Universidade do Minho
Escola de Psicologia

Marta Maria Saraiva Bonjardim

The Need for Speed: Investigating the Influence of Audio Speed on Listening Comprehension

junho de 2024

The Need for Speed: Investigating the Influence of Audio Speed on Listening Comprehension
Marta Maria Saraiva Bonjardim
UMinho | 2024

Universidade do Minho
Escola de Psicologia

Marta Maria Saraiva Bonjardim

The Need for Speed: Investigating the Influence of Audio Speed on Listening Comprehension

Dissertação de Mestrado
Mestrado em Cognição Humana

Trabalho efetuado sob a orientação da
Professora Doutora Ana Paula de Carvalho Soares
e da
Doutora Helena Mendes Oliveira Baldassarre

junho de 2024

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição
CC BY

https://creativecommons.org/licenses/by/4.0/

Agradecimentos

"O Principezinho", de Antoine de Saint-Exupéry, foi uma de múltiplas obras que marcaram a minha infância. Em particular, Saint-Exupéry escreveu a célere frase que me acompanha até aos dias de hoje: "Aqueles que passam por nós, não vão sós, não nos deixam sós. Deixam um pouco de si, levam um pouco de nós". É esta a frase que me remete para a gratidão que sinto por todos os que passaram pela minha vida, em especial pelos que acompanharam a minha jornada académica e contribuíram para o sucesso da mesma.

Obrigada, Professora Ana Paula e Professora Helena, por tudo o que me ensinaram ao longo destes últimos cinco anos, e em especial neste último ano. Obrigada pelo apoio incansável, pela compreensão, pelos conselhos, e por confiarem sempre em mim. Não acabaria esta etapa com o mesmo sentimento de concretização se não fosse pelas Professoras.

Certamente também não teria sido um percurso tão enriquecedor sem a minha turma do mestrado em Cognição Humana, pelo que agradeço a cada um deles pela camaradagem e união que foi essencial para enfrentar todos os desafios que foram surgindo. Em especial, agradeço à Bela, por ser um pilar na minha vida académica e pessoal. Obrigada, Bela, pela ajuda fundamental à escrita desta dissertação, pelo apoio diário, pelo convívio, e pela amizade.

Agradeço também aos meus amigos que estiveram sempre presentes ao longo dos últimos anos. Obrigada, Luana, pela amizade que criamos desde o primeiro dia de universidade, obrigada por me veres e fazeres crescer. Obrigada, Nuno, Rafa e Foca, por todos os momentos de galhofa. Obrigada, Sara, pelo companheirismo e por todas as trocas de livros – hei de estar contigo, onde quer que vás. Obrigada, Mariana, por todas as "chamadas de estudo" e pela cumplicidade. Obrigada, Benny e Paulo, por serem como irmãos, sei que posso contar com vocês para tudo.

Um agradecimento especial ao Bruno, por ser o meu braço direito desde o início. Obrigada, Bruno, por seres a constante dos meus dias, por me acompanhares e cuidares de mim em todos os momentos e em todas as viagens que fazemos.

Por último, e mais importante, escrevo um agradecimento profundamente sentido à minha família. Obrigada, mãe e pai, pelos sacrifícios e pelo amor incondicional, devo-vos tudo. Obrigada Nuno, por seres um irmão preocupado e presente. Obrigada, padrinho e madrinha por acreditarem no meu percurso e estarem sempre do meu lado. Obrigada, avó Lola, pelos ensinamentos. Obrigada, avó São, por ser a minha segunda mãe. E obrigada, avô Custódio, por ser o meu anjo da guarda.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

MARTA MARIA SARAIVA BONJARDIM

Assinado de forma digital por MARTA MARIA SARAIVA BONJARDIM
Dados: 2024.06.05 21:04:02 +01'00'

The Need for Speed: Investigating the Influence of Audio Speed on Listening Comprehension

Abstract

With the emergence of Massive Open Online Courses (MOOCs) and the rise of online classes, more students accelerate video and audio. While it allows faster content consumption, its impact on listening comprehension remains unclear. In this study, we aim to understand how listening comprehension is affected by listening to audio at fast (1.5x) and very fast (2x) speeds, compared to its baseline speed (1x), while controlling for individual learning factors and participants' perception of the stimuli. In each of the three sessions, 48 students listened to one audio sample, selected from a previous pilot study, at one of the three speeds. Afterwards, participants filled out an evaluation test and a perception questionnaire. In the last session, we administered a sociodemographic questionnaire and the Reading Span Test (Gaspar, 2001). Logistic mixed-effects models showed that when audio speed increases, there is a gradual decrease in performance on the evaluation tests, although it is only significant between 1x and 2x speeds. Importantly, this effect of speed persists even when accounting for individual learning factors and content perception factors. These findings bear implications for individuals who accelerate audio content, shedding light on the search for an optimal playback speed.

Keywords: Accelerated speech, human-machine interaction (HMI), listening comprehension, playback speed.

“The Need for Speed”: Investigar a Influência da Velocidade de Áudios na Compreensão Auditiva

Resumo

Com o surgimento dos Cursos *Online* Abertos e Massivos (MOOCs) e o aumento das aulas *online*, mais alunos aceleram a velocidade de visualização de vídeo e áudio. Embora este comportamento permita o consumo mais rápido de conteúdos, o seu impacto na compreensão auditiva não é claro. Este estudo visa compreender como a compreensão auditiva é afetada por áudios a velocidade rápida (1.5x) e muito rápida (2x), comparando com a velocidade original (1x), enquanto controlamos fatores individuais de aprendizagem e de perceção de conteúdo. Em três sessões distintas, 48 alunos ouviram um áudio a uma das três velocidades em estudo. Posteriormente, administrámos um teste de avaliação e um questionário de perceção de conteúdo. Na última sessão, acrescentámos um questionário sociodemográfico e o *Reading Span Test* (Gaspar, 2001). Modelos logísticos de efeitos mistos revelaram que, aumentando de velocidade, há um decréscimo gradual de desempenho nos testes de avaliação, embora seja significativo apenas entre 1x e 2x. Este resultado persiste quando considerados fatores individuais de aprendizagem e de perceção de conteúdo. Estes resultados têm implicações para perceber como a aceleração de conteúdo de áudio pode afetar a compreensão do mesmo, fornecendo pistas sobre a velocidade ideal na reprodução de discurso acelerado.

Palavras-chave: Compreensão auditiva, discurso acelerado, interação homem-máquina (IHM), velocidade de reprodução.

Index

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS ii
Agradecimentos iii
STATEMENT OF INTEGRITY iv
Abstract v
Resumo vi
Figures Index viii
Tables Index viii
Introduction 9
Method 20
Participants 20
Materials 20
Procedure 23
Results 25
Discussion 30
References 34

Figures Index

Figure 1 *Percentage of Correct Responses (Accuracy) in the Listening Comprehension Test per Speed Condition*..... 26

Figure 2 *Distribution of the Participants According to Previous Experience Manipulating Speed*..... 26

Figure 3 *Distribution of the Participants According to their Average Working Memory Span* 27

Tables Index

Table 1 *Means, Standard Deviations (in brackets), and Friedman's ANOVA for the Audio Characteristics of the Pilot Study (N = 9)* 22

Table 2 *Mean and standard deviation (in brackets) for Familiarity, Interest, Arousal, and Valence Ratings for Each Audio per Speed* 28

