity for entrainment to occur. This study included one ***repeated speaker*** and two ***non-repeated speakers***. The following three categories of measures were examined in this study: speech timing (rate, duration, and pauses), spectral information (vowel formant values and vowel space), and prosody, primarily fundamental frequency (F0) (mean F0, and F0 range). Results showed a difference in the ***repeated speaker's*** speech timing measures, vowel space area, and F0 measures across the two conversation tasks in the interactions with both ***non-repeated speakers***. In addition to the potential effects of entrainment, task effects and effects of familiarization were considered as well.

Introduction

Entrainment, defined by Borrie and Liss (2014, p. 1) as "conversational partners naturally adapt[ing] their verbal and nonverbal actions to more closely resemble one another," is essential for effective communication because it allows for conversations to flow smoothly, resulting in fewer disruptions and breakdowns during human interactions (Borrie & Liss, 2014). Gill (2012) considers entrainment to be an essential part of successful conversations and even necessary for our survival as social creatures because it allows for one to connect more fully with their communication partners. Researchers have also referred to entrainment as *synchronization* (e.g. Louwerse et al., 2012), *coordination*, and *alignment* (e.g., Levitan & Hirschberg, 2011; Cummins & Port, 1998; Babel, 2009). The phenomenon of entrainment has been observed in human interactions including gestures (e.g., Oben & Brone, 2016), behavior matching (e.g., Louwerse et al., 2012), syntactic (e.g., Branigan, Pickering, Cleland, 1999) and lexical (e.g., Oben & Brone, 2016) coordination, fundamental frequency (e.g., Manson et al. 2013), rhythm (e.g., Borrie, Lubold, & Pon-Barry, 2015), timing (e.g., Fusaroli, Raczaszek-Leonardi, 2014), and accent (Babel, 2009).

From the perspective of language sciences, researchers have examined entrainment of linguistic characteristics, such as syntactic and lexical entrainment. Branigan, Pickering, and Cleland (1999) found that when speakers were presented with a specific syntactic form, or sentence structure, the speakers frequently repeated that form in the next interaction. In addition to syntactic entrainment, Oben and Bronte (2016) found that when two communication partners interacted with each other the speakers used the same word, to refer to an object and even used the same gestures as their communication partner. Pickering and Garrod (2004) proposed that entrainment is an automatic process and that when communication partners synchronize at one level, say syntactically, they will likely synchronize at other levels as well, such as lexically and gesturally (Pickering & Garrod, 2004).

Researchers have also examined entrainment with regard to speech characteristics. Fusaroli, Raczaszek-Leonardi, and Tylén (2014) point to the importance of coordination of pace and rhythm to allow for conversations to flow smoothly. Also related to the timing aspects of speech production, Cummins and Port (1998) studied the synchronization of speech rhythm with synthetic auditory stimuli, or a series of beeps at different tones. They found that none of their participants synchronized their speech with that stimulus. In another study, Manson et al. (2013) studied rhythm and pace by examining mean syllable duration in conversational speech. They found that mean syllable duration did converge throughout the interaction. Additionally, Street (1984) studied speech convergence during interviews and found that, on average, the speech rate of both conversation partners converged. The contrasting results of the Cummins and Port (1998) study with the Manson et al. (2013) and Street (1984) studies may suggest that speech signals, as opposed to synthetic stimuli, encourage natural and efficient entrainment that facilitates smooth conversation and limits breakdowns. Importantly, this also shows that further investigation into the effects on timing aspects of speech in entrainment is necessary.

Suprasegmental characteristics of speech, such as fundamental frequency (F0), have also been of interest in entrainment research. Babel and Bulatov (2012) examined the importance of F0 in the ability of participants to entrain to recorded speech stimuli. Two groups of speakers were presented with single word speech stimuli, one group receiving the unmodified stimuli and another receiving the same stimuli with a high-pass filter at 300Hz. Findings indicated that participants had a tendency to produce an F0 closer to that of the model speaker in the unmodified condition and deviated from the model speaker's F0 in the filtered condition. In another study, Cummins (2009) altered the availability F0 and other rhythmic features of prerecorded speech to examine the role of F0 and other cues for speech synchronization. Cummins (2009) found that while F0 is not necessary for synchronization, when more frequency information is available, especially when other cues were degraded, the speaker displayed improved synchronization with the model talker. Manson et al. (2013) examined F0 of same-sex triads interacting in conversations about topics of their choice for 10 minutes. In comparison to previous findings, this study did not find a significant relationship between the F0 of the female participants and found that, the male participants' F0 diverged from that of their communication partners. That is, if one male partner had a high F0 as compared to his communication partners, they generally maintained their respective F0s throughout the conversation (Manson et al., 2013). The differences presented above indicate a need for further investigation of synchronization of F0, particularly throughout conversational interactions.

In addition, more fine-grain spectral characteristics, such as those associated with vowels, should be considered as well. Dialect is one sociolinguistic factor that may be associated with fine-grained phonetic changes, such as vowel production changes, during entrainment. Babel (2009) examined vowel space associated with entrainment and showed that, when participants with a New Zealand dialect were directed to repeat words after a recording of a model speaker with an Australian dialect, their vowel spaces converged with the model speaker's vowel space. However, not all vowels converged to the same extent. For example, the vowel /ɛ/ found in the word "dress" was imitated more closely to the Australian accent than the vowel /ɔ/ found in the word "thought." Nonetheless, little is known regarding entrainment of spectral characteristics of speech production.

Human communication is complex, multidimensional, and dynamic, with communication partners constantly adapting to meet the needs of their communication partner (e.g. Gill, 2012; Borrie & Liss, 2014). Verbal and nonverbal characteristics associated with entrainment have been studied, but limited information is known about variations in speech production characteristics, particularly fine-grained spectral characteristics, during entrainment in interactive conversations. The current study examined speech acoustic characteristics in conversations associated with the context of entrainment and adopted a within-speaker approach to evaluate potential effects of entrainment for a given individual, who was defined as the ***repeated speaker***. Specifically, potential effects of entrainment on speech timing (e.g., rate, duration, and pauses), spectral information (e.g., vowel formant values and vowel space), and prosody (e.g., mean fundamental frequency (F0), and F0 range) in conversations were examined because they capture a range of speech characteristics. The ***repeated speaker*** (i.e. speaker coded as “r” speaker) interacted with multiple speakers, defined as ***non-repeated speakers*** (i.e. speakers designated as “nr” speakers), in one-to-one interactions through various tasks which were meant to elicit and facilitate communication. Data from the first conversation (henceforth *Conversation 1*), and the last conversation (henceforth *Conversation 2*) during each interaction were analyzed. The two conversation tasks, occurring toward the beginning (*Conversation 1*) and end (*Conversation 2*) of the interactions, were meant to allow for the evaluation of conversational speech characteristics of the ***repeated speaker*** when entrainment with the communication partner (i.e., ***non-repeated speaker***) was hypothesized to be minimal and after an opportunity for entrainment had occurred. Additionally, the within-speaker approach was chosen because this study, unlike many others, examined the potential effects of entrainment throughout conversations without a model (e.g., digital recording as in studies reviewed above) for the speakers to entrain to. Thus, the within speaker approach allowed for the examination of speech production changes relative to the same individual. That is, the ***repeated speaker*** served as her own reference.

This study seeks to provide a more comprehensive picture of entrainment in conversations and to answer the following questions:

1. How do a given individual's speech production characteristics vary over time during conversation?
2. What are the potential effects of entrainment on speech characteristics?

This line of work also has potentially important clinical implications concerning clinician behaviors that facilitate effective communication. This information may also contribute to further theoretical development of speech production.

Design and Methods

This study was approved by the Institutional Review Board (IRB) at James Madison University (JMU). As discussed earlier, the study adopts a within-speaker repeated measure approach. Table 1 outlines the design of the study. Details will be discussed in the following sections.

Table 1. Design and ordering of tasks for each session

Timing	Speaker(s)	Events/Tasks
Immediately before session	***Repeated speaker***	1. Consent 2. Record reading tasks: o Caterpillar Passage (Patel et al., 2013), Zoo Passage (Kuo & Weismer, 2016), and list of sentences
In session	***Non-repeated speaker***	1. Consent
	Non-repeated speaker	2. Reading tasks o Caterpillar Passage (Patel et al., 2013), Zoo Passage (Kuo & Weismer, 2016), and list of sentences
	Repeated speaker with Non-repeated speaker	3. 3-7 minute Conversation 4. 5-7 minute interactive picture description and problem-solving task (Baker & Hazan, 2011; Van Engen et al., 2010) 5. 3 minute interactive game 6. 5-7 minute interactive picture description and problem-solving task (Baker & Hazan, 2011; Van Engen et al., 2010) 7. 3-7 minute Conversation

Participants

This study used a within-speaker approach. The ***repeated speaker*** worked with participants across sessions, whereas the ***non-repeated speakers*** only participated in a single session. The data of the ***repeated speaker*** were of interest in this study.

Data were collected from one (1) ***repeated speaker*** and eight (8) ***non-repeated speakers*** to develop a database of interactive conversational speech. For the current report, data from only one ***repeated speaker*** (i.e. r7) who interacted with two ***non-repeated speakers*** (i.e. nr6 & nr8) were analyzed with consideration for the scope of the study. All participants were female native speakers of American English between the ages of 18 and 25 with no history of speech, language, and/or hearing disorder(s) by self-report. Additionally, both parents of the participant were reported to be native speakers of American English and, participants were born and raised within the same dialectic base (Clopper & Pisoni, 2006; Westbury, 1994) before coming to college. Speakers r7 and nr8's dialect bases were Virginia, and speaker nr8's dialect base was Michigan. Participants were recruited through undergraduate and graduate courses at JMU and postings around campus. Because previous research shows that sex affects speech characteristics, only females were included in this study to avoid sex differences as a possible confounding factor (Byrd, 1994).

