Effects of Rhythmic Consistency on Perceived Speech Effectiveness

Harrison Davis, Jenner Fox, Alex Roth

Yale University, Cognition of Musical Rhythm, Virtual Lab

1. BACKGROUND AND AIMS

1.1 Introduction [Jenner Fox]

Musical rhythm provides a framework and context through which we can understand other aspects of music. Elements like pitch, harmony, and dynamics would just be jumbled sounds without an underlying rhythmic pattern. The present study aims to investigate whether rhythm also provides a framework for understanding meaning in speech. Rhythm in speech is different from musical rhythm in that musical rhythm depends on a consistent pulse that is precise down to milliseconds, and rhythm in speech is far more loosely realized. For the purpose of this experiment, speech rhythm is defined as the consistency of time between accented syllables, or the inter-accent-interval time. Speakers who had consistent amounts of time between accented syllables were judged to be more rhythmically consistent than speakers whose times were varied.

1.2 Previous Research [Jenner Fox]

In our review of the relevant literature, we encountered several overarching themes. The first was on the connection between motor areas in the brain and speech. Using fMRI, Yuena et al. found that the areas in the brain associated with movement become active during the perception of speech, especially when speech is more rhythmic (Yuena, Davis, Brysbaerta, and Rastlea, 2010). Their findings suggest a sort of speech tempo. Munhall et al. found that people naturally move their heads when they speak, and that this head movement helps to convey linguistic information (Munhall et al., 2004). Subjects correctly identified more syllables when the natural head movements were shown, then when they were altered. In our paradigm we chose not to introduce a visual element to the speech in an attempt to further isolate the variable of rhythm.

The second pattern in the empirical findings was that there seems to be a sort of macro rhythmic structure in speech that revolves around accented syllables (Walter, 1907). Walter found that like music, accented syllables in speech are points of attention. Pitt and Samuel also found that attention is preferentially allocated to stressed syllables, and perhaps, more stressed syllables could cause heightened listening attention, and therefore more retention (Pitt & Samuel, 1990). Upon analyzing thousands of speeches for rhythmic structure, Notebloom suggests that we should focus our rhythmical analysis not on the periodically alternating stressed and unstressed elements in sound, but rather on the larger patterns of speech accents. For the purpose of our study, we do not have the technology or data to focus on macro rhythmic structures, but nonetheless, it is an interesting corollary to our study.

Finally, Jungers, Palmer and Speer discuss tempo as a coordination tool that allows for enhanced communication (Jungers, Palmer, and Speer, 2002). They compare tempo in speech to tempo in music, saying that tempo is necessary in music to allow for people to coordinate together, suggesting that tempo in speech could similarly allow people to coordinate better and understand the message being conveyed. Prosody gets defined in there article as, "perceived stress, rhythm, and intonation in spoken sentences." Their findings suggest that there are culturally preferred and primed tempos. In music primed tempos are more influential, and in speech preferred tempos are more influencial (Jungers, Palmer and Speer, 2002). Jungers, Palmer and Speer's findings suggest both that tempo in speech could help achieve coordination and that there are consistent tempos used across speakers.

1.3 Present Research [Jenner Fox]

Our research questions is: do effective speeches have more of a sense of rhythmic consistency than less effective speeches? Specifically, when listening to the same speech performed by speakers with varying levels of rhythmic consistency, how did listeners rate the effectiveness of the speeches, and was there a correlation between rhythmic consistency and perceived effectiveness? This notion of ‘effectiveness’ was based on delivery of the speech instead of content. We controlled content by exposing listeners to the same speech.

2. **METHOD** [Alex Roth]

In order to obtain our stimuli, Yale Debate Association affiliated volunteers were read the following prompt before given the speech they were to record:

"Yale Debate Team Member:

On the following page you will find a high school graduation speech. Please take a few minutes to familiarize yourself with this speech as if you were preparing to give this speech at a high school graduation ceremony. Then, while standing, give this speech as you would in a high school graduation ceremony context, and we will record you giving this speech. Please hold the recording device about one foot away from your face as you deliver this speech. If selected for use in our experiment, this audio file will be used as a stimulus in our rhythm perception experiment.