Introduction

The advancements in technology over the past century have enabled the manipulation of audio speed, which has led researchers to be curious about the ability that humans possess to understand time-compressed speech (e.g., Orr et al., 1965; Orr & Friedman, 1968). Time-compressed speech alludes to audio that has been increased in speed while retaining its audio output and sound quality (Arons, 1992; Friedman et al., 1967). This capability opens up the potential for delivering speech at different rates, both slower and faster, from the original pace, all while preserving vocal pitch integrity (Foulke, 1968). Time-compressed speech finds applications in diverse fields, such as education, human-computer interfaces and even assisting individuals with disabilities (Arons, 1992). For instance, time-compressed speech has been employed to watch school lessons at faster speeds (Jacobson et al., 2018), advertisements and audio commercials (Rodero, 2019), and assistive technologies for the visually impaired (Foulke, 1967). Additionally, social media platforms such as YouTube and WhatsApp have been allowing users to manipulate video and audio speed (Chen et al., 2017; Costa et al., 2022). The rise of online classes following the COVID-19 pandemic (Mo et al., 2022) has further enticed growing interest in studying time-compression techniques in Massive Online Open Courses (MOOCs) and lecture videos. For instance, Murphy and colleagues (2022) surveyed 123 university students, of whom 85% reported watching lecture videos at a speed higher than the original. Students accelerate video and audio playback speed primarily for the convenience of saving time (Kıyak et al., 2023), the perception of enhanced focus, and the belief that it improves their ability to learn more (Cardall & Ulrich., 2008). However, in reality, its impact on learning remains unclear, and individuals who use these tools are often unaware of what might best suit their needs. Therefore, the calibration of audio and video playback speed is among the potential parameters that sound and multimedia platform designers should consider.

The accelerated playback hypothesis, initially proposed by Foulke (1968) and later formalized by Cheng et al. (2022), posits that learning outcomes do not decrease when audio narration speed is accelerated until a cognitive capacity threshold is reached. However, the specific word rate threshold at which listening comprehension unavoidably declines is still unspecified (Lang, 2020; Mo et al., 2022), and it may vary across different contexts (Kıyak et al., 2023). Although speech comprehension is seemingly effortless to humans, it is an extremely complex process. It requires identifying speech sounds, recognizing and retrieving the meaning of each word, and being aware of the sentence structure in order to understand the conveyed message (Nadig, 2013). Speech comprehension also depends on external factors, such as the rate at which the message is produced (Rodero, 2019). This rate is typically

operationalized as the number of words produced per minute (Aschim et al., 2019; Sandnes et al., 2004). It is estimated that humans speak around 150 words per minute (wpm), with individual variations (Benz, 1971). Previous research has proposed that speech comprehension begins to decline approximately between 250 and 275 wpm (Foulke & Sticht, 1969; Vemuri et al., 2004). This decline is considered the maximum information-processing capacity of the human system (Sweller, 1988, 1989).

According to the cognitive load theory (CLT), new information is initially stored in working memory before being transferred to long-term memory (Sweller, 1988, 1989). Since working memory has limited capacity, an excessive cognitive load of information can hinder transfer to long-term memory (Sweller et al., 2011). The complexity of information (intrinsic cognitive load), the manner in which it is presented (extraneous cognitive load), and the mental effort exerted by the learner (germane cognitive load), all contribute to cognitive load and have the ability to hinder effective learning (Sweller et al., 1998). For instance, Reid (1968) proposed that speech comprehension declines as grammar complexity (i.e., intrinsic cognitive load) increases in audio presentations and the respective speech rate (i.e., extraneous cognitive load) also increases, which aligns with the CLT's principles. The application of CLT to the design of multimedia presentations (i.e., information combining different modalities, such as text and images), such as animations and videos, has led to the development of the cognitive theory of multimedia learning (CTML; Mayer, 2009). CTML explores how designers should structure multimedia content and how to implement effective cognitive strategies to enhance viewers' learning efficiency (Sorden, 2012). To understand how accelerated speech impacts listening comprehension, especially when it potentially exceeds the cognitive capacity threshold, it is necessary to overview previous research on the topic.

In 1968, Foulke examined the comprehension of several speech rates to determine the optimal rate at which recorded speech should be presented. Twelve groups of participants, with 30 subjects per group, listened to an unspecified audio selection that varied in word rate, ranging from 125 to 400 wpm. Starting with the slowest word rate (125 wpm), the materials were presented in ascending order of word rate, likely not accounting for a possible order effect of the rate presentation. Afterwards, the participants' comprehension was assessed through a multiple-choice test. The results revealed that comprehension was not significantly hindered when participants were exposed to word rates between 125 wpm (baseline speed – 1x) and 250 wpm (double the speed – 2x). However, when the word rate exceeded 250 wpm, comprehension quickly declined. Foulke (1968) suggested that this decline in comprehension, as wpm increased, might have been due to a decrease in word intelligibility, related to the limited time available for word perception. While the term "intelligibility" commonly refers to the capacity to recognize individual words, the term "comprehension" concerns the interpretation of the content in a given material (Arons,

1992). Thus, as the word rate surpasses a certain threshold, the available time for listeners to perceive and recognize spoken words may be insufficient, leading to a rapid deterioration in comprehension.

In the same year, Goldhaber (1968) also conducted a study to investigate the comprehension of compressed speech. The study aimed to explore how audio difficulty level, rate of presentation, and the sex of the listener, influenced speech comprehension. The stimuli were three audio samples on the history of radio and television, averaging a total of 1894 words and 175 wpm at baseline speed. The samples were also presented at 325, 375, and 425 wpm. The difficulty of each audio sample was attributed by the Flesch Reading Ease Formula, which measures the readability of a text and attributes it to a schooling level (Farr et al., 1951). Thus, the difficulty levels of the audio samples were identified as Eighth Grade level, Graduate School level, and Freshman College level – by order of increasing difficulty. The experiment was conducted in a series of classes, with a total of 240 college students participating. After listening to the audios, at different levels of difficulty and rates of presentation (counterbalanced), participants took a comprehension test containing twenty multiple-choice questions. The findings revealed that the mean scores of the comprehension test were significantly higher at the freshman level of difficulty than at the other two levels. According to the author, this higher performance at the freshman level of difficulty may have resulted from the subjects' familiarity with this level of difficulty in college materials, since it was not measured during the experiment. Nevertheless, increasing the rate of presentation led to a quick decrease in comprehension, regardless of the difficulty level: comprehension scores significantly declined between 175 and 325 wpm (approximately 1.86x acceleration). This study also found significant differences in comprehension between male and female listeners, with male listeners outscoring female listeners at all rates of presentation and difficulty levels. However, this result should be interpreted with caution, considering that other studies did not report sex differences in learning from speeded speech (e.g., Orr et al., 1965). Moreover, Goldhaber (1968) justified that the difference in results could have been due to a fundamental difference in interest in the topic. Specifically, the history of radio and television could have been a subject which appealed more to males than females – which was an individual characteristic that was also not accounted for, during the experiment.

Reid (1968) aimed to further investigate how comprehension was affected by the rate of speech presentation across different levels of grammatical complexity. The audio material consisted of two forms (i.e., form A and form B), each with two levels of grammatical complexity (i.e., high and low complexity), and four presentation rates (i.e., 175, 275, 325, and 375 wpm). Resorting to four features of language construction, each form (A and B) was rewritten to have a less grammatically complex version: deleting unnecessary words, using the active voice, placing “because-if-when” clauses at the beginning, and

keeping sentences short and concise. A sample of 160 university students were randomly assigned to one form of the material presented at one specific speech rate and one level of grammatical complexity. After listening to the material, participants took a multiple-choice test to assess comprehension. The findings revealed that simplified versions of both material forms led to higher overall comprehension compared to the original, more grammatically complex versions. However, results suggested a lack of interaction between speech rate and grammatical complexity. This means that comprehension decreased as the presentation rate increased, regardless of grammatical complexity. Contrasting with Goldhaber's (1968) findings, comprehension levels remained similar for all groups as the rate of presentation varied from 175 wpm to 325 wpm (approximately 1.86x acceleration), suggesting that the moderate degree of speech compression could have resulted in heightened attention. However, speech comprehension sharply declined from 325 wpm to 375 wpm (approximately 2.14x acceleration), supporting the hypothesis that listening comprehension rapidly deteriorates as the word rate surpasses a certain threshold nearing double the original speed (2x). However, this study employed a between-subjects design, meaning that the subjects did not participate in all experimental conditions, hence not providing complete assurance that the necessary equivalence between groups was attained (Christensen et al., 2011).