Materials

For clarity, all the materials used in the experiment are discussed here. Their use in the experiment will be discussed further in the procedures (also refer to Table 1).

Reading materials. The Caterpillar Passage (Patel et al., 2013), the Zoo Passage (Kuo & Weismer, 2016), and a series of carrier phrase sentences containing target words (i.e., "It's a ___ again.") were used to gather baseline speech production data for each participant at the beginning of the session. The sentence lists contained 13 consonant-vowel-consonant word contexts and four to nine target words that were selected from each of the DiapixUK picture tasks (Baker & Hazan, 2011) (see appendix A for words used in the sentence readings). The target words selected from the DiapixUK picture tasks (Baker & Hazan, 2011) were strategically chosen to include the four corner vowels that make up the American-English quadrilateral vowel space.

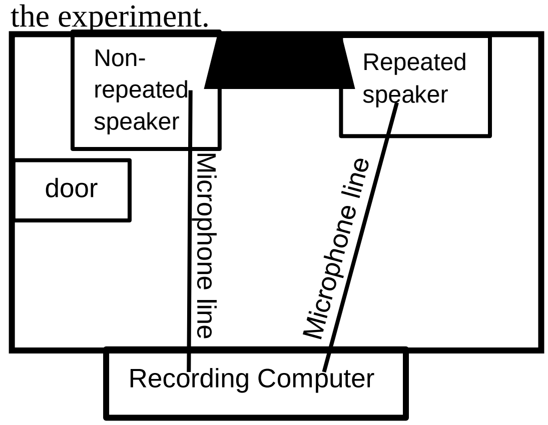
Interactive materials. The participants completed two DiapixUK tasks (Baker & Hazan, 2011) during the session. The DiapixUK (Baker & Hazan, 2011) is a "spot the difference"

(Baker & Hazan, 2011) task developed to facilitate spontaneous conversation between two individuals and is an extension of the original Diapix (Van Engen et al., 2010) picture description task. In the DiapixUK picture description task (Baker & Hazan, 2011) each participant has a picture that varies slightly from their partner's picture. The pair must work together through talking to figure out the differences between the two pictures without actually seeing the other person's picture. Additionally, a deck of cards was used to play Go Fish during the session. This game was included to help build rapport between the speakers during the session.

Procedures

Physical set-up. The experiment was conducted in the Speech Acoustics Lab in the Communication Sciences and Disorders Department at JMU. All data were recorded in a sound booth with participants seated across from each other at a table against the back wall of the booth. Figure 1 illustrates the set-up for the recordings. The table was positioned 30 inches from each side wall. Additionally, there were two chairs, one for each participant, on either side of the table that measured 10 inches from each respective wall. Participants were fitted with Shure SM10A professional unidirectional head-worn dynamic microphones which were positioned a half inch from the participants mouth and processed via a professional quality audio interface (TASCAM US-2x2) and recorded using Ableton Live Lite, a professional recording software which allows simultaneous recording of two microphone channels (one for the ***repeated speaker*** and one for the ***non-repeated speaker***). The dual channel

Figure 1. Set up of the sound booth for the experiment.



system synced the two channels during the session, which allowed for the potential need for descriptive identification of the timing of events.

Speaker tasks.

Individual speaker tasks. The individual reading tasks were meant to capture each individual's speech characteristics prior to potential entrainment effects (also see Table 1). The speakers performed the reading tasks separately (Caterpillar Passage (Patel et al., 2013), the Zoo Passage (Kuo & Weismer, 2016), and sentence readings), as discussed earlier, and followed the same setup for recording.

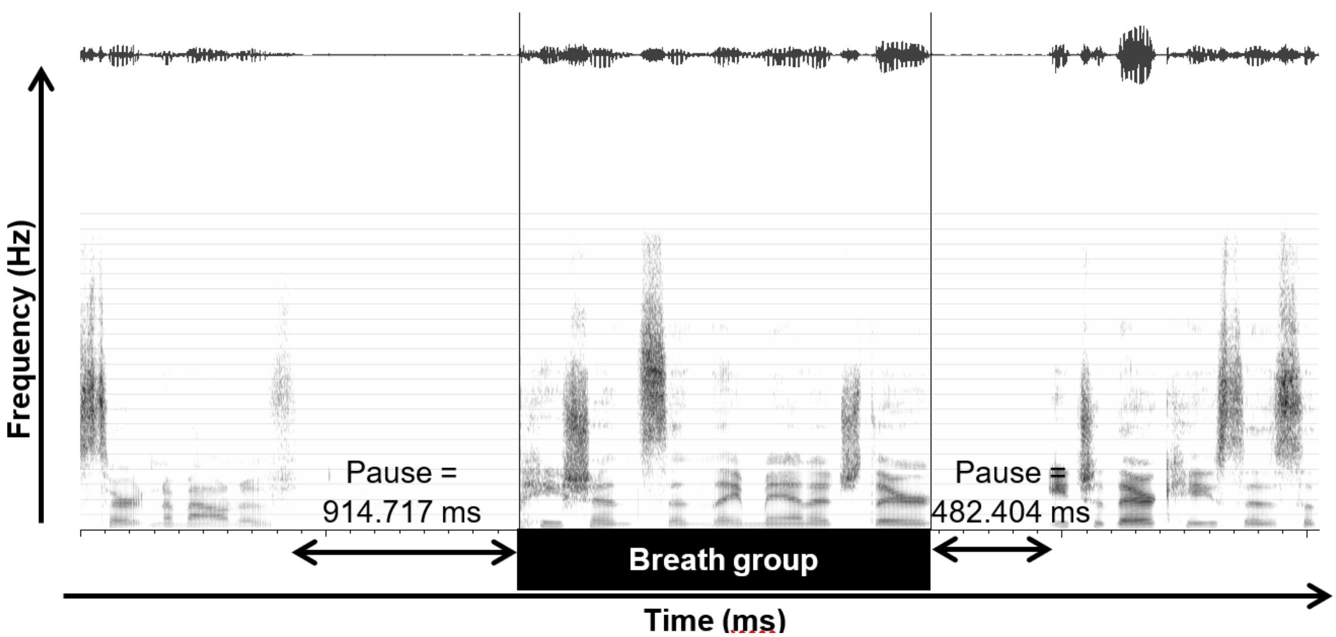
Interactive tasks. The interactive tasks began after both speakers completed the individual reading tasks. As outlined in Table 1, the conversation partners were first instructed to have a conversation for several minutes (i.e., *Conversation 1*). Second, the researcher introduced the participant pairs to the DiapixUK picture description task (Baker & Hazan, 2011). The researcher administered the instructions given from the original DiapixUK task (Baker & Hazan, 2011) and instructed the pair when to start. Third, the speakers were given cards to play a game of Go Fish. The pair played the card game for several minutes and were then instructed to complete a second DiapixUK task (Baker & Hazan, 2011). These interactive communicative opportunities were designed to allow for natural conversations and interactions that facilitated familiarization of the two communication partners. Finally, the participants were again asked to converse for several minutes (i.e., *Conversation 2*).

Measurement and Analysis

Three categories of measures were examined in this study: speech timing (rate, duration, and pauses), spectral information (vowel formant values and vowel space), and prosody (mean fundamental frequency (F0), and F0 range). Each measure was obtained in both *Conversation 1* and *Conversation 2* for comparison. A computer-based speech analysis software program Tf32 (Milenkovic, 2000), was used for measurements and analysis. For the purposes of the present study, as discussed earlier, the analysis focused on the ***repeated speaker's*** data.

Timing measures. Speech was analyzed in units of breath groups. A breath group was operationally defined as a span of continuous speech with no more than 200 ms of silence between the onset and offset (Tjaden & Wilding, 2011). Breath group onsets and offsets were identified using conventional acoustic criteria, such as the first or very last glottal pulse, stop bursts, the beginning or ending of frication, and nasal energy (Tjaden & Wilding, 2011). Any non-linguistic vocalizations, such as laughter, audible breathing, or the use of filler words, such as "um" and "uhuh" that stood alone as a single breath group were excluded from analysis. Figure 2 shows an example breath group for analysis.

Figure 2. The waveform and wideband (300Hz) spectrogram are shown for an example breath group.