Do you have any questions?

Thank you for your participation in our experiment."

All recordings took place in a quiet music practice room in Ezra Stiles College, Yale University. The experimenter stepped out of the music practice room for the duration of the recording period, including the allotted speech familiarization period.

In order to obtain our data:

All study participants took the study through the New York University Music Perception and Cognition Experiments on the Web website. Volunteers were given the URL to the study and asked to complete the study on their own time. The tasks and procedures of this online platform are more fully elucidated upon in "1.3: Task & Procedure" below.

1.1 **Participants** [Alex Roth]

Speakers from whom recordings were obtained were drawn from a population of male members of the Yale Debate Association. Male members were asked to volunteer instead of female members in order to control for the potential confounding variable of sex of speaker impacting perceived speech effectiveness. Yale Debate Association members were commissioned due to their professional speaking backgrounds, which include much experience and demonstrated skill in giving speeches. All speakers were aged 18-21 years. Five speakers were recorded, and as a result of pretest stimuli analysis, two YDA members were dropped from the volunteer pool; recordings from the remaining three YDA members were used in order to have stimuli of evenly diverse calculated corrected values (standard deviations of inter-accent-intervals divided by corresponding averages).

Study participants were drawn from a population of Yale University undergraduates. They were of varying ages, genders and backgrounds in musical training.

Therefore, the populations to which we intend to generalize are:

1. Speakers (those having their speech effectiveness being evaluating): English-speaking, university-aged professional speakers.

2. Speech evaluators (those evaluating speech effectiveness): English-speaking, university-aged individuals.

1.2 Stimuli [Harrison Davis]

Our experimental materials involved, beyond access to an operational testing medium, exclusively the speeches used for stimuli. More specifically, we chose to use brief (approximately thirty seconds) excerpts from member Alex Roth's high school graduation speech, which involves entirely non-controversial and non-polarizing subject matter. These excerpts, the individual last three paragraphs of the speech, were recorded by three members of the Yale debate team, who are all considered experienced at public speaking. The recordings were then analyzed using the software Sonic-visualizer, the time points for rhythmic accents (decided upon with the counsel of a linguistics professor and aided by waveform information such as amplitude and length) were recorded, the average inter-accent-interval was calculated, the standard deviation of the I-A-Is were also calculated, and the coefficient of variation was finally found by dividing the standard deviation of the speech(es) by its mean I-A-I. This lengthy process resulted in a value that could report rhythmic consistency with the tempo factored in. Lower coefficients of variation were equated with higher speech consistency (for specific values of stimuli, see Analysis & Figure 1). In all, we had nine speeches from three speakers with a range of coefficients of variation which were all heard in a random order by our participants. The "three groups of speeches", identified by speech excerpt used, were the three trials in our experiment.

1.3 Task & Procedure [Harrison Davis]

Our experimental paradigm, or the specific task our participants had to perform, was to listen to nine speeches (sound files), or three groups of three speeches by three speakers (three per speaker), subsequently in a random order. The three speakers per speaker were the last three paragraphs of group member Alex Roth's graduation speech (for more stimuli information, see above). The participants were instructed, on the testing website, to:

"Click on the play button to hear the recording. [After the recoding has finished playing, above the 1-7 scale:] How effective was this speaker in communicating with his audience? Rate the speaker’s speech delivery on a scale from 1 to 7, with 1 being very poor and 7 being excellent."

These instructions were presented with each stimuli, and similar instructions were given prior to each trial:

"In this part of our study, you will be presented with three different recordings of a short section of a high school graduation speech. Each recording was recorded by a different speaker. After you listen to each recording, you will be asked to rate the speaker’s speech delivery on a scale from 1 to 7, with 1 being very poor and 7 being excellent."

As stated before, the order in which the specific speakers were heard were randomized to control for potential biases.

1.4 Data Collection & Analysis [Alex Roth]

The collection and transformation of our stimuli are described in the "Stimuli" section above.

Data was collected via the online platform aforementioned in the "Method" section above. Quantitative values for perceived speech delivery effectiveness ranged from 1 to 7 for each trial, where 1 was defined as "very poor" perceived speech delivery effectiveness and 7 was defined as "excellent" perceived speech delivery effectiveness.

