



Intelligibility and acceptability of time-compressed utterances: an experimental study with blind and sighted listeners

Miguel Oliveira Jr.¹, Ayane Nazarela S. de Almeida^{1,2}, René Alain S. de Almeida^{1,2,3}, Oyedeji Musiliyu¹

¹ Faculdade de Letras, Universidade Federal de Alagoas, Brazil

² CAPES Foundation, Ministry of Education of Brazil, Brazil

³ Universidade Federal de Sergipe, Brazil

miguel@fale.ufal.br, ayanesantos@hotmail.com, renealain@hotmail.com, bodeses@yahoo.fr

Abstract

This paper reports an experimental study on the impact of time-compressed speech on acceptability and intelligibility of utterances in Brazilian Portuguese. For the experiments, short audio sentences containing warning messages were used as stimuli. These sentences were recorded in a natural speech rate and then digitally manipulated to faster rates in a scalar fashion (from 9 to 19 syllables per second). Intelligibility and acceptability tests were then conducted with blind and sighted subjects. The results indicate that time-compressed speech has a significant impact on both acceptability and intelligibility of utterances for both groups of participants and that while blind subjects tended to give slightly higher acceptability rates across all speech rate conditions, sighted subjects performed better in the intelligibility experiment, what contradicts a trend that is often reported in the literature.

Index Terms: speech rate, acceptability, intelligibility

1. Introduction

In regulating pharmaceutical drug advertisements on television, the Brazilian Health Surveillance Agency (ANVISA) warns that, depending on the chemical composition of a drug, the commercial should mandatorily be accompanied by warnings about risks related to its administration. Furthermore, it determines that when written warnings are not displayed for a period of time that is adequate for reading, they should come also in spoken form [1]. One obvious importance for such regulation is that, by following it, vital messages reach an audience that cannot read, such as illiterates, visually impaired and blind people.

The regulation, however, sets no stipulation of speech rate for such warnings. A preliminary analysis of some recent common pharmaceutical drug advertisements in mass media tv broadcasts indicates that speech rate tends to be 100% faster in warnings, if compared to that of the entire advertisement.

The average speech rate of fluent native speaker of Brazilian Portuguese (BP), according to the literature [2], [3] varies from 3.2 to 5.5 syllables per second (syll/s), depending on several factors, such as age, context and content of the utterance. There is, however, to our knowledge, no cut-off point in terms of speech rate that is associated to intelligibility/acceptability in BP.

Taking into account all the above mentioned facts, as well as the hypothesis that speech rate plays a crucial role in acceptability and intelligibility of utterances, this study aims to explore, by means of a perceptual experiment, the impact of manipulated speech rate on the acceptability and intelligibility of utterances in BP with blind and sighted subjects, in order to

identify speech rate cut-off points that are associated to acceptability and intelligibility in BP.

There are already some studies on the impact of speech rate manipulation in both intelligibility and acceptability for languages such as German [4], [5], [6], [7], Dutch [8] and Japanese [9]. Most of these studies consider visually impaired individuals as subjects because of the fact that many of these people make use of time compressed speech in a number of applications that are specifically designed for them. Until the elaboration of the present study, no known research has been done on the impact of speech rate manipulation in both intelligibility and acceptability for Brazilian Portuguese that considers blind persons as subjects. In order to guarantee that regulations imposed on language usage that are essential to the safety of people with special needs, such as the one we mentioned above, are indeed useful, surveillances agencies must take into consideration the particularities of the language itself. In that sense, the present study also stands as a contribution for the elaboration of more precise regulations that deal with the compression of speech in warnings that are obligatorily conveyed in commercials.

2. Background

The establishment of methods for artificial compression of speech took place as early as 1953, first by mechanical [10] and later by electronic [11] means. Since then, several applications for time-compressed speech were developed and are widely employed in different areas. In education, for example, it is used as a study aid; in speech technology, as a means of reading the content of a computer screen; in telephony, to enable the rapid review of messages in high-traffic systems; in advertisement, as a means of allowing more information to be included in an ad and reducing its cost.

As it happens to any speech communication, processing, or synthesis systems, applications that use time-compressed speech need to be adequately evaluated. This is often done by means of intelligibility or acceptability testing involving human listeners.