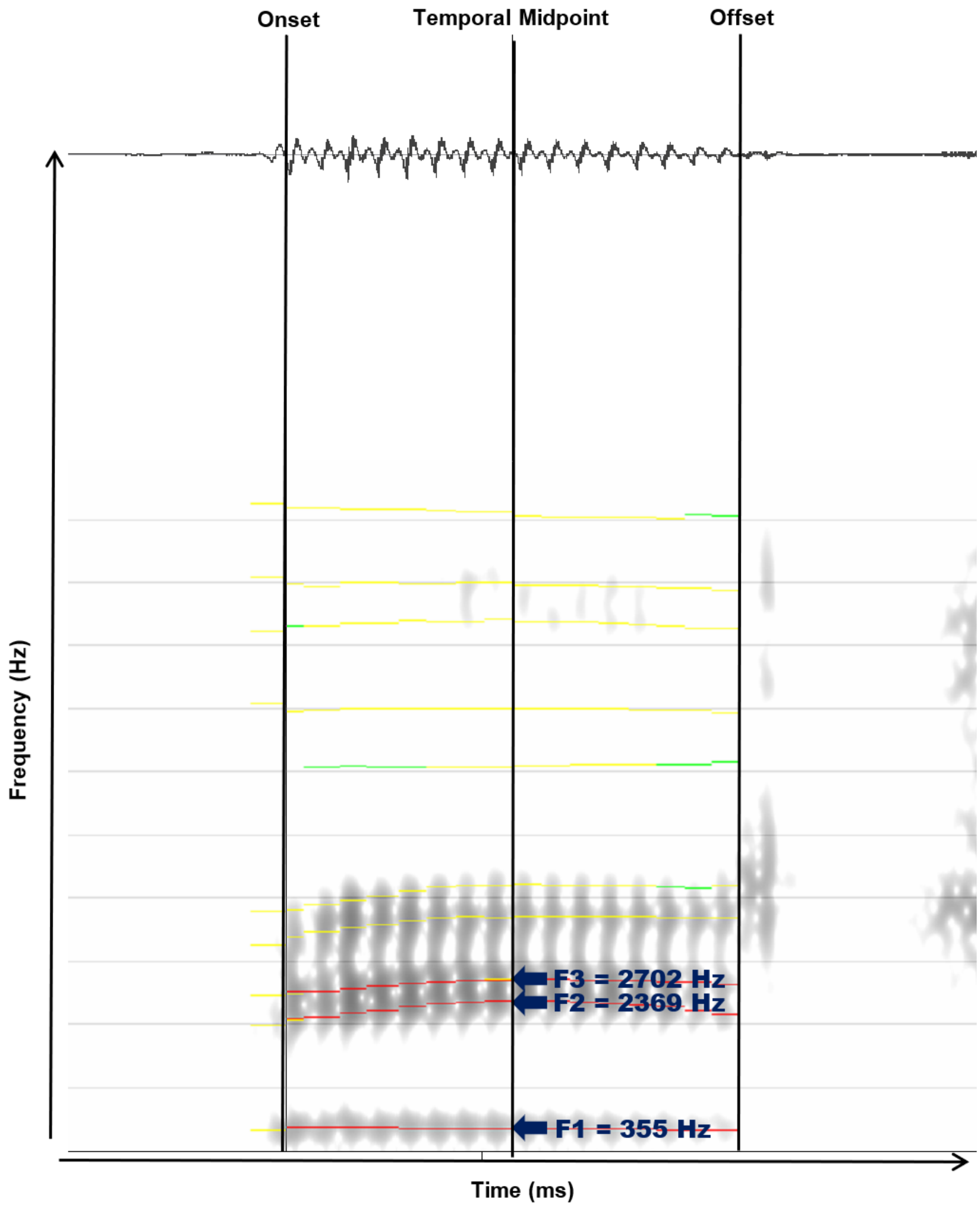


The first 100 breath groups of *Conversation 1* and *Conversation 2* for the ***repeated speaker*** from two interactions with the ***non-repeated speakers*** (i.e 400 total breath groups) were identified using wideband spectrograms and waveforms in Tf32 for analysis. Duration for each breath group was obtained, and an average was generated for *Conversation 1* and *Conversation 2*. Each breath group was then transcribed, and an Excel code was used to determine the number of words and syllables produced in the given breath groups for the calculation of rate. Articulation rate (syllables/second) was calculated by dividing the number of syllables per breath group by the duration of each breath group in seconds. Finally, an Excel code was used to calculate the average articulation rate and the average words per breath group for *Conversation 1* and *Conversation 2*.

Vowel measures. The four corner vowels, /i/, /æ/, /ɑ/, and /u/, were examined. Sixteen to twenty words produced by the ***repeated speaker*** containing the four corner vowels, approximately three to five words per corner vowel, were identified from the breath groups selected for measurement. These words are reported in Appendices C and D. Only vowels in stressed syllables were used. Additionally, four words containing the four corner vowels (one word per vowel) were identified from the sentence task for both the ***repeated speaker*** and ***non-repeated speakers*** for analysis. These words are reported in Appendices C and D.

For the vowel measures, the first three formant (F1, F2, F3) frequency values in Hertz (Hz) were manually corrected in Tf32 as needed. Temporal midpoint measures were obtained for the ***repeated speaker*** for the sentence condition, *Conversation 1*, and *Conversation 2* and for the ***non-repeated speakers'*** sentence condition. The temporal midpoint was identified by calculating the duration of the vowel and then dividing that by two (2) to find the vowel midpoint. The vowel formant frequencies were obtained at this timepoint. F1 and F2 temporal midpoint values were also used to construct vowel spaces.

Figure 3. The waveform and wideband (400 Hz) spectrogram for the vowel /i/ in the word “speech” are shown with the first three vowel formant (F1, F2, F3) frequencies for the temporal midpoint measure.



Fundamental frequency (F0) measures. For the analysis of F0, twenty breath groups from each conversation (i.e., 80 breath groups in total) were randomly selected using a random number generator (i.e., random.org). Tf32 was used to generate F0 traces for each of the selected breath groups. The breath groups were visually examined and corrected as needed on a pitch period-by-period basis (Tjaden & Wilding, 2011). Pitch traces for periods of silence, periods without clear voicing energy in the spectrogram, and/or periodic waveforms were eliminated. The average F0 for each breath group, the overall mean F0 for each conversation, and the F0 range for each conversation were calculated for analysis using Excel codes.

Results

The results are organized by the two interactions with two ***non-repeated speakers*** and in the order of speech timing measures, vowel measures, and fundamental frequency measures. This study examined the ***repeated speaker's*** speech production characteristics during *Conversation 1*, when entrainment with the communication partners was hypothesized to be minimal, and during *Conversation 2* after there had been an opportunity for entrainment to occur. The differences between the two time points (*Conversation 1* vs. *2*) were of interest.

Speech Timing Results

Interaction 1: with non-repeated speaker nr6. Table 2 reports average breath group (BG) duration in milliseconds (ms) and average articulation rate in syllables/second for the ***repeated speaker's*** (r7) interaction with the first ***non-repeated speaker*** (nr6). The r***epeated speaker's*** average breath group duration increased slightly from 1461.16 ms, with a large standard deviation of 1181.32 ms, in *Conversation 1* to 1467.42 ms, with a large standard deviation of 1366.91 ms, in *Conversation 2* (Figure 4). The average articulation rate shows a slight increase from 4.11 syllables/second, with a standard deviation of 1.32, in *Conversation 1* to 4.20 syllables/second, with a standard deviation of 1.46, in *Conversation 2* (Figure 5). The average word count decreased slightly from 6.31 words/breath group (SD = 5.75 words/breath group) in *Conversation 1* to 5.36 words/breath group (SD = 5.25 words/breath group) in *Conversation 2* (Figure 6).

Table 2. Timing measures for the ***repeated speaker*** for interaction 1: with ***non-repeated speaker nr6.***

Conversation	Average of BG Duration (ms)	Average of Artic Rate (syllables/second)	Average of Word Count (Words/BG)
1	1461.16 (1181.32)	4.11 (1.32)	6.31 (5.75)
2	1467.42 (1366.91)	4.20 (1.46)	5.36 (5.25)

Figure 4. The average breath group duration is shown for the two conversations of the ***repeated speaker*** interacting with ***non-repeated speaker nr6***.

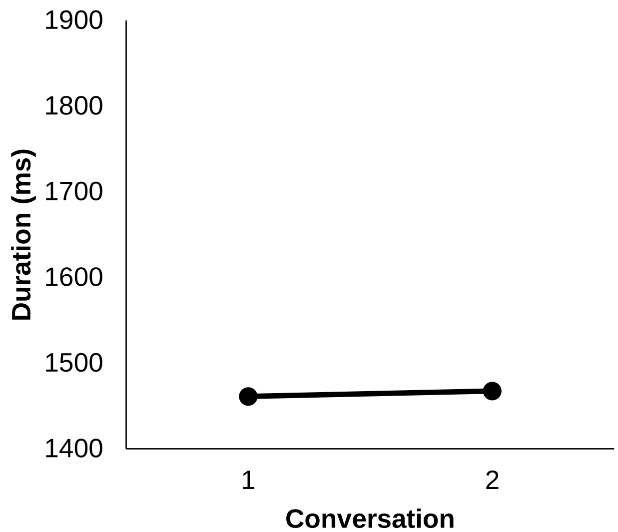


Figure 5. The mean and one standard deviation for articulation rate are shown for the two conversations of the ***repeated speaker*** interacting with ***non-repeated speaker nr6***.

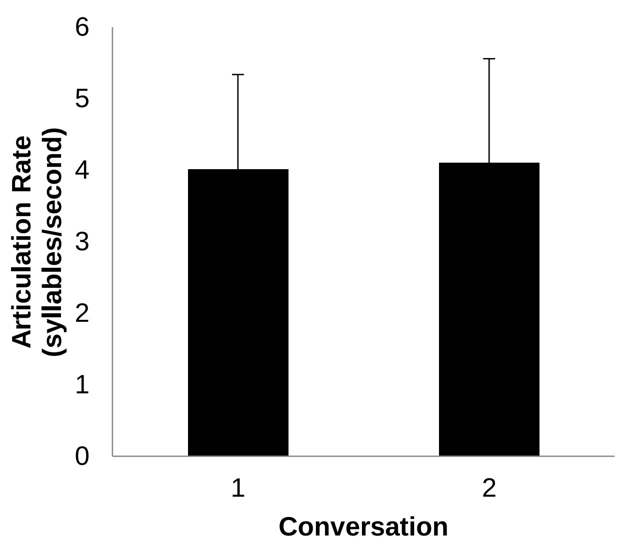
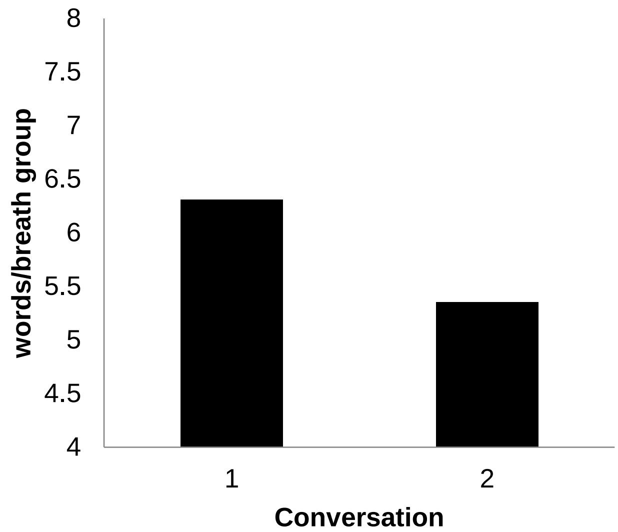


Figure 6. The average word count is shown for the two conversations of the ***repeated speaker*** interacting with ***non-repeated speaker nr6***.