As we are able to understand, pioneering studies on the effects of accelerated speech on listening comprehension date back to over 50 years ago. In the last few decades, there has been a scarce body of literature that uses auditory material to explore the listening comprehension of accelerated speech. The study of Yang and colleagues (2020) was one of those few contributors. The authors investigated the impact of three auditory lecture speeds on 25 graduate students' attention, cognitive load, and learning: One auditory lecture, entitled "Mastering the skills of practice" was spoken at the normal speed (1x); one lecture entitled "Office designs for creativity" was spoken at a fast speed (1.5x); and one lecture entitled "Peek performance" was spoken at a very fast speed (3x). Thus, although each participant listened to each audio in a different order, each audio had its specific speed. The authors resorted to pre- and post-comprehension tests (i.e., the same ten multiple-choice questions on both tests), a cognitive-load questionnaire (i.e., The Cognitive Load Scale, developed by Paas, 1992), and an electroencephalography (EEG) device. The whole procedure was performed in a single session. No significant differences were found in students' attention, cognitive load, comprehension, and memory between the normal (1x) and fast (1.5x) speed conditions, but overall scores were significantly lower for the faster speed (3x), and the participants' brainwaves recorded by EEG device also showed that the participants' attention was three times more intensive at this speed.

Most studies that studied the listening comprehension of accelerated speech have, therefore, proposed that increasing by 25%, or slightly more, the speed of narration in electronic media does not hamper the clarity of the narration, influence learning results, or increase perception of effort (e.g., Foulke, 1968; Friedman et al.,1967; Orr et al., 1965; Reid, 1968; Yang et al., 2020). However, it is important to address some limitations of these studies. For instance, the audio selection was often unspecified, which does not ensure that audio-related characteristics were controlled for, such as similar speech cadence, output quality of the audio, and the complexity of the topic being presented (i.e., Foulke, 1968; Reid, 1968; Yang et al., 2020). Some authors did not further contemplate the characteristics of the participant sample (i.e., individual characteristics) in its experimental design, thus not accounting for their influence on listening comprehension (i.e., Foulke, 1968; Goldhaber, 1968; Reid, 1968). For example, prior knowledge of the topic, and interest in it (Goldhaber, 1968), or factors related to participants' cognitive abilities, such as cognitive load and working memory capacity (Daneman & Merikle, 1996). Moreover, although Yang et al. (2020) assessed the participants' prior knowledge of the auditory lectures through the pre-test employed, they did not focus on the individual differences pertaining to attention, cognitive load, previous experience manipulating audio or video content, or memory capacities. Regardless, this field of research significantly slowed down in the decades following the work of Foulke, Goldhaber and Reid (1968). With the emergence of videos and their growing accessibility, research regained an interest in studying the listening comprehension of accelerated speech, mostly on multimedia material.

Studying time compression in multimedia material is a relatively recent approach to listening comprehension of accelerated speech, especially with online lectures coming to life with e-learning and video streaming platforms (e.g., YouTube; MOOCs) that allow users to manipulate the playback speed of videos and promote online learning (Mo et al., 2022; Murphy et al., 2022). In a recent study, Murphy and colleagues (2022) presented two lecture videos – one on real estate appraisals and the other on the history of the Roman Empire – to 231 undergraduate students. Participants were randomly assigned to watch both videos at only one speed: either its baseline (1x) speed (M = 162 wpm; DP = 6.36), 1.5x speed (M = 222 wpm; DP =16.97), 2x speed (M = 323 wpm; DP = 12.73), or 2.5x speed (M = 405 wpm; DP = 16.26). Listening comprehension was tested immediately and delayed (1 week) using multiple-choice and true-or-false questions. To control for prior expertise on either video topic, data from individuals with prior expertise on the content of either video were excluded from the analysis. The authors found that altering video speed had minimal impact on both immediate and delayed comprehension for participants watching videos at 1.5x and 2x speed. However, both types of comprehension significantly declined in participants watching at 2.5x speed. This learning impairment aligns with previous results,

suggesting that when the material is presented faster than double the original speed (2x), the cognitive load exceeds learners' limited cognitive resources, while the double speed itself does not seem to significantly hinder speech comprehension (e.g., Foulke, 1968).

Murphy and colleagues (2023) conducted three experiments with young adults (between 18 and 25 years old) and older adults (between 64 and 94 years old). In the first two experiments, the authors manipulated one of the lectures used by Murphy et al. (2022; i.e., on real estate appraisals) to play at three different speeds, for which each participant viewed at only one speed: 0.75x, 1x, or 2x. In the first experiment, participants viewed the lecture, predicted their performance on the video comprehension test, and then completed the comprehension test. The second experiment replicated the method and included mind-wandering probes every 60 seconds while watching the video (i.e., asking participants to describe their current cognitive state: focused on the task or distracted). Mind-wandering was defined as being distracted by information in the room or zoning out, whereas not mind-wandering was characterized by being fully focused on the task or thinking about task performance. The third experiment aimed to replicate the results of the first two experiments, by employing the same methodology as in the second experiment with another lecture from Murphy et al. (2022; i.e., on the history of the Roman Empire). Results of experiment one observed an interaction between video speed and age, such that the 2x speed did not negatively impact younger adults, but older adults' test performance was significantly hindered. The third experiment reproduced these results. Meanwhile, the second experiment revealed that mind-wandering overall decreased at 2x speed, and older adults reported less mind-wandering than younger adults. The third experiment did not replicate the mind-wandering patterns previously observed. Specifically, mind-wandering increased as the video endured for younger and older adults regardless of its speed, and both age groups reported similar rates of mind-wandering at 1x and 2x speed. Thus, the results highlight that watching videos at faster speeds may not induce cognitive overload on younger adults and can be an effective study strategy (in line with Murphy et al., 2022). However, this strategy does not seem to benefit older adults, where 2x speed consistently impaired comprehension throughout the three studies. The authors (Murphy et al., 2022, 2023), did not further account for the familiarity or interest in the lecture topics, their complexity, neither consider previous exposure to time-compressed videos, and these variables seem to be relevant when studying the listening comprehension performance of accelerated speech (Jacobson et al., 2018). For example, Tobias (1994) proposed that there is a substantial linear relationship between interest in a topic and prior knowledge of it, meaning that people know more about topics related to their interests than they do about others. In line with this, personal interest enhances learning and effort invested in tasks (Renninger et al., 1992), which benefits

comprehension of accelerated playback among learners with higher familiarity and competence in the subject (Malmberg et al., 2022). Moreover, pioneering research has found that comprehension of accelerated speech is possible after exposure to substantially increased rates of speed (e.g., Orr et al., 1965; Orr & Friedman, 1968), both leading to improved intelligibility and comprehension of the given material (Arons, 1992). Orr and Friedman (1965) studied the trainability of listening comprehension of speeded discourse. While the experimental group received systematic practice in listening to progressively increased rates of speech, the control group did not receive training. The authors demonstrated that individuals can increase listening speed to 425 wpm after ten hours of training, with a maximum 20% comprehension loss. Since there is a growing number of students who frequently accelerate video speed (Mo et al., 2022), this practice could have contributed to younger adults outperforming older adults in the comprehension of accelerated speech in the study of Murphy et al. (2023). Nevertheless, other scientific studies have reported identical research outcomes to Murphy et al. (2022, 2023) (e.g., Jacobson et al., 2018; Lang et al., 2020; Nagahama & Morita, 2017; Wilson et al., 2018), encouraging students to save time and learn more efficiently by watching videos at faster speeds with relatively little trade-off (Lang et al., 2020). Thus, according to these authors, increasing time compression – up to double their original speed – may not result in cognitive overload and should not hinder listening comprehension, especially in young adults.