The general concept of intelligibility refers to the extent to which speech can be properly understood by listeners. According to [12] and [13], the ability to repeat a word, phrase, or short sentence accurately, is often taken as an index of intelligibility. Intelligibility tests are a very important evaluative procedure for speech technology development and have been in use since as early as 1910 [12].

A procedure typical for testing intelligibility of time-compressed speech is one in which words are presented, one at a time, to a listener. The listener's task is to reproduce them orally, or in writing, and his score is the correctly identified

fraction of those words. This procedure is sometimes referred to as an articulation test [12]. A similar procedure was adopted by [14] while observing the intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Intelligibility of time-compressed natural-sounding sentences was measured using a paradigm that alters speech-energy rhythm over a range of frequencies. It is important to note, though, that the intelligibility of time-compressed speech can be affected by listener-related factors, including age, auditory ability, language fluency, and familiarity with synthesized speech [15], [16].

Acceptability, on the other hand, deals with the subjective opinions of how a speech sounds. In contrast to measurement methods of speech intelligibility, which can be scored objectively, speech acceptability assessments are based on listener's subjective judgement. Evaluators listen to speech samples and assess speech quality using a numerical scale or categorical labels (a Likert-type scale), which could be later converted into numbers [13]. A very common procedure used to evaluate speech acceptability is the Mean Opinion Score (MOS), which is expressed on a five-point scale defined by category labels, 1 being the worst and 5 the best.

Because of the widespread use of time-compressed speech in computer applications (such as screen readers), that are most of the time designed for visually impaired persons, a number of studies have been conducted on acceptability and intelligibility of time-compressed speech with blind subjects. What most of these studies report is that, in general, visually impaired persons perform better on comprehension of time-compressed speech than control subjects.

In a study with German native speakers, for example, [4] demonstrated that blinds can understand synthetic speech at a rate of up to 17 syll/s. Sighted individuals, on the other hand, only understand speech at a maximum rate of 9 syll/s. In another study, also on German, [5] investigated the extent to which speech could be time-compressed and still understood by blind people who have daily practice with speech synthesis, compared to sighted persons without such training. The results showed that blind listeners could understand synthetic speech far better than sighted listeners. According to [5], the fact that blind people are able to understand speech at 22 syll/s may be related to a particular ability of these individuals to use parts of their visual system to enhance speech processing.

To further elucidate the neural mechanism underlying this particular skill, [6] and [7] conducted an experiment with blind subjects using functional magnetic resonance imaging (fMRI) while listening to ultra-fast speech. The results showed a significant hemodynamic activation of right-hemispheric primary visual cortex, contralateral fusiform gyrus and bilateral pulvinar, what indicates that central-visual structures are recruited during auditory comprehension of ultra-fast speech.

Speech perception has often been considered as audiovisual in nature, and the visual system seems to be connected to the auditory system in various ways. Thus, it seems plausible that blind people are able to use parts of their visual system to enhance speech processing [17].

3. Methods

A total of 50 warning texts that came from various sources was preselected for this study. The criteria used for the selection of this preliminary corpus were (i) ecological

validity: the texts had to be real warning messages that are used for different purposes and (ii) size: the texts had to be on average 12 words long, with a range between 11 and 13 words. In order to avoid an effect of familiarity on the intelligibility experiment, we ran a familiarity test conducted with twelve students/specialists in linguistics. They were instructed to assign to each of the warnings a grade of 1 (less familiar) to 5 (more familiar). From this familiarity test, a total of thirty-six sentences that were considered less familiar was selected for the acceptability and intelligibility tests. The remaining sentences were used in a training session that was done before the actual experiment.

All the stimuli used in this study were produced by an adult male native speaker of Brazilian Portuguese with a normal speech rate (around 5.6 syll/s). It must be observed here that the utterances used in this study did not present any instance of pauses - defined here as a "period of silence" greater than 150ms [18]. The recordings were then temporally manipulated in Praat [19] with the Vocal Toolkit script [20]. The script uses the PSOLA synthesis method, that preserves most of the physical characteristics of speech signal, even in high temporal compression ratios.

Six different speech rates were used for the perceptual experiments: 9, 11, 13, 15, 17, 19 syllables per second. One of these rates (13 syll/s) is commonly used in the advertisements we have analyzed.

