The Effects of Time Expansion on English as a Second Language Individuals

John S. Novak, III[1], *Daniel Bunn*[1], *Robert V. Kenyon*[1]

[1]University of Illinois at Chicago, United States of America
Jnovak5@uic.edu, Dbunn2@uic.edu, Kenyon@uic.edu

Abstract

When speaking to second language learners, talkers often reduce their rate of speech to assist their listeners' understanding and comprehension. This study grants English as a Second Language subjects fine-grained, real-time control over the playback rates of lengthy audio tracks of conversational speech, and tests the subjects' listening comprehension at their desired playback speeds and at unmodified speeds. We find evidence that slower playback rates are preferred, but no evidence that such playback rates affect listener comprehension.

Index Terms: Perception of prosody; ESL; human-computer interaction

1. Introduction

The ability to actively and successfully listen for meaning and context-- in other words, to understand or comprehend the spoken language-- is crucial to the acquisition of spoken language [1] and may be the most important skill in second language acquisitions [2]. However, it may also be the most difficult of the four basic language skills (listening, reading, writing, and speaking) to learn [3]. It is also, of course, important in the day to day application of those second language skills.

However, not all speech is created equal. "Clear Speech" is an umbrella term for several related speech styles, all of which are adopted by talkers on behalf of their listeners to accommodate some form of adversity, including not only incomplete mastery of a language, but also noisy environments, hearing loss, cognitive decline, or combinations of these and related factors. The distinctions between Clear Speech and Casual Speech include (most obviously) a reduction in speech rate, modification of pitch, expansion of vowel space, and an increase in consonant to vowel energy ratio.

Previous research has shown that naturally produced Clear Speech does indeed benefit the listeners through increased intelligibility and understanding. One study [4] shows that when Clear Speech is presented in a noisy background, intelligibility improves for both native and non-native listeners. However, this study also shows that non-native listeners derive a smaller relative benefit than native listeners. A subsequent study [5] controls the listening test with high-predictability sentences (i.e., sentences whose final words are readily apparent from prior context, allowing language-proficient listeners to predict them) and low-predictability sentences (i.e., sentences whose final words cannot be anticipated.) This study shows that native listeners benefit from Clear Speech for both types of sentences, while non-native speakers derive benefit almost entirely from high-predictability sentences.

However, just as all speech is not created equal, neither are all talkers and listeners created equal. There is evidence in [6] that the benefits of Clear vs Casual Speech vary from individual talker to talker. However, [4] provides evidence of the dual, that the benefits of Clear vs Casual Speech vary from listener to listener. Statistically, Clear Speech provides intelligibility benefits, but those benefits may be highly idiosyncratic with individual talker-listener pairings.

Several studies have attempted to replicate or convert Casual Speech to Clear Speech by computer intermediation, most often by manipulating speech rate and investigating the effects on intelligibility or comprehension. Numerous studies show a deleterious effect on comprehension when speech rate increases [7, 8, 9]. The results of slowing speech are conflicting: One study [10] shows an increase in intelligibility when non-native speakers are allowed to select from one of a small number of pre-selected speech expansion rates. However, other similar experiments [11, 12] show no statistical improvement of speech comprehension.

Motivated by prior contradictory results and by evidence of idiosyncrasy, we consider the possibility that ideal speech rates may be highly personalized, especially, e.g., due to individual language proficiency. To that end we have designed and implemented an experiment to explore fine grained control of delivered speech rate, with real time responsivity. We used this tool in conjunction with TOEFL iBT tests to study non-native speakers' preferred speech rates and the effects of those preferred rates on comprehension.

2. Materials

2.1. Audio and Quiz Material

To test listening skill, we used the Official TOEFL iBT Tests, Vol. 1, comprising 106 separately prepared audio tracks with complementary multiple-choice quizzes. We selected nine of these audio tracks, such that all nine matching quizzes are of the same format: Four multiple choice questions each with one single correct selection out of four, followed by a final multiple-choice question with two jointly correct selections out of four. Questions with two answers were labeled as such.

Of these nine audio tracks, five were fictitious lectures and four were fictitious conversations. Each audio track was copied to Wav audio format, and each quiz was transcribed into a text file.

2.2. Experimental Software

To test our hypotheses that research subjects would use audio expansion, and would show increased listening comprehension as a result, we designed custom software to administer pairs of audio clips and quizzes in a way which afforded listening rate control to the research subjects.