Although the previously described body of research suggests that video time compression up to 2x speed promotes time-saving and even more efficient learning, mostly in undergraduate populations, other researchers suggest that speech comprehension is hindered before reaching the 2x speed (e.g., Pastore, 2009, 2012) or even before reaching 1.5x speed (e.g., Song et al., 2018). Pastore (2009) studied the effects of diagrams and time-compressed instruction on factual recall, problem-solving, and perception of cognitive load in a multimedia environment. A sample of 216 university undergraduates listened to an audio instruction of the heart. The study employed a between-subjects design: half of the participants listened only to the audio, while the other half received non-verbal support (i.e., static images) alongside the audio. The material was presented to each participant either at 1x speed (164 wpm), 1.5x speed (219 wpm), or 2x speed (328 wpm). A pretest measured students' general factual understanding of the topic to ensure that participants' prior expertise was low and equivalent across the treatments. After stimuli exposure, participants reported perceived cognitive load by responding to a seven-point Likert scale question asking how easy or difficult the materials were to understand (i.e., how much mental effort was required). Comprehension was measured through a series of achievement tests, including multiple-choice questions that assessed participants' listening comprehension. The study found that 1x and 1.5x

speed instruction in a multimedia environment produced similar scores on the comprehension test, while at 2x speed scores were significantly lower, regardless of the presence of visual support. The same pattern emerged for the self-reported cognitive load, whereas it remained similar between 1x and 1.5x, but increased substantially at 2x, suggesting that 'participants had to use more mental effort to understand the material and make connections among concepts in the instruction' (Pastore, 2009; p.502). Afterwards, Pastore (2012) employed a similar methodology, which proposed to examine the effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. In this study, Pastore increased the speed of audio narration in multimedia instruction (i.e., a lecture video on the human heart and its parts), with and without subtitles, for a participant sample of 154 university undergraduates. Similar to the previous study, Pastore employed a cognitive load measure and compared participants' performance on a comprehension test (i.e., factual and problem-solving knowledge tests) for three speeds: 1x, 1.5x, and 2x. Aligning with Pastore's (2009) previous results, no significant differences were reported between 1x and 1.5x speeds, and there was a similar self-reported level of cognitive load for both speeds. Meanwhile, watching the lectures at 2x significantly diminished performance on the comprehension test and increased self-reported cognitive load.

Song et al. (2018) further aimed to determine if the acceleration of a video lecture impacted comprehension. A sample of 54 medical students were divided into two groups, A and B. Each group watched one video on ultrasonography artefacts, and one video on transducers: Group A started with the artefacts video at 1.5x speed, then the transducers video at 1x. Group B started with the transducers video at 1.5x speed, then the artefacts video at 1x. Both videos were followed by a comprehension test consisting of open-ended and multiple-choice questions. The authors observed decreased learning at 1.5x video speed compared to baseline speed. Although students with prior exposure to subject material were excluded from the study, all participants watched their first video at 1.5x speed, not accounting for the order effect that speed might have had on the consequent video, watched at baseline speed. A possible order effect is highlighted by the fact that both videos were visualized and assessed on the same day. Regardless, similar findings in the literature suggest that video time compression may impact participants' cognitive load by reducing the time available to absorb the material (e.g., Costa et al., 2022; Ness et al., 2021; Pastore, 2010; Vemuri et al., 2004). Thus, although some authors find that some video speed acceleration do not seem to hinder learning, other authors propose that increasing speed may have the capacity to compromise the encoding of visually presented information, overwhelming attention and reducing comprehension outcomes.

Despite the differing findings, the body of research revised consistently shows that there appears to be a threshold at which video acceleration irrevocably impairs listening comprehension, similar to what pioneering studies have found (e.g., Foulke 1968; Goldhaber, 1968; Reid, 1968). The exact threshold still remains unclear in the literature, with some research suggesting it occurs when the speed is increased to over double the original speed (e.g., Murphy et al., 2022, 2023), while other studies indicate it happens before or near the double speed (e.g., Pastore et al., 2009, 2012; Song et al., 2018). This may be partially attributed to differences in stimuli selection (e.g., multimedia, complexity, emotional load) the research design employed (e.g., between-subjects), and the participant sample (e.g., familiarity, interest, previous experience manipulating audio or video speed, working memory capacity), throughout the different studies. For instance, factors related to the presence of visual stimuli (i.e., extraneous load – CLT), which require attention-diving processes, can result in higher cognitive processing and hinder learning (Foulke & Sticht, 1969; Mousavi et al., 1995; Sweller, 2010). Studies from the past decade have commonly resorted to multimedia stimuli, specifically videos with visual information, often overlooking the impact of multimedia characteristics of the participants' cognitive processing of the material (e.g., Costa et al., 2022; Jacobson et al., 2018; Mo et al., 2022; Nagahama & Morita, 2017). Moreover, research on the comprehension of accelerated speech solely through audio samples has not been accompanying the drastic growth of the use of auditory playback software as a teaching tool, such as podcasts (Pastore, 2010). Therefore, further research is needed to isolate the multimedia characteristics of video lectures, such as resorting to audio samples. Additionally, when selecting the materials (i.e., two or more audio or video samples) it is essential to ensure that the topics and their respective assessments (i.e., listening comprehension test) are equivalent in complexity, since it is considered a reliable indicator of mental load (Pastore, 2009). The selected materials should further be similar in emotional load, since it can influence information retention through the working memory (e.g., Rączy & Orzechowski, 2021). For instance, stronger feelings in a learning environment may increase attention and improve memory for learning content (e.g., Joëls et al., 2006). Emotion can be described as being composed of two main dimensions: valence and activation, the latter also commonly referred to as arousal (Pekrun et al., 2011). Valence is a subjective assessment describing how pleasant a stimulus is, whereas arousal concerns the subjective level of activation or intensity that a stimulus elicits (Stadthagen-Gonzalez et al., 2017). Chung and colleagues (2015) explored how video clips' emotional valence and arousal influenced learning performance and mental effort. The stimuli included animated instruction with different modalities (written text versus spoken text). A sample of 206 participants was randomly assigned to eight groups, experiencing different emotional states (i.e., calm positive, calm negative, arousing positive, arousing

negative) across written and spoken modalities, and learning performance was measured through free-recall tests. Results showed that arousing emotions significantly improved recall scores regardless of emotional valence, although recall scores were higher for positive emotions (M = 5.19) than negative ones (M = 4.58). Other variables, such as working memory itself, should also be acknowledged. For instance, an extensive body of research suggests that working memory capacity is a strong predictor of listening comprehension (e.g., Fay & Buchweitz 2014; James et al., 2014; Jiang & Farquharson, 2018; Molloy, 1997) and the differences between listening comprehension performance among individuals might be accounted by their working memory capacity (Murphy et al., 2023). The verbal working memory span (assessed by the Reading Span Test – Daneman & Carpenter, 1980) further correlates with a listening comprehension span, with subjects that have higher verbal working memory spans seemingly having a higher listening comprehension span (Daneman & Carpenter, 1980). Therefore, we cannot discuss auditory comprehension without addressing verbal working memory, as they are closely interconnected (Daneman & Merikle, 1996). Further acknowledgement of the research design employed is also fundamental. Similar to pioneer research in the comprehension of accelerated speech, recent studies are mostly still employing between-subject designs, although making efforts to control for participants' topic expertise of the selected audio and video samples (e.g., Murphy et al., 2022, 2023; Pastore 2009, 2012). Therefore, to allow for more experimental control, research should aim to employ within-subject designs, where all participants take part in every condition through a complete counterbalance of the materials and the speeds being studied.