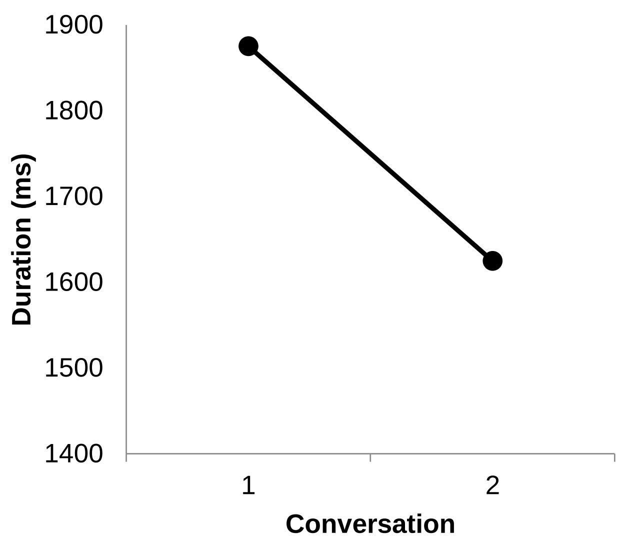


Interactions 2: with non-repeated speaker nr8. Table 3 reports average breath group (BG) duration in milliseconds and average articulation rate in syllables/second for the ***repeated speaker's (r7)*** interaction with the second ***non-repeated speaker (nr8)***. The ***repeated speaker's*** average breath group duration decreased from 1875.21 ms, with a large standard deviation of 1549.44 ms, in *Conversation 1* to 1624.79 ms, with a large standard deviation of 1318.64 ms, in *Conversation 2* (Figure 7). The average articulation rate shows a slight decrease, from 4.52 syllables/second, with a standard deviation of 1.32, in *Conversation 1* to 4.38 syllables/second, with a standard deviation of 1.51 in *Conversation 2* (Figure 8). The average word count decreased slightly from 7.26 words/breath group (SD = 6.28 words/breath group) in *Conversation 1* to 6 words/breath group (SD = 5.29 words/breath group) in *Conversation 2* (Figure 9).

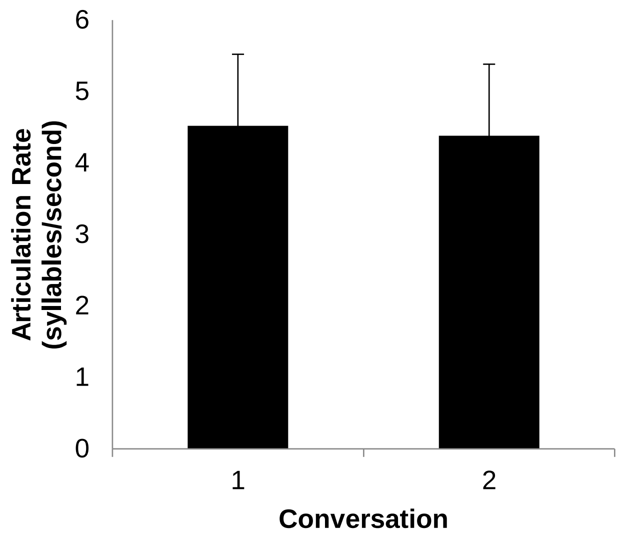
Table 3. Timing measures for the ***repeated speaker*** for interaction 2: with ***non-repeated speaker nr8***

Conversation	Average of BG Duration (ms)	Average of Artic Rate (syllables/second)	Average of Word Count (Words/BG)
1	1875.21 (1549.44)	4.52 (1.32)	7.26 (6.28)
2	1624.79 (1318.64)	4.38 (1.51)	6 (5.29)

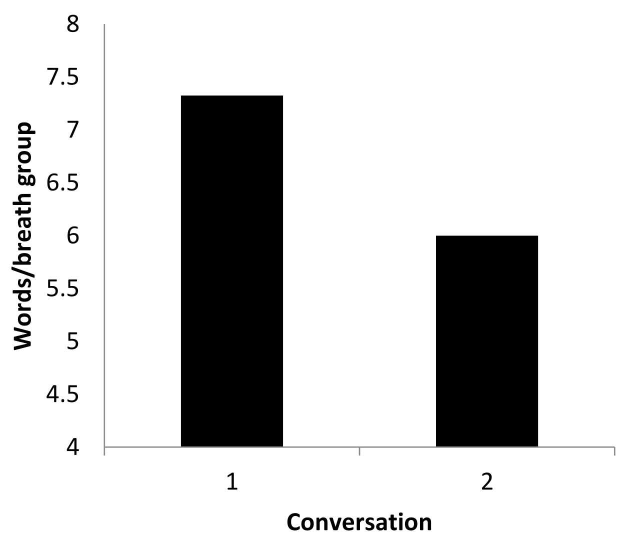
Figure 7. The average breath group duration is shown for the two conversations of the ***repeated speaker*** interacting with ***non-repeated speaker nr8***.



<u>Figure 8.</u> The mean and one standard deviation for articulation rate are shown for the two conversations of the ***repeated speaker*** interacting with ***non-repeated speaker nr8***.



<u>Figure 9.</u> The average word count is shown for the two conversations of the ***repeated speaker*** interacting with ***non-repeated speaker nr8***.



Vowel Characteristics Results

Interaction 1: with non-repeated speaker nr6. Table 4 reports the average F1 and F2 midpoint values (Hz) for each of the four vowels analyzed in the sentence condition, *Conversation 1*, and *Conversation 2*. The average F1 midpoint for /i/ was higher in *Conversation 1*, 380.00 Hz, than in *Conversation 2,* 341.00 Hz. The average F1 midpoint for /æ/ was higher in *Conversation 1,* 867.63 Hz, than in *Conversation 2,* 864.17 Hz. The average F1 midpoint for /ɑ/ was lower in *Conversation 1*, 821.67 Hz, than in *Conversation 2,* 833.67 Hz. The average F1 midpoint for /u/ was higher in *Conversation 1*, 570.33 Hz, than in *Conversation 2,* 443.67 Hz. The average F2 midpoint for /i/ was lower in *Conversation 1*, 2227.60 Hz, than in *Conversation 2,* 2321.67 Hz. The average F2 midpoint for /æ/ was lower in *Conversation 1*, 1610.75 Hz, than in *Conversation 2*, 1647.00 Hz. The average F2 midpoint for /ɑ/ was lower in *Conversation 1,* 1292.33 Hz, than in *Conversation 2*, 1480.33 Hz. The average F2 midpoint for /u/ was lower in *Conversation 1,* 1335.00 Hz, than in *Conversation 2,* 1873.67 Hz. The values presented in Table 4 were used to calculate the vowel space area (Area = 0.5 x [(F2i x F1ae + F2ae x F1a + F2a x F1u + F2u x F1i) - (F1i x F2ae + F1ae x F2a + F1a x F2u + F1u x F2i)] (Vorperian & Kent, 2007)) for the sentence condition (Figure 10) and for *Conversation 1* and *Conversation 2* (Figure 11). The vowel space area for the ***repeated speaker's*** sentence condition was 419,234 Hz^2. The vowel space area for the ***repeated speaker*** for *Conversation 1* was 199918.28 Hz^2. The vowel space area for the ***repeated speaker*** for *Conversation 2* was 121054.78 Hz^2.

Table 4. Vowel measures for the ***repeated speaker*** for interaction 1: with ***non-repeated speaker nr6***

Vowels	Average of F1 Midpoint (Hz)	Average of F2 Midpoint (Hz)
Sentence Condition		
/i/	314	2616
/æ/	1000	1629
/ɑ/	973	1309
/u/	361	1624
Conversation 1		
/i/	380.00 (85.03)	2227.60 (222.20)
/æ/	867.63 (84.50)	1610.75 (84.21)
/ɑ/	821.67 (59.58)	1292.33 (28.54)
/u/	570.33 (141.21)	1335.00 (445.63)
Conversation 2		
/i/	341.00 (27.22)	2321.67 (156.15)
/æ/	864.17 (67.93)	1647.00 (109.10)
/ɑ/	833.67 (90.18)	1480.33 (91.95)
/u/	443.67 (63.54)	1873.67 (177.97)

Figure 10. The vowel space for the ***repeated speaker's*** (r7) sentence condition before interacting with the ***non-repeated speaker*** (nr6) is shown.