It must be noted here that although a variety of units of measurement is used in research on speech rate, the most appropriate unit of measurement, according to many authors, is that of the syllables per unit of time [21], [22], [23]. It is important to note, however, that this unit of measurement has some flaws, the main one, perhaps, is the fact of not taking into consideration the processes which could result in syllable omission, a phenomena often found in rapid speech, such as assimilation and syllabic deletion [24].

The present study opted for the widely used unit in time-related research on speech for comparison purposes. The authors acknowledge the problems related to that option, but it is assumed that they are not significant enough to invalidate the interpretation of data. Speech rate will be analyzed in this study, thus, using the unit of syllables per second (syll/s).

The stimuli were randomly presented to the participants of the experiments through a PC computer. The participants used noise-canceling headphones for the entire procedure and had only access to the audio version of the stimuli. The experiments were carried out in a quiet area. Written instructions were given prior to starting the task and a training protocol was also completed.

For the acceptability experiment, a set of eighteen warning messages was presented to the participants, one at a time. For each stimulus, the participants had to speak out their responses using a numerical scale of 1 to 5, that represents the following concepts: 1 "terrible", 2 "poor", 3 "fair", 4 "good" and 5 "excellent". This is the same procedure adopted by [25], and is based on the MOS protocol. Experiments that employ the MOS measure often consider values above 4.0 as an indication of good quality, while values dropping below 3.5 are termed *unacceptable* [26].

For the intelligibility experiment, another set of eighteen warning messages, different from the previous one, was presented, one at a time, to the participants, who were instructed to repeat what they heard while their verbal

response was recorded. After the experiment was over, transcriptions of the recorded voices were made and the recall rate was defined as the number of correct words compared with the total number of words in the sentence. For this study, a sentence was considered intelligible when the recall rate was higher than 90%, following the criteria adopted in [9].

It is important to note that, according to studies that considered the amount of items that can be stored in working memory, an average individual can recall up to 9 isolated words on a list. However, when words are part of a sentence, working memory capacity increases up to 16 words [27]. As it was mentioned above, the stimuli used in the present study are composed of sentences that contain up to thirteen words. This was done in order to prevent any interference of working memory capacity on the intelligibility experiment.

Thirty blind subjects (20 men, age's mean=39.36, age's standard deviation=13.64) and thirty sighted subjects (13 men, age's mean=25.66, age's standard deviation=6) participated in this study. All subjects are native speakers of Brazilian Portuguese, literate and of age. Individuals with auditory, motor and/or cognitive impairments, as self-declared, were not included in the study. All participants read and signed a Free and Clarified Consent Term. Blind subjects had access to a version of the term in Braille.

4. Results

4.1. Acceptability

Means (and standard deviations) of acceptability scores (1–5) for the two groups of participants across the six speech rates are shown in Table 1.

Table 1. Means (and sd) of acceptability scores.

Rates	Subjects	Blind	Sighted
9 syll/s		4.6 (0.54)	4.4 (0.68)
11 syll/s		3.9 (1.03)	3.6 (1.04)
13 syll/s		3.2 (1.04)	2.6 (0.87)
15 syll/s		2.3 (1.01)	1.9 (0.86)
17 syll/s		1.8 (0.87)	1.5 (0.55)
19 syll/s		1.4 (0.45)	1.2 (0.50)

As it can be seen in Table 1, the higher the speech rate, the less acceptable the utterance is for both groups of participants. In general, the utterances with a speech rate of 9 syll/s were well-evaluated, while those with a speech rate of 11 syll/s were regarded, on average, as either “fair” or “good”, by both blind and sighted subjects. Utterances with a speech rate greater than or equals to 13 syll/s were considered, on average, either “fair” or “bad” for the blind subjects, and “bad” or “terrible” for the sighted subjects. It should be pointed out here again that this is precisely the rate that is commonly employed for the warning message in most TV and radio ads that we have analyzed.

The values in Table 1 also indicate that blind subjects tend to give, on average, higher acceptability rates in all cases of time-compressed utterances. However, that difference was not statistically significant across all speech rates, as observed through a paired t-test done in order to compare means between groups of participants in each speech rate condition.

The results of paired t-tests show a significant difference between the acceptability scores presented by blind and sighted subjects at speech rates 13 syll/s, $t(89)=3.15$, $p<0.05$, 15 syll/s, $t(89)=2.81$, $p<0.05$ and 17 syll/s, $t(89)=2.78$, $p<0.05$. At other speech rates, there are some differences between the two groups of participants, but they are not significant.