Given the limitations highlighted in the reviewed research and the various factors that may influence comprehension at accelerated speeds, we designed an experimental study with a set of highly controlled audio selection. Our aim was to investigate how listening comprehension is affected by exposure to three different audio speeds: baseline (1x), fast (1.5x), and very fast (2x). This was done without any potential confounding video-related variables that could contribute to an increase in participants' cognitive load. Moreover, we employed a within-subjects design, whereby all participants listened to the same audios and were subjected to the same audio speeds, albeit in varying order, through the complete counterbalancing of the audios and the speeds. Furthermore, this study sought to acknowledge individual variables that could exert an influence on the listening comprehension of accelerated speech. These included prior experience manipulating playback speed in audio or multimedia content (i.e., assessed in a sociodemographic questionnaire), verbal working memory span (i.e., measured through the Portuguese adaptation of the Reading Span Test), previous knowledge and interest in the audio selection, and the potential arousal and valence evoked (i.e., measured in a perception

questionnaire). Other possible extraneous variables that were controlled for include characteristics of the stimuli (i.e., sound quality, speech cadence, topic complexity) and characteristics of the listening comprehension test (i.e., test accuracy and complexity), which were addressed in a pilot study for the stimuli selection. With these variables accounted for, and based on the reviewed research, we expected to observe a significant decrease in listening comprehension at the very fast speed.

Method

Participants

Forty-eight students (40 women, M_{age} = 20.91, SD_{age} = 4.08) from the University of Minho were recruited to participate in the experiment. All of them were European Portuguese native speakers, with normal hearing, and no history of neurological and/or language disorders. The experiment was approved by the local Ethics Committee (CEICSH 107/2023). Before the experiment, each participant signed a written informed consent. Their participation was rewarded with academic credits.

Materials

Three audios (i.e., A, B, and C) from the podcast "Antena 2 Ciência" were used in the current study. This Portuguese podcast uncovers, weekly, Portuguese research projects and highlights their technological applications relevant to contemporary society. Each episode is usually narrated by the researchers themselves. Specifically, Audio A is narrated by Maria João Feio, the main researcher of the study conducted by Feio et al. (2023) regarding the biological condition of rivers globally. In Audio B, Pedro Horta, holder of the Renewable Energy Chair of the University of Évora, describes a project foreseeing the implementation of floating photovoltaic solar panels in the Iberian Peninsula. Audio C is narrated by Alice Nunes, Post-Doctorate in a study ongoing until the year 2025, which describes how ecosystems are being altered by grazing in Portugal. The audios were chosen on the basis that these topics are not covered in the psychology school curriculum and, therefore, possibly unfamiliar to the majority of the participants. This would adequately control for the influence of prior knowledge on the study's results.

The episodes had similar duration (M_{min} = 10.36, SD_{min} = .09) as well as wpm (M_{1x} = 142.28, SD_{1x} = 3.11; $M_{1.5x}$ = 213.42, $SD_{1.5x}$ = 4.66; M_{2x} = 284.57, SD_{2x} = 6.22). Wpm were calculated by transcribing the episodes and dividing the total word count by the respective time of the episode in minutes. Before the experiment, the three audio samples were subjected to a pilot study with nine students from the same student population (who did not participate in the main experiment) to ensure the audio's equivalence in various variables that could affect participants' performance. This pilot was also an occasion to test the questions used to assess participants' listening comprehension. Namely, in the pilot study, each participant listened to the three audios, one for each of the three different experimental sessions, which were separated by a week. In each session, participants first listened to one

of the three audios (counterbalanced across participants) at normal speed and then completed a listening comprehension test on Qualtrics. As in previous studies (e.g., Goldhaber, 1970; Kıyak et al., 2023; Murphy et al., 2022; Ness et al., 2021; Yang et al., 2020), the test comprised 10 multiple-choice questions with four options each. Each question had only one correct response, and adopted a binary response scoring system, with participants receiving one point for each correct response, and no points for incorrect ones. Thus, the maximum score for each test was 10 points. Afterwards, participants filled in a perception questionnaire, likewise, on Qualtrics. This questionnaire included Likert scales for participants to rate audio quality (i.e., the stream quality allowing for full speech comprehension), speech cadence (i.e., the perceived speed at which the narrator was speaking), topic complexity (i.e., the perceived complexity of the audio content), test complexity (i.e., the perceived complexity of the listening comprehension test), familiarity (i.e., how familiar the participant is with the topic), interest (i.e., how interested the participant is in the topic), arousal (i.e., the activation elicited by the topic) and valence (i.e., the emotion elicited by the topic). The Likert scale for each measure went from zero to ten, with zero being the lowest score (i.e., low sound quality, speech cadence, complexity, familiarity, interest, and arousal; negative valence) and the highest score being ten (i.e., high sound quality, speech cadence, complexity, familiarity, interest, and arousal; positive valence). The scores obtained in the listening comprehension test and in each of the variables assessed by the perception questionnaire in the pilot study are presented in Table 1.

Table 1

Means, Standard Deviations (in brackets), and Friedman's ANOVA for the Audio Characteristics of the Pilot Study (N = 9)

Audio Characteristics	Audio			Friedman's ANOVA	
	A	B	C	$\chi 2(2)$	*p*
Test Accuracy	6.00 (1.86)	6.00 (1.56)	7.00 (1.09)	6.69	.035*
Sound Quality	8.78 (1.20)	9.00 (.87)	9.33 (1.00)	3.58	.167
Speech Cadence	5.33 (.50)	6.11 (1.27)	5.22 (.97)	4.67	.097
Topic Complexity	4.22 (1.79)	5.11 (1.36)	4.56 (1.51)	3.00	.223
Test Complexity	4.11 (2.77)	3.67 (2.21)	3.78 (2.39)	1.18	.554
Interest	4.33 (2.35)	4.22 (2.39)	3.56 (2.07)	1.19	.552
Familiarity	2.56 (2.40)	2.44 (2.19)	2.22 (1.99)	.07	.968
Arousal	3.33 (1.66)	2.78 (2.11)	2.22 (1.72)	.93	.629
Valence	3.56 (1.51)	6.33 (1.00)	4.00 (1.12)	10.94	.004**

* $p < .05$.

** $p < .01$.

The statistical analyses (Friedman test), conducted in SPSS Software over the listening comprehension performance and perception ratings, revealed no significant statistical differences across audios, except for test accuracy ($\chi^2 = 6.69$, $p = .035$) and valence ($\chi^2 = 10.94$, $p = .004$). Posthoc comparisons revealed that participants had more correct responses on the comprehension test for audio C compared to audio B ($p = .025$). Participants also rated audio B as more positive than audio A ($p = .003$) and audio C ($p = .018$). Despite the observed differences in these two variables, we chose to proceed with these three audios in the current experiment. The within-subject design and complete counterbalancing of the audios and speed sought to mitigate the potential impacts of these variables on the results.

In the current experiment, we further employed the Reading Span Test, in order to measure the participants' verbal working memory span. The Reading Span Test was created by Daneman and Carpenter (1980) and adapted for the Portuguese population by Gaspar (2001). Participants were instructed to read progressively longer series of sentences while trying to retain only the last word of each sentence. Each series comprised three sets of the series number. For instance, a series of two consisted

of three sets of two sentences, and a series of three comprised three sets of three sentences. The series ranged from two to six sentences, and at the end of each set, a question mark appeared on the screen – which prompted the participants to recall, in presentation order, the last words of each sentence read. For example, in a series of two, the participant should read aloud two consecutive sentences, such as "The teacher asked the students to be quiet for once; The gardener watered the flowers daily and watched them bloom" and the set would end. The screen would show a question mark, and participants should recall, aloud, the words "once" and "bloom" in that respective order.