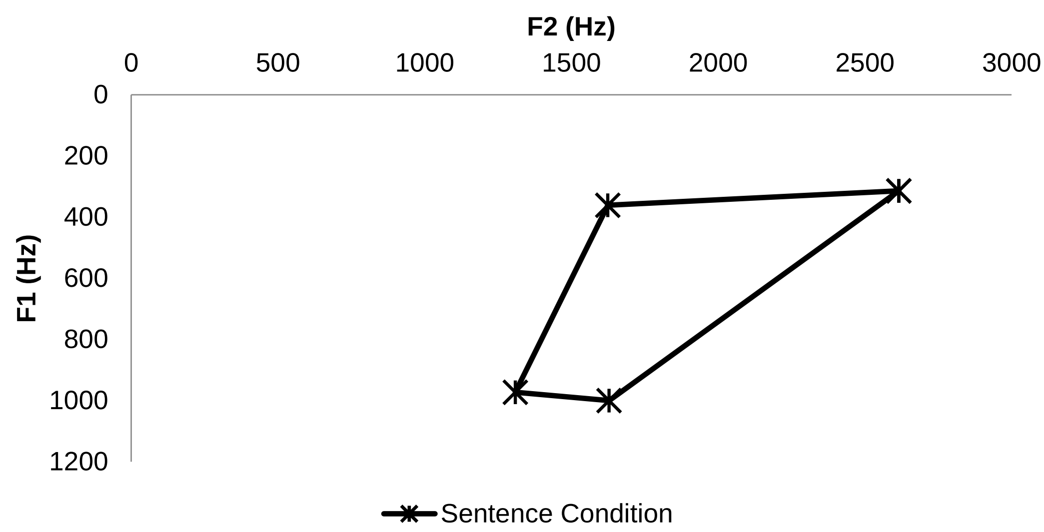
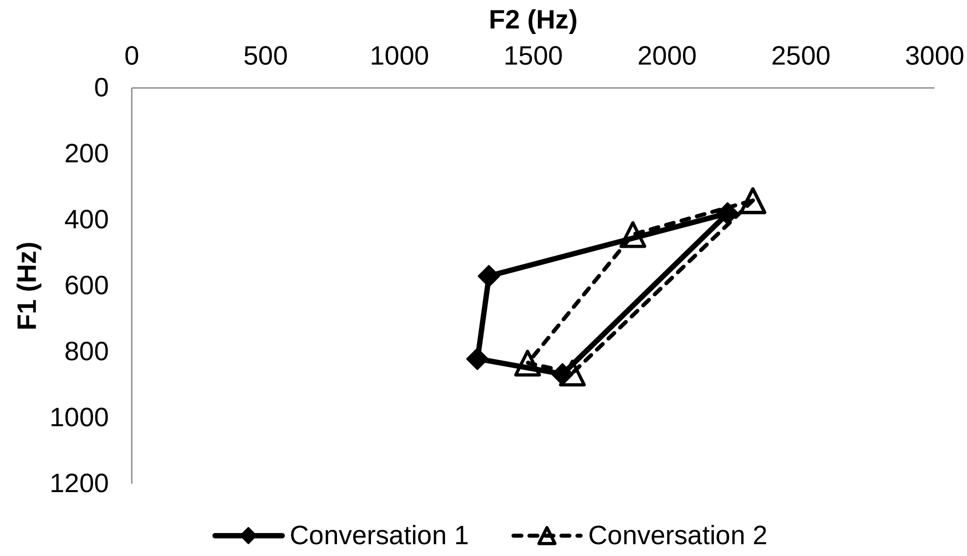


Figure 11. The vowel space for the ***repeated speaker*** (r7) is shown for the two conversations of the ***repeated speaker*** interacting with ***non-repeated speaker nr6***.



Interaction 2: with non-repeated speaker nr8. Table 5 reports the average F1 and F2 midpoint values (Hz) for each of the four vowels analyzed in the sentence condition, *Conversation 1,* and *Conversation 2*. The average F1 midpoint for /i/ was lower in *Conversation 1*, 389.57 Hz, than in *Conversation 2,* 403.6 Hz. The average F1 midpoint for /æ/ was higher in *Conversation 1*, 893 Hz, than in *Conversation 2,* 874.5 Hz. The average F1 midpoint for /ɑ/ was higher in *Conversation 1*, 897.5 Hz, than in *Conversation 2,* 862.33 Hz. The average F1 midpoint for /u/ was lower in *Conversation 1*, 446.25 Hz, than in *Conversation 2,* 452 Hz. The average F2 midpoint for /i/ was higher in *Conversation 1*, 2141.86 Hz, than in *Conversation 2,* 2114 Hz. The average F2 midpoint for /æ/ was lower in *Conversation 1*, 1602 Hz, than in *Conversation 2,* 1647.17 Hz. The average F2 midpoint for /ɑ/ was lower in *Conversation 1*, 1244 Hz, than in *Conversation 2,* 1603.33 Hz. The average F2 midpoint for /u/ was lower in *Conversation 1*, 1297 Hz, than in *Conversation 2,* 1392.5 Hz. The values presented in Table 5 were used to calculate the vowel space area (Vorperian & Kent, 2007) for the sentence condition

(Figure 12) and for *Conversation 1* and *Conversation 2* (Figure 13). The vowel space area for the ***repeated speaker's*** sentence condition was 510494.5 Hz^2. The vowel space area for the ***repeated speaker*** for *Conversation 1* was 278017.95 Hz^2. The vowel space area for the ***repeated speaker*** for *Conversation 2* was 166290.38 Hz^2.

Table 5. Vowel measures for the ***repeated speaker*** for interaction 2: with ***non-repeated speaker nr8***

Vowels	**Average of F1 Midpoint (Hz)**	**Average of F2 Midpoint (Hz)**
Sentence Condition		
/i/	297	2775
/æ/	1027	1642
/ɑ/	940	1390
/u/	333	1550
Conversation 1		
/i/	389.57 (76.07)	2141.86 (273.48)
/æ/	893 (46.08)	1602 (53.62)
/ɑ/	897.5 (13.44)	1244 (223.45)
/u/	446.25 (25.27)	1297 (510.93)
Conversation 2		
/i/	403.6 (93.99)	2114 (358.26)
/æ/	874.5 (108.36)	1647.17 (144.99)
/ɑ/	862.33 (91.53)	1603.33 (96.46)
/u/	452 (56.41)	1392.5 (581.97)

Figure 12. The vowel space for the ***repeated speaker's*** (r7) sentence condition before interacting with the ***non-repeated speaker*** (nr8) is shown.

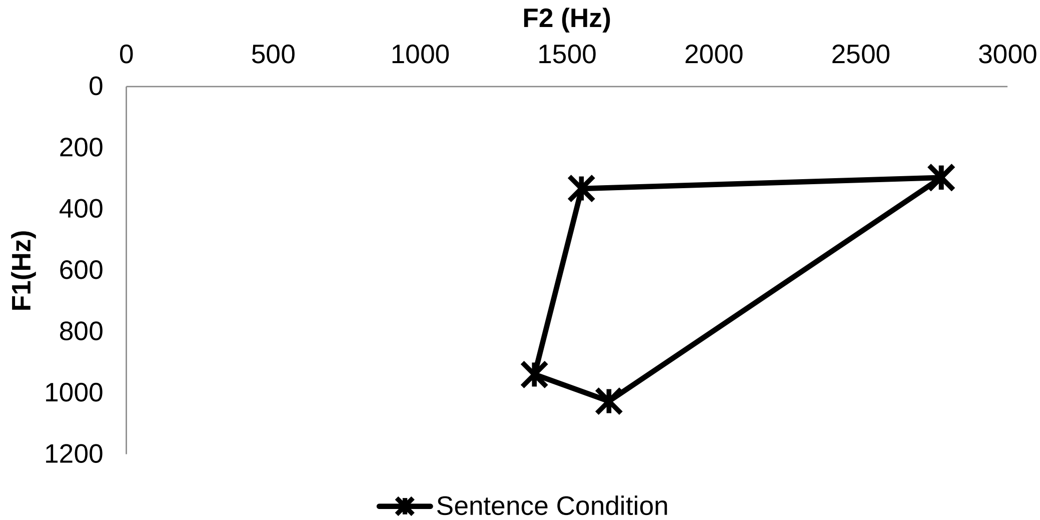
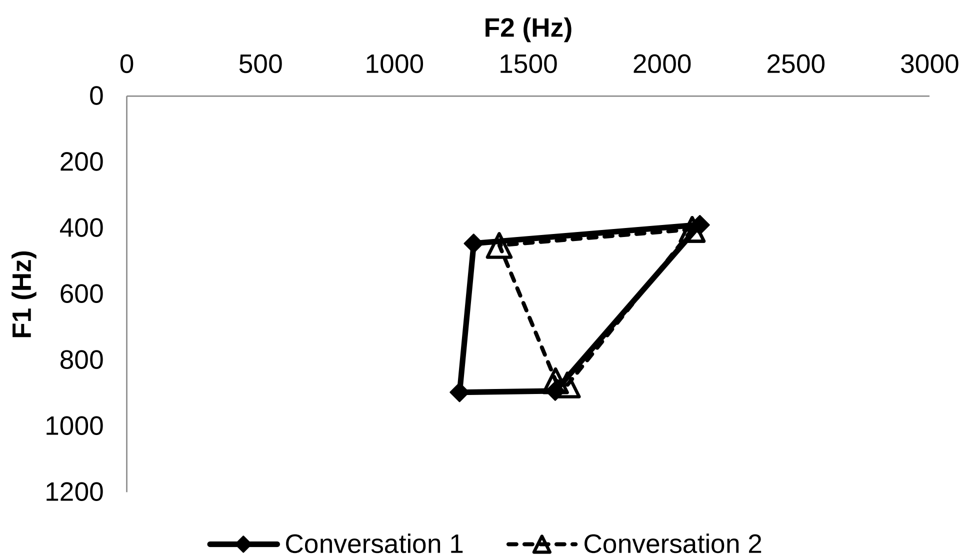


Figure 13. The vowel space for the ***repeated speaker*** (r7) is shown for the two conversations of the ***repeated speaker*** with ***non-repeated speaker*** nr8.