In attempting to determine whether there was a correlational relationship between rhythmic consistency and perceived speech delivery effectiveness, corrected values (standard deviations of inter-accent-intervals divided by corresponding averages) were plotted against speech delivery effectiveness ratings. This analysis did not produce a line of best fit of statistical significance.

3. RESULTS

1.1 Figures 1A, 1B & 1C & Analyses: Distributions of Speech Effectiveness Rating Scorings for Three Yale Debate Team Speakers in Paragraphs 1, 2 and 3 [Alex Roth]

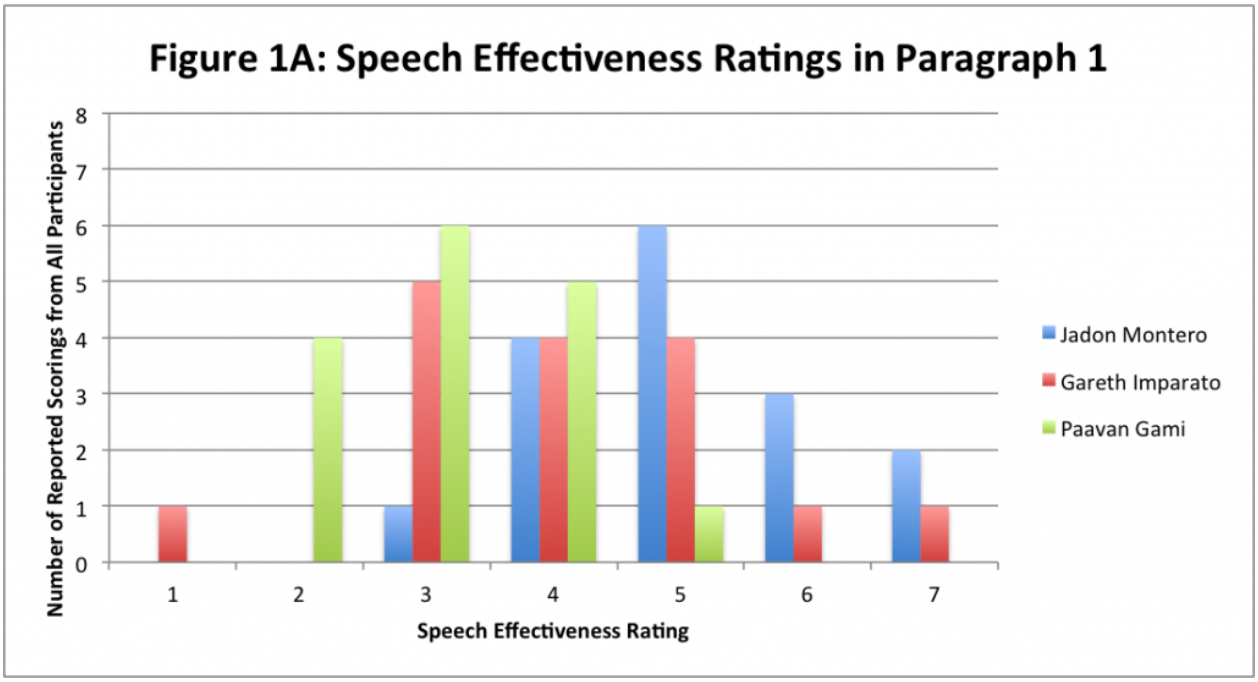


Figure 1A shows perceived speech effectiveness data for Jadon Montero's speaking in Paragraph 1 clustering around a speech effectiveness rating of approximately 5, with an approximately normal distribution; for Gareth Imparato's speaking in Paragraph 1 clustering around a speech effectiveness rating of approximately 4, with an approximately normal distribution; and for Paavan Gami's speaking in Paragraph 1 clustering around a speech effectiveness rating of approximately 3, with an approximately normal distribution. This data visually shows Jadon Montero's speaking rated most effective on average in Paragraph 1 and Paavan Gami's and Gareth Imparato's speaking rated such that their average speech effectiveness rating values are close.