Finally, with the goal of characterizing the participant sample and assessing previous experience manipulating audio or video content, we employed a sociodemographic questionnaire. Besides collecting sociodemographic data (i.e., gender, age, native language, and the last level of education obtained), it assessed the frequency that they resort to acceleration techniques to see and/or hear videos and audios in their daily lives, reported through a Likert scale from zero (i.e., never) to five (i.e., always).

Procedure

Participants' data were individually collected in the soundproof booths of the Human Cognition Laboratory of the University of Minho in three experimental sessions, each separated by one week, through the Qualtrics platform. Before the first session began, participants provided their written informed consent. In each session, participants were instructed to pay maximum attention to the following audio, since afterwards they would be asked to answer some questions about it (i.e., to perform a listening comprehension test) and a follow-up perception questionnaire where only their opinion was relevant. After listening to the audio, participants completed the listening comprehension test, which included the 10 multiple-choice questions from the pilot study. An attention check was added: the last question of each test instructed participants on which option to select, regardless of the audio content, ensuring they were paying attention to the questions.

Afterwards, participants filled out the perception questionnaire: they rated the audio on perceived familiarity, interest, arousal, and valence, using the same scales as the pilot study. Only these characteristics were assessed, since they are more individual-specific and could yield a wider range of responses. At the end of the last experimental session, they were asked to answer the Portuguese version of the Reading Span Test. They were also asked to answer some sociodemographic questions, as well as the frequency with which they use acceleration techniques to see and/or hear videos and audios in their daily lives. All these variables were used in the analyses reported below to test their potential influence on the listening comprehension scores obtained by participants at each speed rate.

It is also worth noting that, since the current study employs a within-participant design, all participants were exposed to the same three audios (i.e., A, B and C) and the three different audio speeds (i.e., 1x, 1.5x, and 2x), albeit in varying orders (i.e., counterbalance of the audios and speeds across participants). In the first session, audios were presented at their normal speed, which functioned as the baseline condition to which the other speed conditions (1.5x and 2x), were compared with. In all the sessions, participants were only allowed to adjust the audio volume and were instructed not to go back and forth in the audio, pause it, or listen to it more than once. The entire procedure lasted about 45 minutes per participant: 15 min in the first session; 10 to 15 min in the second (depending on the speed condition presented – 2x or 1.5x, respectively); and 15 to 20 min in the last session (depending on the speed condition presented – 2x or 1.5x, respectively).

Results

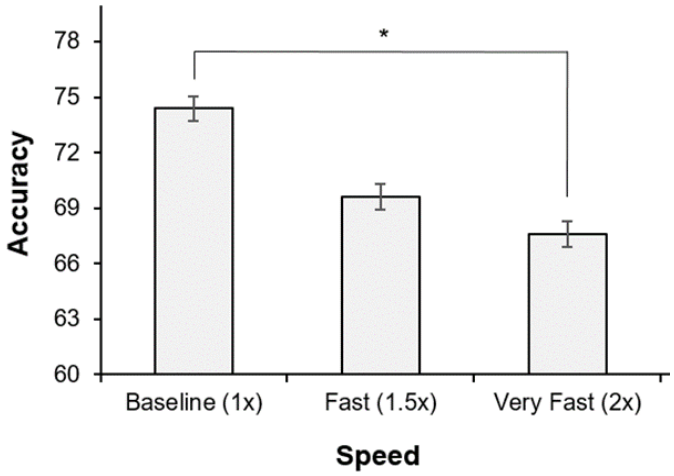
Listening comprehension (i.e., performance on the test) in the three audio speed conditions was analyzed using logistic mixed-effects (lme4) models in the R software (Bates et al., 2011; Bates et al., 2015). Before conducting the analyses, data from three participants were excluded from the dataset (6.25% of the data), one for scoring below one-third of the maximum points possible at the baseline speed (1x), another for failing to attend the attention check question, and the third for dropping out before completing the entire procedure.

The analyses reported followed a three-step approach using the R software. Firstly, we tested the effect of audio speed (baseline [1x] vs. fast [1.5x] vs. very fast [2x]) on listening comprehension, while all the other variables collected from participants were not considered in the analyses. Secondly, participants' previous experience with the acceleration technique (experience), and participants' verbal memory span (memory) were introduced as additional variables in the model. Finally, in the third approach, other variables collected from the participants (i.e., familiarity, interest, arousal, and valence) were also introduced in the model. With this procedure, we aimed to test how the successive introduction of these variables in the model might modulate the effect of audio speed on participants' listening comprehension. Thus, testing the role that individual characteristics might have on the results.

The initial logistic mixed effects (lme4) model was fitted using maximum likelihood with Laplace Approximation, with random effects for subject and item, defined as follows: accuracy ~ speed + (1 | subject) + (1 | item). The results revealed a main effect of audio speed, χ^2 (2) = 6.05, p = .049, showing that participants were more accurate in answering the listening comprehension test in the baseline condition (M = 74.4, SD = 4.37) than in very fast speed condition (M = 67.6, SD = 4.69, p = .049). However, the difference between the baseline and the fast speed condition in accuracy levels remains non-significant (M = 69.6, SD = 4.61, p = .176), as well as between the fast and very fast speed conditions (p = .482). Figure 1 depicts the percentage of the correct responses (accuracy) obtained by participants in listening comprehension tests, per audio speed conditions.

Figure 1

Percentage of Correct Responses (Accuracy) in the Listening Comprehension Test per Speed Condition

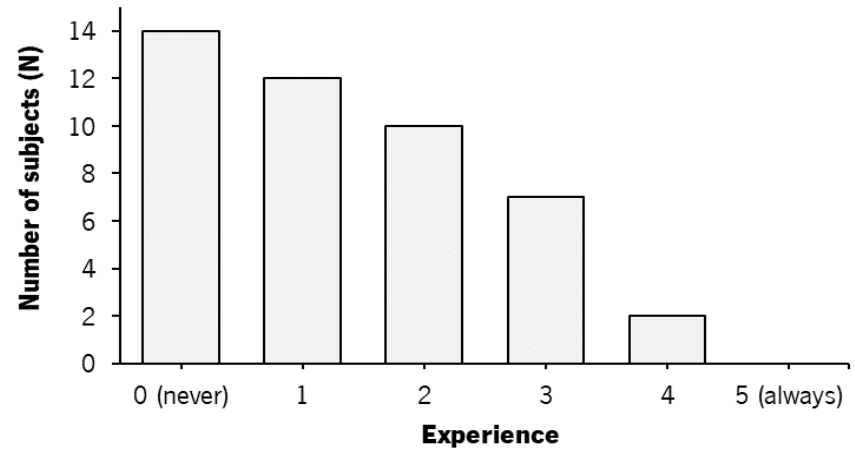


Note. Bars represent the standard error of the mean (SEM).
* $p < .05$.

In the second logistic mixed-effects model, participants' experience manipulating audio or video speed and the reading span test results were introduced in the model as co-variables (accuracy ~ speed + experience + memory + (1 | subject) + (1 | item). Experience (Figure 2) was assessed on a scale ranging from zero to five ($M = 1.36$, $SD = 1.03$), where zero represents participants who never manipulate audio or video speed, and five represents participants who always do.