Specifically, we developed a phase vocoder with frame-wise magnitude interpolation and frequency-wise phase advanced as described in [13, 14]. An appropriate choice of interpolation in this vocoder acts to stretch or expand audio without altering the pitch or tonal qualities of the audio our user interface used a simple on-screen slider interface so that, while playing an audio clip, moving the slider left would slow the audio down and the opposite would speed the audio up. Time expansion ranged from 1.0 to 2.5 of the original length (1.0 to 0.4 of original speed, or increasing instantaneous playback time by 0 to 150%) with 61 possible slider settings. The fine granularity of the intervals and responsiveness combined to provide a smooth interface analogous to a volume control.

We used this interface to present participants with alternating audio tracks and comprehension quizzes according to the protocol described below. In addition, subjects were free to adjust the audio playback rate throughout the applicable audio clips; these changes of the speed settings along with appropriate timestamps were stored for analysis.

3. Methods

3.1. Participants

We recruited twenty-six young (age 18 to 30 years), healthy (no self-reported diagnoses of hearing problems), English as a Second Language (ESL) individuals to participate in this study. These participants all self-reported prior TOEFL scores between 60 and 110 at a time no greater than twelve months prior to their participation date. We offered no incentive for participation.

This study was approved by the UIC Office for the Protection of Research Subjects. Participants were recruited by announcements to UIC student mailing lists and in lectures. Participants provided written statements of informed consent prior to participation.

3.2. Procedures

We installed our custom software on a Windows laptop, connected to Sennheiser HD 598SE over-ear headphones. We calibrated audio track sound levels to present audio at approximately 65 dB SPL. Participants then engaged in a four-phase experiment as described below. The researcher interviewed each subject directly after the experiment. The experiment took approximately 45 minutes per subject.

Instruction Phase: In this phase, the researcher seated the subjects in front of the research laptop, explained the idea of audio time expansion to the subjects, and demonstrated the use of the slider interface. The researcher then explained that the subjects would be asked to listen to several audio clips and immediately answer quiz questions about them, and that during several of these clips they would have the ability to control the rate of audio. In these cases, the subjects were asked to use the slider to best increase their ability to understand the clips and answer the quiz questions. No guidance was given as to what slider setting or speed might achieve this.

Training Phase: After instruction, the subjects were allowed to experiment with the interface by listening to an audio clip of an actor reciting the Gettysburg Address [15]. The slider was active during the training phase, but no quiz questions were asked. Subjects were allowed to repeat the training phase as often as desired, to feel fully familiar with the interface.

Experimental Phase: After training, the subjects engaged in a sequence of six trials. In each trial, the software presented an audio clip, randomly drawn from the nine clips described above. However, in all cases the 1st, 3rd, and 5th trials gave subjects the opportunity to control the audio rate, while the 2nd, 4th, and 6th trials did not. We refer to these as "treatment trials" and "control trials," respectively. During each treatment trial, the initial position of the control slider was randomized, to prevent historicity. Immediately after each audio clip, as part of each trial, the software administered a written multiple-choice quiz to test comprehension of the passage in question, which the subjects answered using a mouse-based interface.

Unlike a true TOEFL test, subjects were not allowed to make written notes during any of the trials.

Survey/Interview Phase: After subjects completed the six trials, the researcher presented a Likert scaled survey, with the following three items, each with five selections (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree):

- "In general, audio expansion for listening skill was useful."
- "Slowing speech for the audio files I listened to was useful."
- "The ability to control audio expansion was easy to use."

Finally, immediately on completion of the survey, the researcher conducted a free form interview with each participant.

3.3. Experimental Records

The test software maintained automatic, anonymized records of user activities and responses through the experimentation and survey phase, in addition to the self-reported TOEFL score described above. This includes the identity and order of the audio clips which were selected for each trial, and the corresponding quiz responses. In addition, for the treatment trials the software recorded the initial position of the slider (i.e., the initial rate of speech) and the time of each audio rate adjustment relative to the start of the experiment. This record is detailed enough to reconstruct each audio clip as each subject heard it, to calculate expansion factor statistics, and to facilitate manual inspection of user behavior.

After all six trials, the software administered the three Likert-scale questions described above, and recorded the answers. Finally, during the interview phase, the researcher took notes and direct quotes from subjects, and transcribed them immediately to electronic storage.