One-way ANOVA procedures for repeated measures indicated that the acceptability ratings were significantly different across conditions (rates) for both the blind subjects, $F(5,534)=102.6$, $p<0.05$, and the sighted subjects, $F(5,534)=143.7$, $p<0.05$.

The results of acceptability test were examined using a two-way (Participant x Rate) Analysis of Variance (ANOVA) with repeated measures for one factor (Rate). There was a significant difference between groups of participants, $F(1,58)=3.9$, $p=0.05$, and rates, $F(5,290)=271.7$, $p<0.05$. However, the interaction observed between rate and group of participant was not significant, $F(5,290)=1.2$, $p>0.05$.

4.2. Intelligibility

Means of intelligibility scores (percent correct) for the two groups of participants across the six speech rates are displayed in Figure 1. As it can be seen, the tendency observed in the acceptability test is also present here: the higher the speech rate the less intelligible the utterance is considered, for both groups of participants.

Figure 1 also highlights that no group of participants achieved 90% of intelligibility score, not even at speech rate of 9 syll/s. If 90% is to be taken as a minimum score in a perceptual test in order for an utterance to be considered intelligible, as the literature proposes [9], the results of the present study suggest that from 9 syll/s up, any time-compressed utterance will be considered unintelligible in Brazilian Portuguese, regardless the condition of the hearer: being s/he either blind or sighted.

Besides, the results of this perceptual experiment point to a different direction from the one commonly reported in previous studies. As mentioned above, visually impaired people systematically perform better on comprehension of time-compressed speech than sighted subjects in perceptual experiments. Figure 1 clearly indicates that this is not the case in the present study: sighted subjects performed better in the intelligibility experiment than blind subjects, across all speech rate conditions. The results of paired t-tests show a significant difference between the intelligibility scores presented by blind and sighted subjects at speech rates 13 syll/s, $t(89)=5.15$, $p<0.05$; 15 syll/s, $t(89)=4.81$, $p<0.05$; 17 syll/s, $t(89)=5.45$, $p<0.05$ and 19 syll/s, $t(89)=6.0$, $p<0.05$.

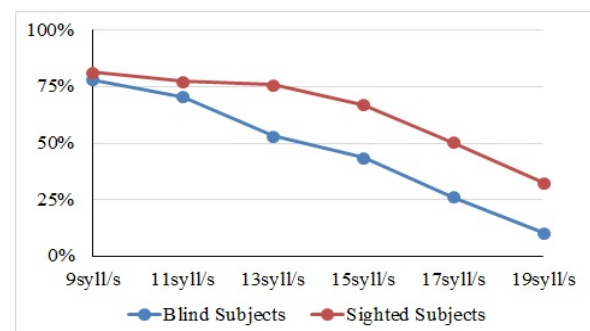


Figure 1: Intelligibility experiment results.

A two-way ANOVA revealed a significant difference between subjects, $F(1,58)=20.7$, $p<0.05$, and between rates, $F(5,290)=271.7$, $p<0.05$. A significant interaction was observed between rate and subject, $F(5,290)=3.6$, $p<0.05$. Post-hoc tests revealed that intelligibility scores differ significantly from 15 syll/s on for the sighted subjects, and from 13 syll/s on for the blind subjects ($p < 0.05$).

As a result of the interaction effect, all participants from each group that were able to repeat, on average, more than 90% of the words in the utterances were considered separately for each speech rate condition. Table 2 below shows the number of individuals in each group who performed best, on average, among all subjects on the intelligibility experiment.

Table 2. Participants who repeated more than 90% of the words on the intelligibility experiment.

Rates	Blind		Sighted	
	No.	%	No.	%
9 syll/s	6	20	10	33.3
11 syll/s	5	16.7	9	30
13 syll/s	3	10	8	26.7
15 syll/s	3	10	4	13.3
17 syll/s	-	-	-	-
19 syll/s	-	-	-	-

The results in Table 2 indicate: i) a reduction in intelligibility as speech rate increases, ii) a greater number of sighted subjects, in comparison to blind subjects, who performed better on the intelligibility experiment, across all speech rate conditions, iii) a very low representativity, in terms of percentage, of both blind and sighted subjects who could actually repeat more than 90% of the words in time-compressed speech, and iv) an absence of subjects who could repeat more than 90% of the words in the utterances associated with higher speech rates (17 and 19syll/s).