Prosodic Characteristics Results

Interaction 1: with non-repeated speaker nr6. Table 6 reports fundamental frequency measures for the ***repeated speaker*** (r7) for *Conversation 1* and *Conversation 2* for interaction 1: with ***non-repeated speaker nr6***. The average mean fundamental frequency was lower in *Conversation 1*, 183.49 Hz (SD = 33.99 Hz), than in *Conversation 2*, 195.77 Hz (SD = 45.34 Hz). The average maximum F0 was lower in *Conversation 1*, 235.87 Hz than in *Conversation 2*, 262.86 Hz. The average minimum F0 was higher in *Conversation 1*, 123.81 Hz, than in *Conversation 2*, 116.58 Hz. The average range in F0 was lower in *Conversation 1*, 112.06 Hz, than in *Conversation 2*, 146.28 Hz.

Table 6. Fundamental frequency measures in Hz for the ***repeated speaker*** for interaction 1: with ***non-repeated speaker nr6.***

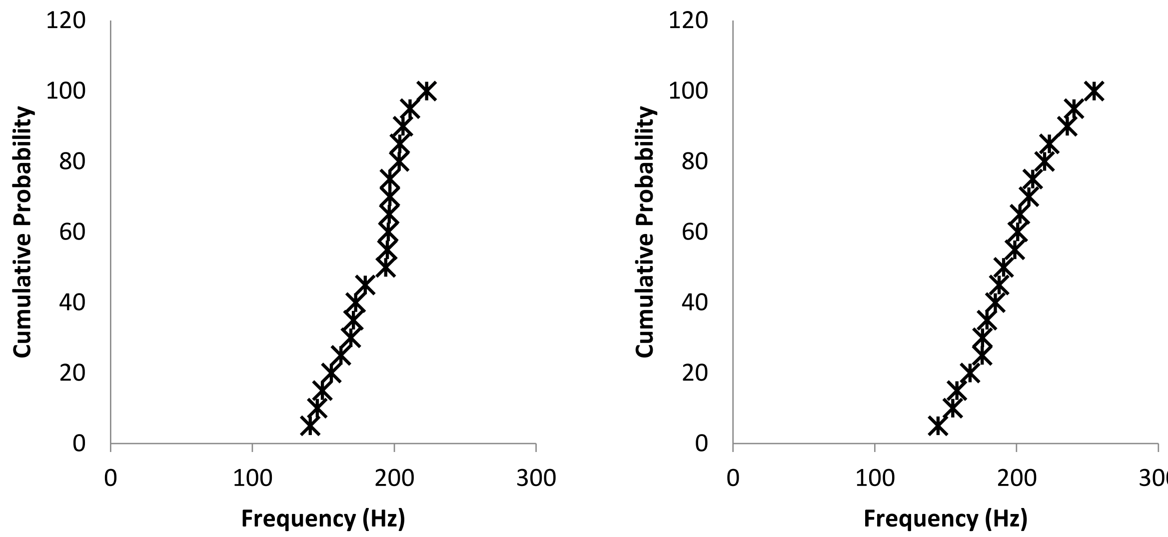
	Mean	STDev	Maximum	Minimum	Range
Conversation 1	183.49	33.99	235.87	123.81	112.06
Conversation 2	195.77	45.34	262.86	116.58	146.28

Table 7 shows the mean fundamental frequency values from 20 breath groups in *Conversation 1* and 20 breath groups in *Conversation 2* for the ***repeated speaker*** in interaction 1: with ***non-repeated speaker nr6***. The values from Table 7 are shown in Figure 14. The cumulative probability plots of *Conversation 1* and *Conversation 2* illustrate the distribution of the F0 values for the 20 breath groups sampled. The plots in Figure 14 show that, for the breath groups sampled in this interaction, the F0 distributions were quite comparable.

Table 7. Mean fundamental frequency values in Hz for 20 randomly selected breath groups and one standard deviation shown in parenthesis are shown for the ***repeated speaker*** in each conversation in interaction 1: with ***non-repeated speaker nr6***.

Conversation 1	Conversation 2
140.87 (48.20)	144.70 (39.83)
145.70 (56.28)	154.97 (43.43)
149.36 58.84)	157.87 (52.53)
155.68 (47.92)	167.26 (75.67)
162.47 (38.43)	175.70 (61.25)
169.34 (55.87)	175.96 (53.62)
171.28 (42.07)	179.22 (62.67)
172.75 (32.02)	184.99 (55.11)
179.62 (25.04)	187.67 (30.61)
194.01 (11.54)	190.93 (26.36)
195.12 (62.39)	198.70 (20.36)
195.77 (7.89)	200.73 (68.62)
196.54 (16.43)	202.51 (11.88)
196.96 (15.00)	208.76 (20.86)
197.01 (36.46)	211.39 (22.82)
203.37 (20.75)	219.80 (20.70)
203.89 (61.47)	223.19 (70.80)
206.12 (1.40)	235.88 (17.45)
211.08 (26.48)	240.48 (101.12)
222.87 (14.94)	254.66 (51.10)

Figure 14. The F0 distribution for *Conversation 1* on the left and *Conversation 2* on the right is shown for the ***repeated speaker's*** interaction with ***non-repeated speaker nr6***.



Interaction 2: with non-repeated speaker nr8. Table 8 reports fundamental frequency measures for the ***repeated speaker*** (r7) for *Conversation 1* and *Conversation 2* for interaction 1: with ***non-repeated speaker nr8***. The average mean fundamental frequency was lower in *Conversation 1*, 188.47 Hz (SD = 40.06 Hz), than in *Conversation 2*, 194.05 Hz (SD = 30.95 Hz). The average maximum F0 was higher in *Conversation 1*, 308.26 Hz than in *Conversation 2,* 241.7 Hz. The average minimum F0 was lower in *Conversation 1,* 101.69 Hz*,* than in *Conversation 2*, 129.42 Hz. The average range in F0 was higher in *Conversation 1*, 206.57 Hz, than in *Conversation 2*, 112.25 Hz.

Table 8. Fundamental frequency measures in Hz for the ***repeated speaker*** for interaction 2: with ***non-repeated speaker nr8.***

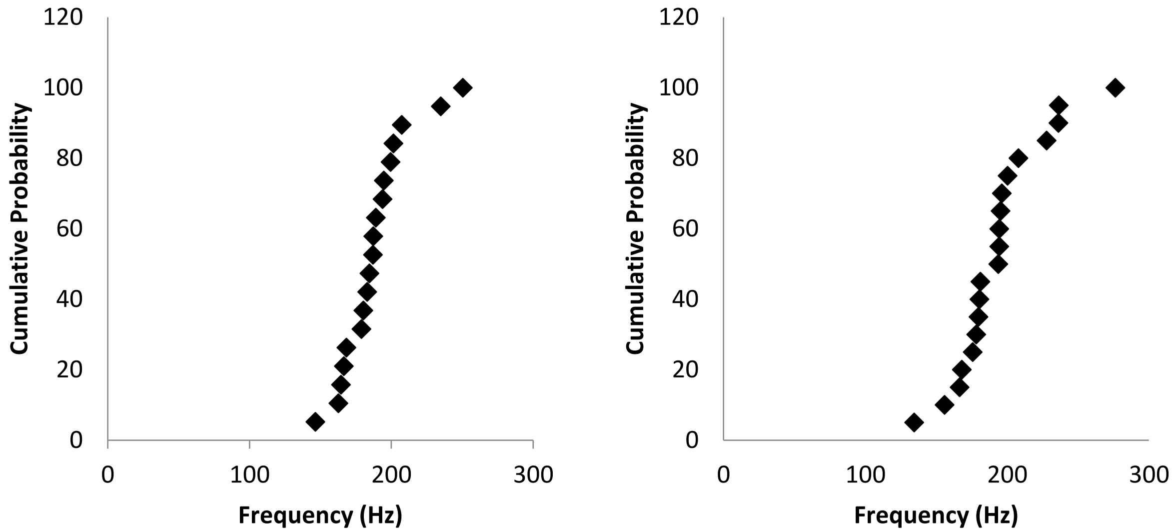
	Mean	STDev	Maximum	Minimum	Range
Conversation 1	188.47	40.06	308.26	101.69	206.57
Conversation 2	194.05	30.95	241.7	129.42	112.25

Table 9 shows the mean fundamental frequency values from 20 breath groups in *Conversation 1* and 20 breath groups in *Conversation 2* for the ***repeated speaker*** in interaction 2: with ***non-repeated speaker nr8***. The values from Table 9 are shown in Figure 15. The cumulative probability plots of *Conversation 1* and *Conversation 2* illustrate the distribution of the F0 values for the 20 breath groups sampled. The plots in Figure 15 show that, for the breath groups sampled in this interaction, the F0 distributions were quite comparable.

Table 9. Mean fundamental frequency values in Hz for 20 randomly selected breath groups and one standard deviation shown in parenthesis are shown for the ***repeated speaker*** in each conversation in interaction 1: with ***non-repeated speaker nr8***.

Conversation 1	Conversation 2
146.53 (56.20)	134.58 (45.16)
162.78 (70.29)	156.00 (53.81)
164.47 (59.99)	166.50 (48.53)
166.54 (61.82)	168.14 (48.13)
168.29 (47.85)	175.78 (36.68)
178.96 (54.82)	178.28 (33.95)
180.31 (51.03	179.91 (35.68)
182.93 (53.39)	180.56 (61.31)
184.51 (14.55)	181.12 (44.72)
186.97 (36.70)	193.91 (2.70)
187.27 (42.65)	194.49 (4.49)
189.12 (8.82)	194.49 (10.10)
193.75 (10.06)	195.48 (15.41)
194.84 (9.34)	196.39 (7.32)
199.36 (55.58)	200.41 (7.29)
201.54 (39.29)	207.96 (32.85)
207.32 (15.75)	227.92 (40.89)
234.94 (56.48)	236.35 (15.48)
250.40 (16.51)	236.37 (35.36)
	276.42 (39.24)

Figure 15. The F0 distribution for *Conversation 1* on the left and *Conversation 2* on the right is shown for the ***repeated speaker's*** interaction with ***non-repeated speaker nr8***.