3.4. Analysis Techniques

To analyze subject behavior, the recorded data of each treatment trial was inspected, and the following metrics were extracted:

An average expansion value was calculated as the ratio of modified to unmodified audio playback time. A final

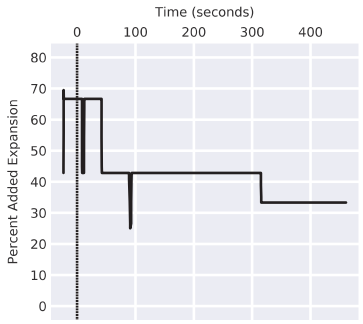

Fig 1a: Time vs Expansion Factor

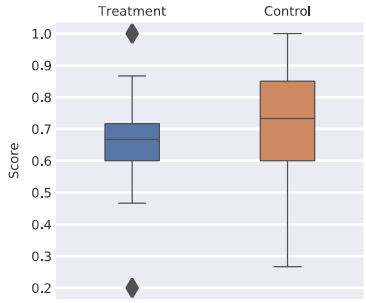

Fig 1b: Treatment vs Control Quiz Scores

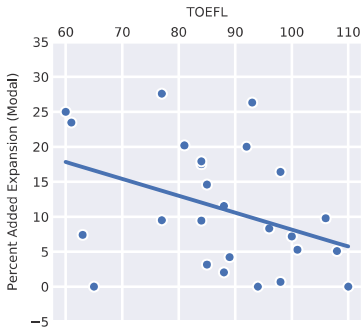

Fig 1c: TOEFL vs Modal Expansion Factor

expansion value was read directly from the experimental records as the expansion value at the end of the audio clip. A modal expansion value was determined from the details of the experimental records, where the modal expansion value is defined as the expansion value for each trial that was experienced for the most amount of time.

Figure 1a depicts one particular trial which began at a randomly selected expansion value of 1.43 (i.e., increasing playback time by 43%). The subject changed the expansion factor to 1.67 [+67%], then began the audio track (denoted by the dashed vertical line), moved back and forth between those two values before remaining at an expansion of 1.43 [+43%] for most of the track, and finally reducing the expansion to 1.33 [+33%] for the remainder of the track. The unmodified length of the track is 324 seconds, which, due to the expansion values applied, played out in 461 seconds. The modal expansion for this trial is 1.43 [+43%], the final expansion is 1.33 [+33%], and the average expansion is 1.42 [+42%].

The results of this paper do not vary with the method of calculation (average, final, or modal), and all discussion following uses modal expansion values unless otherwise noted.

Since each subject experienced three treatment trials, the previously described metrics were averaged into a single score for each subject, i.e., a subject average expansion, a subject final expansion and a subject modal expansion.

4. Results

4.1. Effects on Comprehension

We used a Lilliefors test to examine the control and treatment quiz scores for normality. The control quiz scores ($M = 0.70$, $SD = 0.19$) and the treatment quiz scores ($M = 0.65$, $SD = 0.15$) both rejected the null hypothesis of normal distribution ($p < 0.05$). We therefore used a Kruskal Wallis non-parametric test to analyze the distributions of control and treatment data, and did not find evidence that the distributions of control and treatment quiz scores were statistically significant different ($p > 0.05$). See Figure 1b.

We conclude that in this setting with this audio technique, user control of speech rate neither improved nor degraded subject comprehension.

4.2. Subject Behavior

We examine two aspects of subject behavior during treatment trials. First, we consider whether subjects used audio expansion at all. If an individual treatment trial ended with the audio playing at an unmodified rate (i.e., an expansion value of unity) then that trial was considered a rejection of the expansion technique. Any subject who rejected the expansion technique during only one or two of their treatment trials is considered as partially rejecting audio expansion. Any subject who rejected audio expansion during all three trials is considered as fully rejecting audio expansion.

The 26 subjects experienced a total of 78 treatment trials. As defined above, audio expansion was rejected in 16 of the trials (20.5%). Three of the subjects (11.5%) fully rejected audio expansion, six of the subjects (23%) partially rejected audio expansion, and 17 (65.5%) of the subjects used audio expansion in all three of their treatment trials. Of those subjects who partially rejected the expansion, we found no evidence that the expansion was used in earlier trials and rejected in later trials.

A similar analysis of subject behavior using modal expansion values (i.e., a modal expansion value of unity is considered to be rejecting the technique) gave nearly identical results: rejection in 15 of the trials (19%), five subjects (19%) partially rejecting the technique, and 18 (69%) of the subjects using expansion in all three treatment trials.