Figure 2

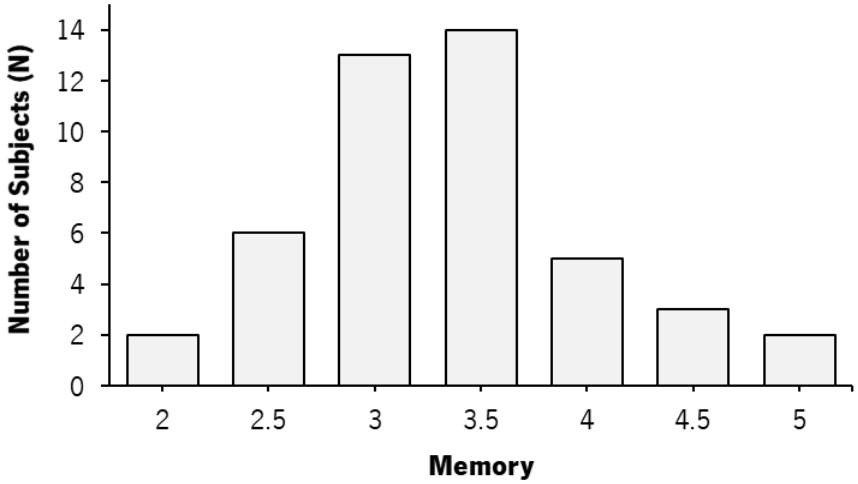
Distribution of the Participants According to Previous Experience Manipulating Speed



Meanwhile, memory (Figure 3) was assessed on a scale ranging from two to five (M = 3.32 SD = .62), each number representing the average working memory span of the participants. Results showed that when these variables were introduced in the analyses, the effect of audio speed remained statistically significant, $\chi^2(2) = 6.02$, $p = .049$. As in the previous model, the posthoc comparisons reveal that differences across speed conditions only reach statistical significance when the performance at the baseline condition was compared with the performance at the very fast speed condition ($p = .049$). Additionally, a main effect of memory was also observed, $\chi^2(6) = 19.78$, $p = .003$.

Figure 3

Distribution of the Participants According to their Average Working Memory Span



Finally, when participants' familiarity with the topic (familiarity), interest in the topic (interest), arousal, and valence evoked by each audio were introduced in the analysis (accuracy ~ speed + experience + memory + familiarity + interest + arousal + valence + (1 | subject) + (1 | item), the results were virtually the same. The mean and respective standard deviation of the data obtained in these variables can be seen in Table 2. Specifically, the results of this third model showed that the main effect of audio speed remains significant, $\chi^2(2) = 6.71$, $p = .035$.

Table 2

Mean and standard deviation (in brackets) for Familiarity, Interest, Arousal, and Valence Ratings for Each Audio per Speed

Variables	Audio								
	A			B			C		
	1x	1.5x	2x	1x	1.5x	2x	1x	1.5x	2x
Familiarity	3.14 (1.65)	2.14 (2.18)	2.14 (1.97)	2.63 (2.41)	2.83 (1.08)	2.86 (2.18)	1.17 (1.79)	1.38 (.86)	1.71 (1.39)
Interest	4.14 (1.65)	4.43 (2.21)	3.43 (1.93)	4.50 (1.23)	4.67 (.95)	4.43 (2.68)	1.83 (1.97)	2.25 (.834)	2.43 (1.19)
Arousal	3.00 (1.78)	4.14 (1.97)	2.43 (2.21)	2.63 (2.30)	2.83 (1.22)	3.71 (2.26)	1.17 (1.48)	2.38 (1.59)	1.29 (1.29)
Valence	4.43 (0.73)	4.71 (1.49)	5.43 (2.21)	6.88 (1.91)	5.33 (.75)	4.14 (1.97)	3.67 (2.30)	4.75 (2.50)	5.29 (.89)

Similar to the previous models, the posthoc comparisons reveal significant differences solely when the performance at the baseline condition is compared with the performance at the very fast speed condition ($p = .029$). A memory effect, observed in the second analysis, also reached significance, $\chi^2(6) = 26.02$, $p < .001$. An interest effect, $\chi^2(1) = 4.72$, $p = .029$ and an arousal effect, $\chi^2(1) = 5.00$, $p = .025$ were also observed.

Discussion

The main purpose of the current experiment was to understand how listening comprehension performance was affected by listening to audio at fast (1.5x), and very fast (2x) speed, when compared to performance at baseline speed (1x). This question has intrigued scientists for decades, and the rise of online learning has intensified concerns about accelerating audio and video speeds for students, due to its implications in educational practices (i.e., possibly less effective learning when accelerating a lecture speed). Regardless, research has often overlooked how individual characteristics may play a part in the listening comprehension of accelerated speech. To address the limitations of previous studies, we further aimed to recognize the influence of individual variables such as prior experience with accelerated speech, verbal working memory span, familiarity and interest in the topic, and affective reactions to it (arousal and valence). By accounting for these variables, we expected to observe a significant decrease in listening comprehension at very fast speed, compared to the baseline.

Materials were exclusively in audio format in order to control for confounding variables introduced by the multimedia materials. For that purpose, after listening to each audio, participants were asked to evaluate their familiarity with the audio's content, their interest in the topic, and their affective reactions (valence and arousal) to each of them. Afterwards, Verbal working memory span was assessed through the Reading Span Test adapted to the Portuguese population (Gaspar, 2001). Participants were also asked to answer a sociodemographic questionnaire that assessed previous experience with accelerated audio. Moreover, the characteristics of the three audio used in the experiment (i.e., sound quality, speech cadence and topic complexity), as well as characteristics of the listening comprehension test (i.e., test accuracy and complexity), underwent a prior pilot study to control for their potential effects in the results, unlike previous studies in this field of research.

The results of the experiment conducted with this highly controlled set of stimuli in a fully counterbalanced within-subject design were clear-cut. Specifically, the results from the logistic mixed-effects (lme4) models analyses conducted on participants' listening comprehension (accuracy) in each of the three experimental conditions (1x, 1.5x and 2x) showed that performance gradually declines as speech rate increases. Although differences between the baseline (1x) and fast (1.5x) conditions were not statistically significant, significant differences emerged between the baseline (1x) and very fast (2x) conditions across all logistic mixed-effects models, as expected. The results persisted even after accounting for all other variables considered in this study. Previous pioneering studies on the comprehension of accelerated speech have also proposed that while comprehension of accelerated speech can be moderately preserved until a certain threshold, it begins to rapidly decrease between 250

and 275 wpm (e.g., Foulke, 1968; Foulke & Sticht, 1969; Vemuri et al., 2004). In the current study, we notice a slight, but not significant, performance decrease at 213 wpm (1.5x speed), and a significant decrease in performance at around 285 wpm (2x speed). This aligns with previous findings that examine the impact of speech acceleration on listening comprehension, and closely resembles the average person's reading rate (i.e., 270 wpm; Monroe & Ehninger, 1974). Therefore, our results support that when the word rate nears 285 wpm, words might not be able to be processed as rapidly as they are listened to, leading to the loss of speech information (Foulke & Sticht, 1969). These results show that learners can increase their learning efficiency by speeding up their auditory lecture to a certain degree (i.e., 1.5x), and are in line with previous studies in the scientific literature (e.g., Foulke, 1968; Foulke & Sticht, 1969; Pastore, 2009, 2012; Ritzhaupt et al., 2015; Song et al., 2018; Yang et al., 2020), thus providing further support to them.

By introducing individual characteristics in the data analysis (i.e., prior experience with accelerated speech, verbal working memory span, familiarity, interest, arousal, and valence evoked by the topic), the variability introduced did not eliminate the effect of speed, whereby performance in the comprehension gradually declines, but only significantly at 2x speed – highlighting our initial prediction. If true, this could suggest that people should be cautioned that listening to or watching content at double speed (about 285 wpm or more) will impair learning. This is particularly true if the familiarity and interest for the topic are low, along with little to no previous experience accelerating audio and video content. Conversely, under the same conditions, students might save time and study more effectively by accelerating audio up to 1.5x speed (approximately 213 wpm).