Discussion

The purpose of the present study was to determine the potential effects of entrainment on a given speaker's speech acoustic characteristics using a within speaker approach. Speech timing, vowel spectral characteristics, and prosodic characteristics, specifically F0, were examined for a context when the effects of entrainment were hypothesized to be minimal (i.e. *Conversation 1*) and after entrainment had the opportunity to occur (i.e. *Conversation 2*). In general, the speaker exhibited changes in speech characteristics across the two conversation contexts. However, the patterns of these changes were complex.

Timing Measures

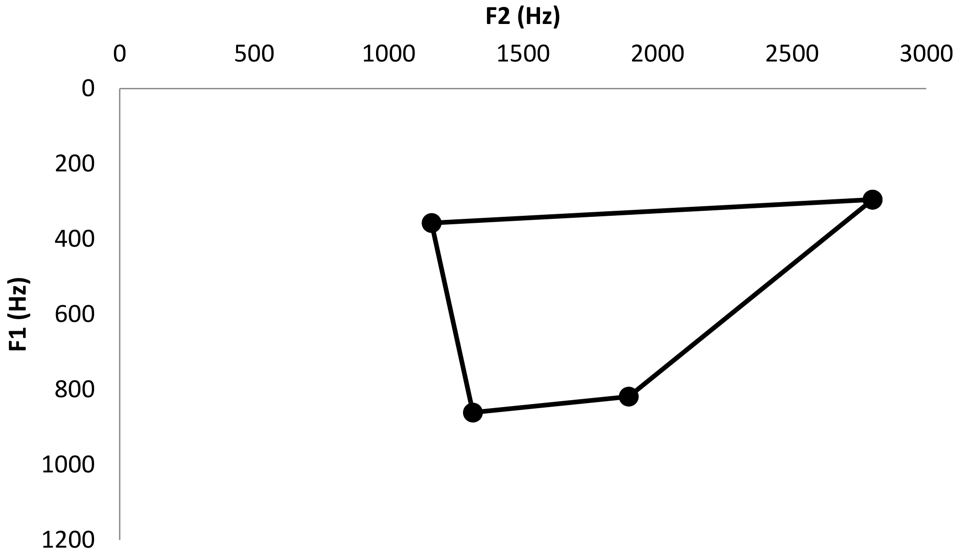
In both interactions with the two ***non-repeated speakers***, the average word count per breath group for the ***repeated speaker*** decreased from *Conversation 1* to *Conversation 2*. On the other hand, the average breath group duration and average articulation rate for the ***repeated speaker*** increased from *Conversation 1* to *Conversation 2* in interaction 1: with ***non-repeated speaker nr6*** and decreased in interaction 2: with ***non-repeated speaker nr8***. The ***repeated speaker's*** average breath group duration and average articulation rate did not change in the same way in both interactions (i.e. they increased during one interaction and decreased during the other). One hypothesis for the differences in the patterns of the timing variations could be entrainment; however, without the data on the timing measures of the ***non-repeated speakers*** it is not possible to confirm this theory and contextual effect from the conversation should also be considered as a possible explanation for these changes.

Another possible explanation for the shift in the measures may be attributed to familiarity with the communication partner. There is a continuum of speech from hyper-speech, or over exaggerated clear speech, to hypo-speech, or a more relaxed, less formal style of speech (Lindblom, 1990). Individuals have the ability to systematically vary aspects of their speech production, but are still able to be understood (Lindblom, 1990). Slower rate is generally associated with more “formal” (hyper) speech, whereas faster rate is generally associated with “casual” (hypo) speech. The changes throughout the interactions from more hyper-speech in *Conversation 1* toward hypo-speech in *Conversation 2* may suggest some level of increased familiarity for the communication partners.

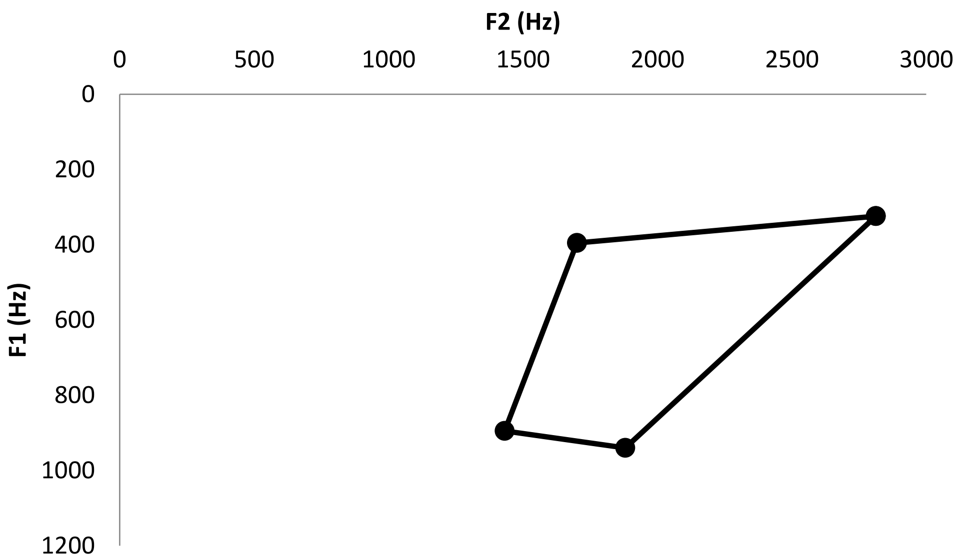
Vowel Measures

In both interactions, the vowel space area for the ***repeated speaker*** was larger in *Conversation 1* as compared to *Conversation 2.* In fact, the *Conversation 1* vowel space area was almost double the size of the *Conversation 2* vowel space area for both interactions (Figures 11 and 13). To further understand these changes, the ***non-repeated speakers’*** vowel spaces are plotted below in Figures 16 and 17. When examining the vowel space from the sentence condition for the ***non-repeated speakers*** before both interactions, the ***repeated speaker’s*** *Conversation 2* vowel spaces do not seem to be similar to the ***non-repeated speaker’s*** sentence condition vowel spaces in either interaction. In other words, the change in the ***repeated speaker’s*** vowel space from *Conversation 1* to *Conversation 2* did not converge toward the sentence reading, or habitual, vowel space of the ***non-repeated speakers.***

<u>Figure 16.</u> The vowel space for the ***non-repeated speaker, nr6's*** sentence condition is shown.



<u>Figure 17.</u> The vowel space for the ***non-repeated speaker nr8's*** sentence condition is shown.



It is also important to note that the ***repeated speaker's*** vowel spaces for *Conversation 1* in both interactions are smaller than the ***repeated speaker's*** sentence reading vowel spaces measured before each interaction occurred. The sentence reading vowel spaces, 419,234 Hz^2 for interaction 1 with ***non-repeated speaker nr6*** and 510,494.5 Hz^2 for interaction 2 with ***non-***

repeated speaker nr8, were within the range of the average vowel space area for adult females as reported by Vorperian and Kent (2007), which is 456K Hz^2.

The reduction in the vowel spaces throughout both interactions may be understood with the literature associated with speaking style variations. Ferguson and Kewley-Port (2007) demonstrated that when speakers used clear-speech, or spoke as though they were speaking with someone with hearing loss, the vowel space area expanded from the person's typical vowel space area. Additionally, Kuo and Weismer (2016) showed that conversational speech is associated with smaller vowel spaces as compared to other structured tasks. The vowel spaces of the ***repeated speaker*** in this study reduced in size as the speakers went from not interacting (i.e. the sentence condition), to interacting as unfamiliar communication partners (i.e. *Conversation 1*), to finally interacting as familiar communication partners (i.e. *Conversation 2*). The three tasks throughout the interaction can be understood on the continuum of hyper- to hypo speech (Lindblom, 1990). The sentence reading condition would represent hyper-speech because it was the most structured and controlled form of speech used in this study and resulted in the largest vowel space area. In comparison, *Conversation 2* resembles hypo-speech because the communication partners are using more casual styles of speaking as they become more and more familiar with each other and resulted in the smallest vowel space area. Finally, *Conversation 1* can be placed in the middle, somewhere between hyper- and hypo-speech.

Fundamental Frequency Measures

In both interactions, the F0 was lower in *Conversation 1* than in *Conversation 2*. Additionally, the ***repeated speaker's*** F0 variability (Figures 14 and 15) from *Conversation 1* to *Conversation 2* in both interactions seemed to be very similar. While a change was observed in the ***repeated speaker's*** F0 from *Conversation 1* to *Conversation 2* in both interactions, again it is not possible to say whether or not entrainment occurred without the ***non-repeated speakers'*** data.

Observations and Importance

It is important to note that changes were observed in both interactions across all variables examined in this study. This supports the need for further investigation into entrainment and its potential effects on speech production characteristics throughout conversation.

Limitations and Future Directions

Due to limitations of the scope of this particular study, the sample size was small. In future work on this study, analyzing data from a larger sample size and including all of the measures for the ***non-repeated speakers*** in addition to the ***repeated speaker's*** measures will allow for further examination of the potential convergence of speech characteristics. Additionally, an interesting future direction would be to include a model or deliberate change in the speaking style of one of the speakers to see if that would induce change from the communication partner. For example, if one speaker deliberately talked faster, would the communication partner speed up too?

In future directions of this study it would be important to also obtain longer speech samples to allow for the identification of more comparable phonetic environments for examination.