Second, we considered if and to what degree the subjects' use of audio expansion correlated with their TOEFL scores. To capture the effect of the audio experiences, we use subject average expansion scores, as described above. We constructed a linear regression model of TOEFL score and subject average expansion, which showed a mild reduction in the use of audio expansion as TOEFL score increased: with an R^2-value of 0.15, an increase of 10 points of self-reported TOEFL led to a 2.4 percent reduction in playback time. See Figure 1c.

4.3. Survey and Interview Data

Finally, we considered the survey and interview data collected from the subjects. The first two Likert survey questions asked for the subjects' opinion of the usefulness of user-controlled audio expansion for listening skill.

The first item, asking about audio expansion and listening skill in general, received 81% positive responses ("Agree" or "Strongly Agree"), 7% neutral responses, and 12% negative responses ("Disagree" or "Strongly Disagree.") The second item, asking about audio expansion in relation to these specific clips received 61% positive, 19% neutral, and 20% negative responses. The third item, asking about the ease of use of the

interface, received 88% positive, 8% negative, and only 4% negative responses.

The spontaneous interview data revealed three recurring concerns among the subjects. First, 30% of the interviewees remarked at the length of the audio passages, and/or the difficulty of remembering information from the beginning of lengthy passages. Second, 40% of the interviewees noted their difficulties with the vocabulary, and that audio expansion does not help with this aspect of a listening skill test. Third, and possibly related, 45% of the interviewees spontaneously remarked that lecture passages were more difficult than conversational passages, often citing vocabulary as a factor.

5. Discussion

Behaviorally, we found strong evidence that subjects will use audio expansion techniques in an attempt to increase their listening comprehension: 88.5% of subjects used audio expansion in at least one of their three treatment trials, and 65.5% used audio expansion in all three of their treatment trials. The average modal expansion across all treatment trials was 1.11 [+11% playback time], and across all trials where audio expansion was not fully rejected was 1.13 [+13%]. These expansion values, while not extreme, are noticeable to an untrained ear.

We also note that our subjects' measured behavior is broadly in line with their survey responses. The second item of our survey ("Slowing speech for the audio files I listened to was useful") asked directly about audio expansion as regards these audio tracks, with 61% agreeing or strongly agreeing, while 65.5% of the subject did use the technique in all three treatment trials.

We also found evidence of personalization, with subject modal expansion factors ranging from 1.0 (i.e., a total rejection of audio expansion across all three trials, no added playback time) to a maximum average expansion of 1.28 [+28%] across three trials for one subject, and a maximum expansion of 1.43 [+43%] for an individual trial. There is also evidence that some of this personalization correlates with TOEFL scores, with greater language proficiency leading to less expansion. However, the effect is modest, with an R^2 value of only 0.15, and a change in expansion factor of only 0.024 per ten points of TOEFL (i.e., a change of 2.4 percent playback time per ten points of TOEFL score.)

However, when measuring objective performance, we find no statistically significant changes to listening comprehension. These results are similar to [16] where native English speakers' preferences for audio expansion were studied under various adverse noise conditions. In that study, when instructed to use (or not use) audio expansion to obtain the best speech intelligibility in background noise, subjects reliably selected increasing amounts of audio expansion with increasing amount of background noise. In that study, subjects also expressed through post-experimental survey a qualified belief that audio expansion was helpful for understanding speech in noise. However, that experiment demonstrated that user-directed audio expansion resulted in statistically significant *degradation* of performance on intelligibility tests. In both studies, the measured behavior and post-experimental survey data is at odds with subject performance.

We speculate several factors may contribute to this effect. First, as in [16], the audio expansion is linear; once a subject selects an expansion value (in the absence of further adjustments) all part of speech are expanded equally. A number of sources [17, 18, 19] observe that Clear Speech is not produced by a process of strictly linear expansion, and that different phoneme classes may experience statistically different values of expansion or even contraction. Other subtle changes are also present, including modifications to vowel pronunciation and vowel space, as well as changing ratios of consonant to vowel energy. This may result in linearly expanded speech sounding somewhat unnatural, as indeed several of our subjects remarked in the interviews. This slightly incorrect cadence may inhibit comprehension.

It may also be the case that audio expansion, by lengthening the playback times of the audio passages, may be making it more difficult to remember information imparted throughout the audio clip. Nearly one third of the subjects expressed a concern about the length of the audio and their ability to recall information. In addition, almost half the subjects noted that while audio expansion may at times make it easier to hear or understand individual words, if the words were unknown to them then no amount of audio expansion would help.