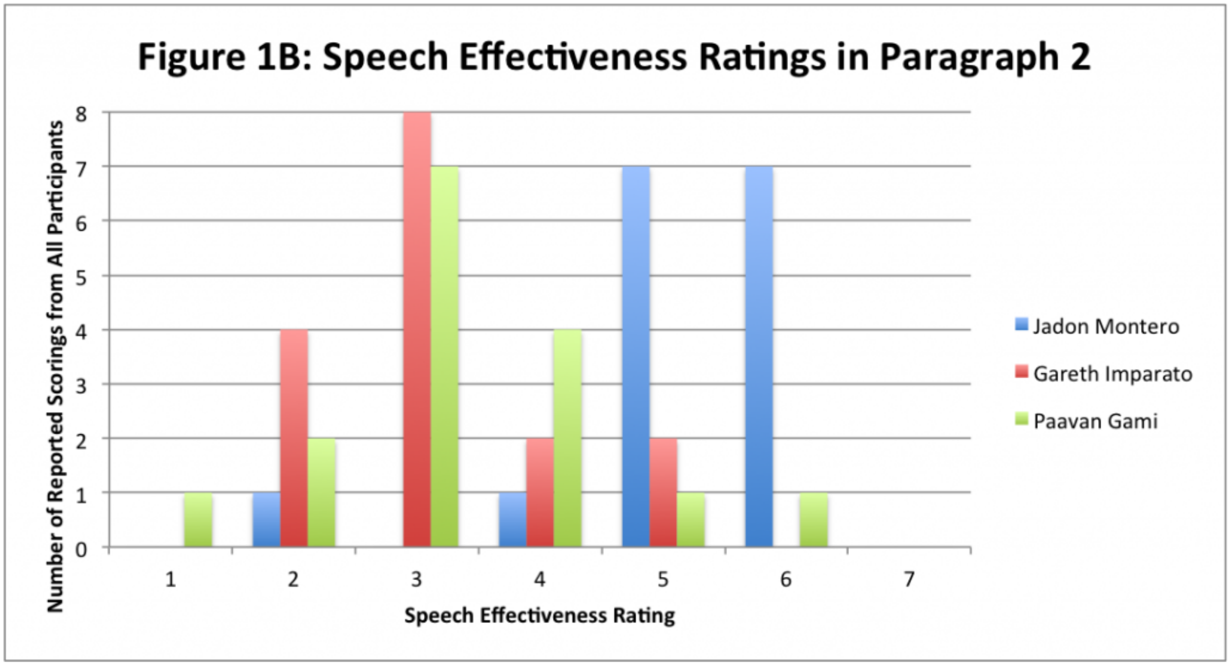


Figure 1B shows perceived speech effectiveness data for Jadon Montero’s speaking in Paragraph 2 clustering around a speech effectiveness rating of approximately 5, with an approximately normal distribution skewed to the left; for Gareth Imparato’s speaking in Paragraph 2 clustering around a speech effectiveness rating of approximately 3, with an approximately normal distribution; and for Paavan Gami’s speaking in Paragraph 2 clustering around a speech effectiveness rating of approximately 3, with an approximately normal distribution. This data visually shows Jadon Montero's speaking rated most effective on average in Paragraph 2 and Paavan Gami's and Gareth Imparato's speaking rated such that their average speech effectiveness rating values are close.