Conclusion

Entrainment has been studied in many other facets of human interaction (e.g., Oben & Brone, 2016; Louwerse et al., 2012; Branigan, Pickering, Cleland, 1999) and is considered essential for interactions to flow smoothly with limited disruptions and communication breakdowns (Borrie & Liss, 2014). Despite the limitations discussed earlier, this study offers a better understanding of the adaptations that communication partners constantly make throughout conversations and lays the groundwork for future studies to examine entrainment of fine-grain phonetic characteristics of speech in conversational interactions.

Acknowledgements

This project was completed as part of the Senior Honors Thesis. The author thanks the James Madison University students who volunteered to participate in this study. The author thanks Dr. Christina Kuo for her mentorship and guidance through this process. She has been an invaluable resource in this project and all of the guidance and support that she has given/continues to provide is appreciated. The author thanks Samantha King, undergraduate research assistant, for her assistance in the data collection and a portion of the analysis. The author thanks Abigail Marodi, graduate research assistant, for her assistance in the transcription of the speech samples. The author thanks Lucy Malenke for her help throughout the writing process. Finally, the author thanks her readers, Dr. Susan Ingram and Dr. Jaime Lee.

References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology*. General, 143(6), 2304-2315. doi:10.1037/xge0000021

Babel, M. (2009). Dialect divergence and convergence in New Zealand English. *Language in Society*, (4). 437. doi:10.1017/S0047404510000400

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, (2), 231.

Baker, R., & Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, 43(3), 761-770.

Borrie, S. A., & Delfino, C. R. (2016). Conversational entrainment of vocal fry in young adult female American English speakers. *Journal of Voice*, doi:10.1016/j.jvoice.2016.12.005

Borrie, S. A., & Liss, J. M. (2014). Rhythm as a coordinating device: Entrainment with disordered speech. *Journal of Speech, Language & Hearing Research*, 57(3), 815-824. doi:10.1044/2014_JSLHR-S-13-0149

Borrie, S. A., Lubold, N., & Pon-Barry, H. (2015). Disordered speech disrupts conversational entrainment: A study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. *Frontiers in Psychology*, Vol 6 (2015), doi:10.3389/fpsyg.2015.01187/full

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13-B25.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, (1/2), 39.

Clopper, C. G., & Pisoni, D. B. (2006). The nationwide speech project: A new corpus of American English dialects. *Speech Communication*, 48633-644. doi:10.1016/j.specom.2005.09.010

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 3716-28. doi:10.1016/j.wocn.2008.08.003

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26(2), 145-171. doi:10.1006/jpho.1998.0070

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *Journal Of Speech, Language & Hearing Research*, 50(5), 1241-1255.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32147-157. doi:10.1016/j.newideapsych.2013.03.005

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5), 3009-3111.

Keating, P. A., Byrd, D., Flemming, E., & Todaka, Y. (1994). Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication*, 14(2), 131-142.

Kuo, C. & Weismer, G. (2016). Vowel reduction across tasks for male speakers of American English. *Journal of The Acoustical Society of America*. 140(1), 369-383. doiL10.1121/1.4955310

Lamel, F.L., Kassel, R.H., Seneff, S. (1986). “Speech database development: Design and analysis of the acoustic-phonetic corpus”. *Proceedings of DARPA Speech Recognition Workshop*, February 1986, pp. 100-109.

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech*, Florence, Italy, 3081–3084.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35(11), 1773-1781.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. *Speech Production & Speech Modelling*, 403-439.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, 36(8), 1404-1426. doi:10.1111/j.1551-6709.2012.01269.x

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Original article: Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, 34419-426. doi:10.1016/j.evolhumbehav.2013.08.001

Milenkovic, P. (2000). Time-frequency analysis for 32-bit Windows (Version revised July 26, 2001) [computer software].

Oben, B., & Brône, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics*, 10432-51. doi:10.1016/j.pragma.2016.07.002

Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., & Russell, S. (2013). 'The Caterpillar': A novel reading passage for assessment of motor speech disorders. *American*

Journal of Speech-Language Pathology, 22(1), 1-9. doi:10.1044/1058-0360(2012/11-0134)

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral & Brain Sciences*, 27(2), 169-226.

Haahr, M. 2006. Random.org: True random number service. Web resource, available at http://www.random.org

Stevens, K. N., House, A. S., & Paul, A. P. (1966). Acoustical description of syllabic nuclei: An interpretation in terms of a dynamic model of articulation. *Journal of the Acoustical Society of America*, 40(1), 123-132.

Tjaden, K., & Wilding, G. (2011). The impact of rate reduction and increased loudness on fundamental frequency characteristics in dysarthria. *FOLIA PHONIATRICA ET LOGOPAEDICA*, (4). 178.

Turner, G., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with Amyotrophic Lateral Sclerosis. *Journal of Speech & Hearing Research*, 38(5), 1001-1013.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The wildcat corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, (4), 510.

Vorperian, H. K. & Kent, R. D. (2007). Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. *Journal of Speech, Language & Hearing Research*, 50(6), 1510-1545.

Westbury, J. R. (1994). *X-ray microbeam speech production database user's handbook*. Madison, WI: X-ray Microbeam Facility.

Appendices

Appendix A: Words used in the sentence reading condition

Final consonant		
	/d/	/t/
Vowels		
/i/		heat
/æ/	had	
/ɑ/		hot
/u/	hood	hoot
/o/	hoed	
aɪ	hide	
ɛ	head	
eɪ		hate
ɔɪ		hoyt
ʌ		hut
aʊ	how did	
ɪ	hid	

Selected Words for Four Corner Vowels				
	/i/	**/æ/**	**/ɑ/**	**/u/**
Diapix Picture Pairs				
ESM1/ESM2	beach	café, crab	xxx	blue
ESM3/ESM4	beach, seagull	shack	bar	food
ESM5/ESM6	beach, seagull	trash, taxi	rock	smoothies
ESM7/ESM8	beach, radio	shack, trash	water	tribute, shoe
ESM9/ESM10	sheep, bee	cat, tractor	shop	sue, peashoots
ESM11/ESM12	sheep	hat, cap, tractor	shop	shoot, fruit
ESM13/ESM14	bee, green	hat, dad, tractor	John	shoot
ESM15/ESM16	bee, tree	hat, at, man	xxx	sue
ESM17/ESM18	Pete's, we	café	doctor	blue
ESM19/ESM20	real	bag, trash	polish	blue
ESM21/ESM22	see, leisure, green	magical, black	spots, shop	blue, two
ESM23/ESM24	casino, celebrity	drag, fashion	shop, dog gossip	two, suit

Appendix B: Reading passages read by the speakers as part of the individual task.

<u>Zoo Passage (Kuo & Weismer, 2016)</u>

The Hoyt Aquarium and Zoo Park had a special exhibition featuring tropical lives. The **hoot** of the great horned owl could be heard meters away. Flowers of different colors surrounded the information booth where maps and guides could be picked up. In the garden with these flowers, a gardener hoed the soil to make it loose and good for new growth. A **hut** around the entrance marked the beginning of an adventure. Many children jumped up and down in excitement. There was a **head** of a fake King Kong on one of the man-made hills where monkeys rested and watched people. Next to the monkeys was the famous red panda. The panda **had** an itch on its leg and was rubbing against a small bush. It would hide around the bush, however, when too many people stood around. There were also a wide variety of sea creatures at the exhibition. The aquarium was home to thirteen sharks along with other smaller fish. The aquarium keeper explained the habitat of sharks to everyone. One shark **hid** behind some seaweed and devoured the food. A child asked, "How did the shark eat so fast?" Following the **heat** to the north side of the exhibition, one could find the "Paradise of Birds." The **hot** air was appealing to the tropical birds, said the self-guided tour. The tour notes said that it could be as hot as being under the **hood** of a running car and the birds would still like it. A couple of stunning toucan birds flew across the palm trees several times. "They hate to be watched closely," said the bird specialist. The wide variety of items at the exhibition, not just the featured tropical animals, but also the information sites and games for the youngsters attracted visitors of all ages. It was definitely a fun and educational day at the park for all.

<u>Caterpillar Passage (Patel et al., 2013)</u>

Do you like amusement parks? Well, I sure do. To amuse myself, I went twice last spring. My most MEMORABLE moment was riding on the Caterpillar, which is a gigantic rollercoaster high above the ground. When I saw how high the Caterpillar rose into the bright blue sky I knew it was for me. After waiting in line for thirty minutes, I made it to the front where the man measured my height to see if I was tall enough. I gave the man my coins, asked for change, and jumped on the cart. Tick, tick, tick, the Caterpillar climbed slowly up the tracks. It went SO high I could see the parking lot. Boy was I SCARED! I thought to myself, "There's no turning back now." People were so scared they screamed as we swiftly zoomed fast, fast, and faster along the tracks. As quickly as it started, the Caterpillar came to a stop. Unfortunately, it was time to pack the car and drive home. That night I dreamt of the wild ride on the Caterpillar. Taking a trip to the amusement park and riding on the Caterpillar was my MOST memorable moment ever!

Appendix C: List of words extracted for the target vowels, in the stressed syllable, from the *repeated speaker* during interaction 1: with *non-repeated speaker nr6*

Vowel	Conversation 1	Conversation 2
/i/	convenient here these people people	these each week
/æ/	imagine past that that's pediatric that happened have	campus have that that have have
/ɑ/	lot shock hospital	o'clock not gosh
/u/	true cool too	to too consuming

Appendix D: List of words extracted for the target vowels, in the stressed syllable, from the *repeated speaker* during interaction 2: with *non-repeated speaker nr8*

Vowel	Conversation 1	Conversation 2
/i/	she she he's feel speech speech appealing	speech be weekend week real
/æ/	past paths bad grad practice	imagine had class bad hadn't last
/ɑ/	swallowing cognitive	not gosh got
/u/	cool removed assuming school	too schools schools room