These two factors may combine to create an illusion of improved performance, whereby the expanded audio is "easier" to listen to, which seems will be of some benefit, but these hypothetical benefits fail to materialize due to vocabulary or memory effects. If this is the case, we further speculate that a physiological cognitive load test may show direct evidence of a decreased load with increasing audio expansion.

Finally, we note during our analysis that the TOEFL scores shown in Figure 1c seem clustered into a low-score and a high score group, with low TOEFL scores characterized by high variance of measured expansion values. This may be similar to results found in [16] which suggest that once difficulties (there, noise-induced difficulties; here, skill-based) pass a certain threshold, the variance of subject behavior becomes very large, possibly because one setting is as good (or bad) as any other. However, this still suggests that what is found in lower difficulty (i.e., lower-noise or higher-skill) regions is a stronger illusion that audio expansion is beneficial.

6. Conclusions

This study gives ESL listeners fine, real-time control of the pace of lengthy audio passages. We used this tool to examine ESL subjects' preferences, performance, and perceived utility of the tool in conjunction with listening comprehension tests. We have shown strong evidence that ESL subjects will use audio expansion for the purpose of increasing their comprehension, and evidence of a small tendency to add more expansion as self-reported TOEFL skills decrease. However, there is no evidence that these subject-directed audio expansions improve listening comprehension.

We believe these results highlight a serious difficulty associated with subject-directed auditory interventions specifically, and subject-directed sensory modification in general. Namely, that users may have preferences and beliefs about the utility of sensory modifications, but that these beliefs may diverge from objective measures of performance.

7. References

[1] S. B. Davis and P. Mermelstein, "Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[2] Meskill, C., 1996. Listening skills development through multimedia. Journal of Educational Multimedia and Hypermedia, 5(2), pp.179-201.

[3] Feyten, C.M., 1991. The power of listening ability: An overlooked dimension in language acquisition. The modern language journal, 75(2), pp.173-180.

[4] Vandergrift, L., 2004. 1. Listening to Learn or Learning to Listen? Annual review of applied linguistics, 24, pp.3-25.

[5] Bradlow, A.R. and Bent, T., 2002. The clear speech effect for non-native listeners. The Journal of the Acoustical Society of America, 112(1), pp.272-284.

[6] Bradlow, A.R. and Alexander, J.A., 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. The Journal of the Acoustical Society of America, 121(4), pp.2339-2349.

[7] Hargus Ferguson, S., 2004. Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. The Journal of the Acoustical Society of America, 116(4), pp.2365-2373.

[8] Foulke, E., 1968. Listening comprehension as a function of word rate. Journal of Communication, 18(3), pp.198-206.

[9] Foulke, E. and Sticht, T.G., 1969. Review of research on the intelligibility and comprehension of accelerated speech. Psychological bulletin, 72(1), p.50.

[10] Carver, R.P., 1973. Effects of increasing the rate of speech presentation upon comprehension. Journal of Educational Psychology, 65(1), p.118.

[11] Zhao, Y., 1997. The effects of listeners' control of speech rate on second language comprehension. Applied linguistics, 18(1), pp.49-68.

[12] Blau, E.K., 1990. The effect of syntax, speed, and pauses on listening comprehension. TESOL quarterly, 24(4), pp.746-753.

[13] Khatib, M., 2010. The Effect of Modified Speech on Listening to Authentic Speech. Journal of Language Teaching & Research, 1(5).

[14] Novak III, J.S., Archer, J., Shafiro, V., Kenyon, R.V. and Leigh, J., 2013. On-line audio dilation for human interaction. In INTERSPEECH (pp. 1869-1871).

[15] Novak III, J.S., Tandon, A., Leigh, J. and Kenyon, R.V., 2014, September. Networked on-line audio dilation. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (pp. 255-258). ACM.

[16] A Reading of the Gettysburg Address [Internet]. NPR.org. 2016 [cited 14 December 2016]. Available from: http://www.npr.org/templates/story/story.php?storyId=1512410

[17] Novak III, J.S. and Kenyon, R.V., Effects of User Controlled Speech Rate on Intelligibility in Noisy Environments.

[18] Picheny, M.A., Durlach, N.I. and Braida, L.D., 1986. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. Journal of Speech, Language, and Hearing Research, 29(4), pp.434-446.

[19] Garnier, M. and Henrich, N., 2014. Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?. Computer Speech & Language, 28(2), pp.580-597.

[20] Gygi, B. and Shafiro, V., 2014. Spatial and temporal modifications of multitalker speech can improve speech perception in older adults. Hearing research, 310, pp.76-86.