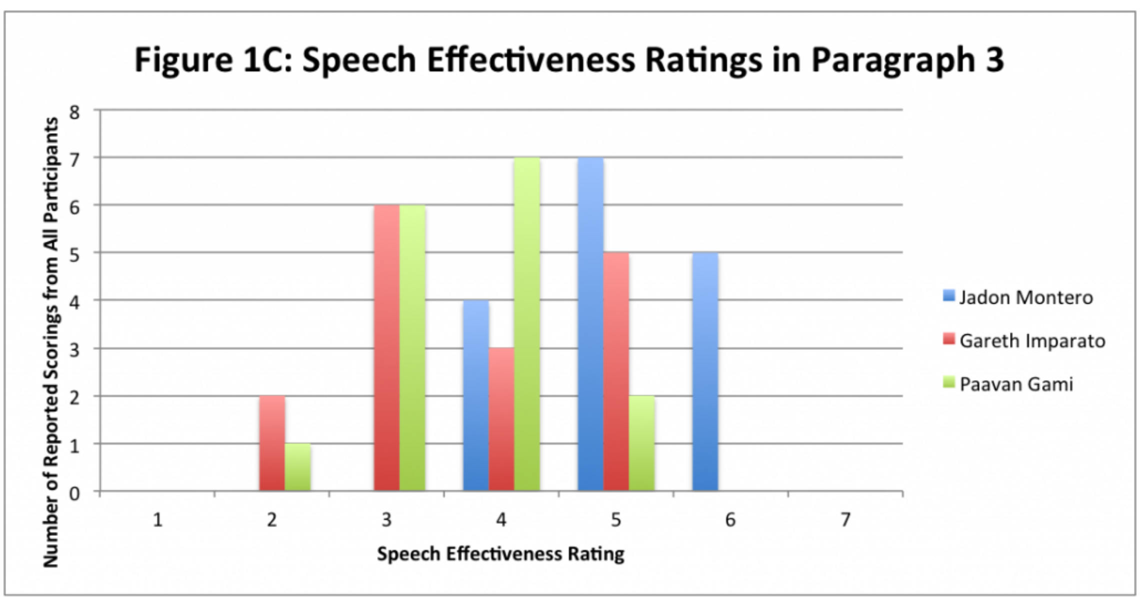
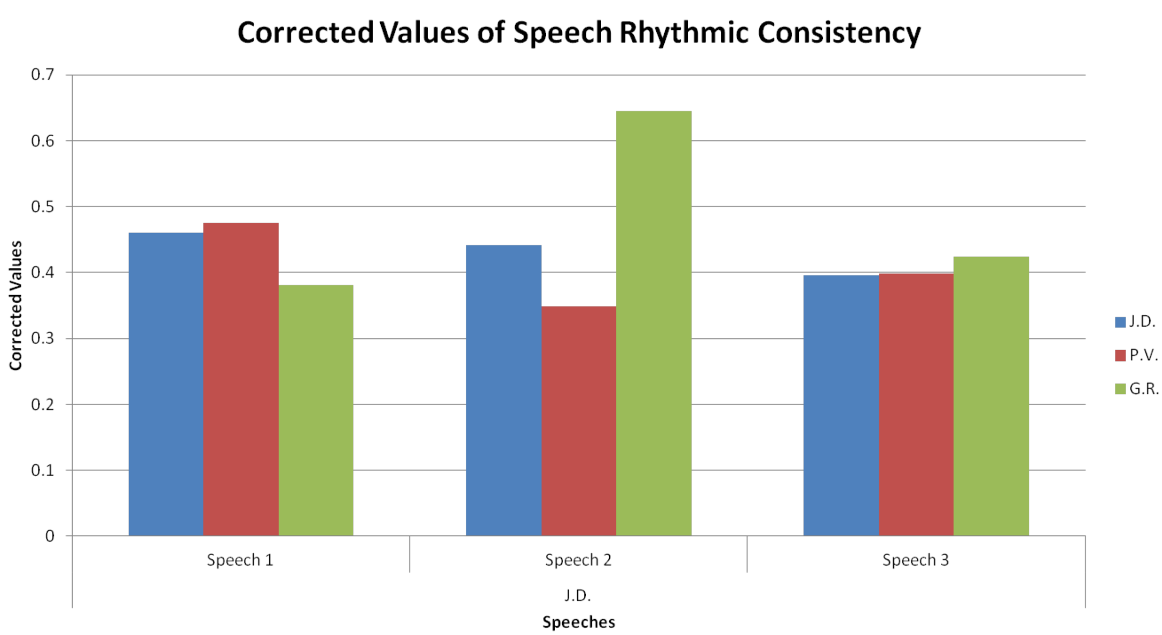


Figure 1C shows perceived speech effectiveness data for Jadon Montero’s speaking in Paragraph 3 clustering around a speech effectiveness rating of approximately 5, with an approximately normal distribution; for Gareth Imparato’s speaking in Paragraph 3 clustering around a speech effectiveness rating of approximately 4, with an approximately normal distribution; and for Paavan Gami's speaking in Paragraph 3 clustering around a speech effectiveness rating of approximately 4, with an approximately normal distribution. This data visually shows Jadon Montero's speaking rated most effective on average in Paragraph 3, with Paavan Gami's speaking being rated second most effective on average and Gareth Imparato's speaking being rated third most effective (least effective) on average.