5. Discussion and Conclusions

The results reported here clearly demonstrate that time-compressed speech has a significant impact on both acceptability and intelligibility of utterances in Brazilian Portuguese. On the one hand, it was found that intelligibility decreases significantly under increased time compression. These findings are in line with those of [28] and [29]. On average, no group of participants achieved more than 82% of intelligibility score in any of the manipulated speech rates that were used in the experiment, what is considered a poor score by intelligibility tests standards.

On the other hand, the results indicate that the more compressed the speech, the less likely it will be for listeners to consider it acceptable. Specifically, utterances with a speech rate of 11 syll/s were rated, on average, 3.9 for a group of subjects and 3.6 for the other group. Utterances with a speech rate of 13 syll/s, that are often associated to warning messages in Brazilian TV and radio ads, were rated, on average, 3.2 for a group of subjects and 2.6 for the other group. As pointed out before, evaluation protocols that use the MOS measure often consider values above 4.0 as adequate and below 3.5 unacceptable. We can thus conclude that neither 11 syll/s or 13 syll/s can be considered adequate speech rates for time-compressed speech material, being the latter rate also unacceptable.

The findings also suggest that intelligibility and acceptability evaluation are very similar for both sighted and blind subjects, despite the fact that blind people are reportedly able to use parts of their visual system to enhance speech processing [17]. While it is true that blind subjects tend to give, on average, slightly higher acceptability rates for time-compressed utterances, as the results of this study indicated and showed to be statistically significant in three speech rates (13 syll/s, 15 syll/s and 17 syll/s), sighted subjects performed better in the intelligibility experiment, across all speech rate conditions, with statistically significant differences in four speech rates (13 syll/s, 15 syll/s, 17 syll/s and 19 syll/s). This latter finding contradicts what is generally observed in the literature for other languages [4], [5], [6], [7]. It is important to point out here that recall rate started to decrease significantly at 13 syll/s for the group of blind subjects, with approximately 20% of observed variance in intelligibility score attributable to compression level.

Future research should examine the impact of time compressed speech in other groups of potential target of applications that use it, such as illiterates. In any case, the results of the present study clearly suggest that it is essential that time-compressed speech is used with caution in applications that demand full comprehension of the message that it conveys. In particular, the results presented here justify the need to make more precise laws as regards the elocution rate of warnings in drug commercials, since these warnings are of vital importance and are meant for individuals that cannot read written warnings.

6. References

- [1] ANVISA. Regulation about pharmaceutical drug advertising. Online: http://portal.anvisa.gov.br/wps/wcm/connect/b12a03004745973d9f9adf3fbc4c6735/rdc_9608_comentada.pdf?MOD=AJPERES, accessed on 04 Feb 2015.
- [2] V. O. Martins and C. F. Andrade, "Perfil evolutivo da fluência da fala de falantes do Português brasileiro," *Pró-Fono Revista de Atualização Científica*, vol. 20, no. 1, pp. 7–12, 2008.
- [3] M. Oliveira Jr., *Prosodic Features in Spontaneous Narratives*. Thesis (PhD). Simon Fraser University, Vancouver, Canada, 2000.
- [4] J. Trouvain, "On the comprehension of extremely fast synthetic speech," *Saarland Working Papers in Linguistics (SWPL)*, vol. 1, pp. 5–13, 2007.
- [5] A. Moos and J. Trouvain, "Comprehension of Ultra-Fast Speech - Blind vs. 'Normally Hearing' Persons," *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken*, 2007, pp. 677–680.
- [6] S. Dietrich, I. Hertrich, and H. Ackermann, "Training of ultra-fast speech comprehension induces functional reorganization of the central-visual system in late-blind humans," *Frontiers in Human Neuroscience*, vol. 7, article 701, 2013.
- [7] S. Dietrich, I. Hertrich, and H. Ackermann, "Ultra-fast speech comprehension in blind subjects engages primary visual cortex, fusiform gyrus, and pulvinar - a functional magnetic resonance imaging (fMRI) study," *BMC Neuroscience*, vol. 14, article 74, 2013.
- [8] E. Janse, *Production and perception of fast speech*. Utrecht: LOT, 2003.
- [9] C. Asakawa, H. Takagi, S. Ino, and T. Ifukube, "Maximum listening speeds for the blind," *Proceedings of the 2003 International Conference on Auditory Display, (ICAD03)*, Boston, 2003, pp. 276–279.
- [10] W. D. Garvey, "The intelligibility of speeded speech," *Journal of Experimental Psychology*, vol. 45, no. 2, pp. 102–108, 1953.
- [11] G. Fairbanks, W. L. Everitt, and R. P. Jaeger, "Method for time or frequency compression-expansion of speech," in S. Duker