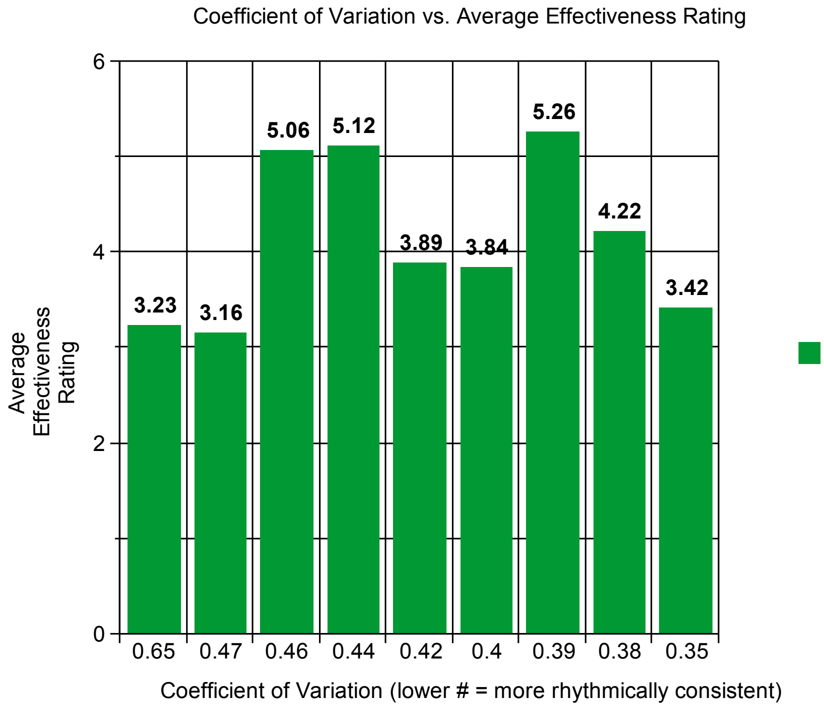
1.2 Analysis & Figure 2 [Harrison Davis]

J.D.			P.V.			G.R.		
Speech 1	Speech 2	Speech 3	Speech 1	Speech 2	Speech 3	Speech 1	Speech 2	Speech 3
0.460716	0.44138	0.395836	0.475949	0.34873	0.39912	0.381168	0.645216	0.423523



This table and graph set present what are essentially the independent variables in our experiment. Our data was interpreted along the lines of these data, which we chose to represent, as explained above (see stimuli), the concept of "rhythmic consistency". Essentially these corrected values are the standard deviations divided by the averages of inter-accent-intervals (space of time between linguistic and performance accents) in our stimuli. This function accounted for the differences in speech tempo and allowed us to generate an objective figure for judging the overall consistency of accents within the speeches. For reference, the lower the corrected value, the more rhythmically consistent the speech.

1.3 Analysis & Figure 3 [Jenner Fox]



Along the horizontal axis, from left to right, the graph shows the different degrees of rhythmic consistency, from least to most consistent. Rhythmic consistency was found by mapping the accents for each speaker in each sound clip and then finding the coefficient of variation between the standard deviations of the inter onset intervals. The vertical axis shows the average effectiveness ratings (on a scale from 1-7) for each degree of rhythmic consistency. Average effectiveness ratings were found by averaging out the effectiveness responses for each participant for each sound excerpt.

There was no pattern whatsoever in the relationship between rhythmic consistency and speech effectiveness. The only correlation that existed was between specific speakers and speech effectiveness. We also ran a test on correlation between rhythmic consistency and effectiveness ratings for individual speakers, and there were no significant results.

4. CONCLUSIONS [Alex Roth]

Our hypothesis was: "There is a positive correlation between rhythmic consistency in speeches and perceived speech delivery effectiveness."As no line of best fit of statistical significance could be drawn when rhythmic consistency was plotted against perceived speech delivery effectiveness, our study is inconclusive.

There were many limitations in our method, both inherent in our research question and execution of our study. Limitations inherent in our research question: Even if a positive correlation were established between our independent and dependent variables, it would not have been impossible to rule out potential lurking variables, i.e. variables that could have been influencing both rhythmic consistency and perceived speech effectiveness. Furthermore, with a small pool of three YDA speakers, a positive correlation could've potentially fallen into place independent of any sort of meaningful relationship between the dependent and independent variables. The second limitation in our research question was that we were not controlling for several variables, e.g. timbre of voice. We wished to be able to generalize our study to English-speaking, university-aged professional speakers in the real world, but in doing so, we sacrificed controlling for the given example variable voice timbre. In order to control for this variable we could computerize our stimuli by recording speeches from YDA members, uploading these speeches to a computer, and having these speeches be outputted in a computerized voice. In doing so we would preserve the rhythmic elements in our recorded speeches but would arguably lose the "human elements" that allow us to generalize our study to real-life speakers.

Limitations deriving from the execution of our study: Moreover, the execution of our study was less than ideal on a couple points. First, in one of our recordings, music was being played in the background in an adjacent music practice room, and we therefore were forced to drop this recording from our stimuli pool because this music could be heard in the recording. This recording corruption could be corrected for by recording stimuli in a professional audio studio or laboratory setting. However, our access to these spaces is limited, and these spaces are oftentimes expensive to use. Second, several of our stimuli contained unintentional "stumblings in speech," which very likely negatively impacted perceived speech delivery effectiveness ratings. These "stumblings" could easily be corrected for by increasing the number of recordings obtained by recruiting more YDA members for our study and/or having YDA members record the same speech multiple times in the same session. Finally, our calculated corrected values (standard deviations of inter-accent-intervals divided by corresponding averages) were not evenly diverse; this unevenness in the diversity across our stimuli could be corrected for by recruiting more YDA members for our study and/or having YDA members record the same speech multiple times in the same session (more potential stimuli to choose from) and/or more strategically selecting passages from the speech that would result in a more evenly diverse set of pretest calculated corrected values of stimuli.

In moving forward, it would be ideal to reconduct our study with the aforementioned suggested corrections: recording stimuli in a more consistently quiet space, recruiting more YDA members, having YDA members record the same speech multiple times in the same session, and more strategically selecting passages from the speech that would result in a more evenly diverse set of pretest calculated corrected values. In doing so, we would hope to evidence a correlation between rhythmic consistency (as we've defined it) and perceived speech delivery effectiveness.

REFERENCES [Harrison Davis]

YUENA, Ivan, MATTHEW H. Davis, MARC C. Brysbaerta, and KATHLEEN Rastlea. “Activation of Articulatory Information in Speech Perception.” 107.2 (2010): n. pag. *PNAS*. Web. 30 Mar.

JUNGERS, MELISSA K., CAROLINE Palmer, and SHARI R. Speer. “Time after Time: The Coordinating Influence of Tempo in Music and Speech.” *International Quarterly of Cognitive Science* 1.2 (2002): 21-35. *Cognitive Processing*. Web. 30 Mar. 2013.

MUNHALL, K.G., JEFFREY A. Jones, DANIEL E. Callan, TAKAAKI Kuratate, and ERIC Vatikiotis-Bateson. “Visual Prosody and Speech Intelligibility: Head Movement Improves Auditory Speech Perception.” *Psychological Science* 15.2 (2004): 133-37. Web.

PITT, MARK A., and ARTHUR G. Samuel. “The Use of Rhythm in Attending to Speech.”*Journal of Experimental Psychology: Human Perception and Performance* 16.3 (1990): 564-73. Web.

SCOTT, Walter D. “Rhythm in Written and Oral Discourse” *The Psychology of Public Speaking* (1907): (135-139). Print.

NOOTEBOOM, Sieb. ”The Prosody of Speech: Melody and Rhythm” *The Handbook of Phonetic Sciences*. Nr. 5 (1997) 15-35. Print.