- (Ed.), *Time-compressed speech*, vol. 1, pp. 172–180. Metuchen, N.J.: Scarecrow, 1974.
- [12] E. Foulke, *The comprehension of rapid speech by the blind, part III*, 1969. Online: <http://files.eric.ed.gov/fulltext/ED034346.pdf>, accessed on 25 Feb 2015.
- [13] A. Schmidt-Nielsen, “Intelligibility and Acceptability testing for Speech Technology,” *Naval Research Laboratory*, 1992. Online: <http://www.dtic.mil/dtic/tr/fulltext/u2/a252015.pdf>, accessed on 24 Feb 2015.
- [14] O. Ghitza and S. Greenberg, “On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence,” *Phonetica*, vol. 66, no. 1–2, pp. 113–126, 2009.
- [15] B. Sutton, J. King, K. Hux, and D. Beukelman, “Younger and older adults rate performance when listening to synthetic speech,” *Augmentative and Alternative Communication*, vol. 11, no. 3, pp. 147–153, 1995.
- [16] E. Janse, M. Van der Werff, and H. Quené, “Listening to fast speech: aging and sentence context,” in J. Trouvain and W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 2007, pp. 681–684.
- [17] A. Moos, I. Hertrich, S. Dietrich, J. Trouvain, and H. Ackermann, “Perception of Ultra-Fast Speech by a Blind Listener – Does He Use His Visual System?,” *Proceedings of the 8th International Seminar on Speech Production, ISSP*, 2008, pp. 297–300.
- [18] S. Kowal, S. Wiese, and D. O’Connell, “The use of time in storytelling,” *Language and Speech*, vol. 26, no. 4, pp. 377–392, 1983.
- [19] P. Boersma, “Praat, a system for doing phonetics by computer,” *Glott International*, vol. 5, no. 9–10, pp. 341–345, 2001.
- [20] R. Corrette, Praat vocal toolkit [Computer Software]. Barcelona. praatvocaltoolkit.com, 2012.
- [21] E. Blaauw, *On the perceptual classification of spontaneous and read speech*. Research Institute for Language and Speech, Utrecht University, 1995.
- [22] J. Fon, “Speech rate as a reflection of variance and invariance in conceptual planning in storytelling,” *Proceeding of the ICPHS*, San Francisco, vol. 14, no. 1, pp. 663–666, 1999.
- [23] M. Van Donzel, *Prosodic Aspects of Information Structure in Discourse*. Thesis (PhD). Faculteit der Geesteswetenschappen, University of Amsterdam, Amsterdam, 1999.
- [24] D. C. O’Connell and T. D. Kowal, “Cross-linguistic pause and rate phenomena in adults and adolescents,” *Journal of Psycholinguistic Research*, vol. 1, no. 2, pp. 155–164, 1972.
- [25] P. A. Dagenais, G. R. Brown, and R. E. Moore, “Speech rate effects upon intelligibility and acceptability of dysarthric speech,” *Clinical Linguistics & Phonetics*, vol. 20, no. 2–3, pp. 141–148, 2006.
- [26] E. B. Carne, *A professional’s guide to data communication in a TCP/IP world*. Boston: Artech House, 2004.
- [27] A. D. Baddeley and G. J. Hitch, “Development of working memory: should the Pascual Leone and the Baddeley and Hitch models be merged?,” *Journal of Experimental Child Psychology*, vol. 77, no. 2, pp. 128–137, 2000.
- [28] P. E. King and R. R. Behnke, “The Effect of Time-Compressed Speech on Comprehensive, Interpretive, and Short-Term Listening,” *Human Communication Research*, vol. 15, no. 3, pp. 428–443, 1989.
- [29] M. J. Beatty, R. R. Behnke, and D. L. Froelich, “Effects of achievement incentive and presentation rate on listening comprehension,” *Quarterly Journal of Speech*, vol. 66, no. 2, pp. 193–200, 1980.