Nonetheless, some variability introduced by these co-variables revealed statistical significance, such as the verbal working memory span (memory), the interest in the topic (interest) and the arousal evoked by the stimuli (arousal). These results should be carefully considered, since we cannot conclude what role they are playing in the results of the current study. For instance, since the distribution of participants across the result scale of the Reading Span Test is largely heterogeneous (Figure 3), we could hardly discuss this memory effect. Moreover, data obtained in interest and arousal (Table 2) do not show a clear pattern of these variables across each audio per speed. The significant variability introduced by these co-variables in our results can be perceived as an opportunity to explore these variables in future research. For instance, the impact of these characteristics on listening comprehension of accelerated speech could be further studied by creating groups that have low and high verbal working memory span; groups with low and high interest in the stimuli's topic; and groups that are exposed to stimuli evoking low and high arousal.

Moreover, in the current study, more people reported having none to little experience with manipulating audio or video speed, than having some or a lot of experience, and the variability introduced by this variable did not show significance. Regardless, the growing popularity of integrating accelerated audio in multimedia programs underlines the importance of determining when this technology places learners at a disadvantage (Ritzhaupt, 2008), especially since previous studies have reported a training effect: exposure to increased rates of speech improved its intelligibility and comprehension (e.g., Arons, 1992; Orr et al., 1965; Orr & Friedman, 1968). Therefore, future investigation should focus on gathering more data to better address previous experience with time-compression techniques. Participants should be screened in advance, in order to determine whether they have low or high experience with manipulating speed, and proceed employing a control and an experimental group, respectively.

Other recommendations for future research to implement include the assessment of listening comprehension through a more elaborate listening comprehension test. Although multiple-questionnaire tests have been widely used for assessing the listening comprehension of accelerated speed (e.g., Goldhaber, 1970; Murphy et al., 2022; Ness et al., 2021; Kıyak et al., 2023), the questions are based only on content recognition (Ritzhaupt, 2008), in which retrieving a memory is facilitated, since people are given clues that act as activators for information in memory (Haladyna, 2004). In order to measure the recall process, further item formats can be adopted, such as the "fill-in-the-blank" item format (Costa et al., 2022) and open-ended questions (Song et al., 2018).

The results of our experiment further suggest that the threshold where speech comprehension declines significantly might lie between accelerating speech by 1.5x and 2x of its original pace (i.e., 213 and 285 wpm, approximately). Future research could examine how different speeds within the given range (e.g., 1.6x, 1.7x, 1.8x, 1.9x) affect listening comprehension in a similar participant sample. Ideally, this sample should have a more balanced gender distribution, as our experiment predominantly involved female psychology students, which might limit the generalizability of the findings (Kıyak et al., 2023). Moreover, future research with a more diverse sample (e.g., with visually impaired individuals) could further yield results to find application in broader settings (i.e., assistive technologies for the individual impaired – Foulke, 1967).

Regardless, the current study stands out in the literature by simultaneously employing a within-participant design and a pilot study, accounting for individual characteristics and the materials, respectively. It is noteworthy to highlight that despite incorporating the individual characteristics in the data analysis, there remains a significant effect of speed, indicating an overall gradual decline in listening comprehension and information retention as audio speed increases. Thus, the current findings shed a

light on how speech acceleration, particularly in audio format, can be conducted without placing the person at a learning disadvantage. Specifically, although the baseline speed yielded the best comprehension scores, students might be able to save significant time and to complete their coursework more efficiently by accelerating learning content up to 1.5x (approximately 213 wpm) without significant listening comprehension loss, especially when there is low familiarity and interest in the topic, coupled with minimal or no prior experience in accelerating audio and video content. The application of these results is especially important in an educational context, since the participant sample of this experiment and the participant sample of the literature reviewed is mostly composed by young adult students. However, these results can also find application in other fields, such as the speech rates of audio commercials (Rodero, 2019). Economic reasons can drive the fast speech rate found in commercials, since longer commercials are more expensive than shorter ones (Rodero, 2019). However, the downside is that overly fast speech might not be well understood by target consumers, especially older ones (Murphy et al., 2023). Thus, audio commercial designers might increase the effectiveness of the message they are trying to convey if more emphasis is made upon the speed at which speech is delivered (i.e., up to 213 wpm), thus improving the processing of advertising messages.

References

Arons, B. (1992). Techniques, perception, and applications of time-compressed speech. In *Proceedings of 1992 Conference* (pp. 169-177).

Aschim, T. B., Gjerstad, J. L., Lien, L. V., Tahsin, R., & Sandnes, F. E. (2019). Are split tablet keyboards better? A study of soft keyboard layout and hand posture. In *IFIP Conference on Human-Computer Interaction* (pp. 647-655). Cham: Springer International Publishing.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation*, *8*(1), 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes [computer software] (R package, Version 0.999375-42). Available at: <http://CRAN.R-project.org/package=lme4>.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Benz, C. R. (1971). *Effects of Time-compressed Speech Upon the Comprehension of a Visually Oriented Televised Lecture*. Unpublished Doctoral Dissertation, Wayne State University.

Canale, M. (1983). From communicative competence to communicative pedagogy. In J. C. Richards, & R. W. Schmidt (Eds.), *Language and Communication (pp. 2–27)*. Longman.

Cardall, S., Krupat, E., & Ulrich, M. (2008). Live Lecture Versus Video-Recorded Lecture: Are Students Voting With Their Feet? *Academic Medicine*, *83*(12), 1174–1178. https://doi.org/10.1097/acm.0b013e31818c6902

Chen, O., Woolcott, G., & Sweller, J. (2017). Using cognitive load theory to structure computer-based learning including MOOCs. *Journal of Computer Assisted Learning*, *33*(4), 293-305. https://doi.org/10.1111/jcal.12188

Cheng, L., Pastore, R., & Ritzhaupt, A. D. (2022). Examining the Accelerated Playback Hypothesis of Time-Compression in Multimedia Learning Environments: A Meta-Analysis Study. *Journal of*

Educational Computing Research, *60*(3), 579–598. https://doi.org/10.1177/07356331211043535.

Chi, C., Chen, H. J. H., Tseng, W. T., & Liu, Y. T. (2023). Efficacy of different presentation modes for L2 video comprehension: Full versus partial display of verbal and nonverbal input. *ReCALL*, *35*(1), 105-121.https://doi.org/10.1017/S0958344022000088

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2011). *Research methods, design, and analysis (11th Edition)*. Pearson.

Chung, S., Cheon, J., & Lee, K. W. (2015). Emotion and multimedia learning: an investigation of the effects of valence and arousal on different modalities in an instructional animation. *Instructional Science*, *43*(5), 545–559. https://doi.org/10.1007/s11251-015-9352-y

Conrad, L. (1989). The effects of time-compressed speech on native and EFL listening comprehension. *Studies in Second Language Acquisition*, *11*(1), 1-16. https://doi.org/10.1017/S0272263100007804

Costa, J., Faria, A., Dias, R., Noriega, P., Vilar, E., & Rebelo, F. (2022). Video Transmission Speed and Impact on Information Retention: A Pilot Study. In *Advances in Design and Digital Communication III: Proceedings of the 6th International Conference on Design and Digital Communication, Digicom 2022, November 3–5, 2022, Barcelos, Portugal* (pp. 191-200). Cham: Springer Nature Switzerland.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge University Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422–433. https://doi.org/10.3758/BF03214546

Farr, J. N., Jenkins, J. J., & Paterson, D. G. (1951). Simplification of Flesch Reading Ease Formula. *Journal of Applied Psychology, 35*(5), 333–337. https://doi.org/10.1037/h0062427

Fay, A., & Buchweitz, A. (2014). Compreensão oral e diferenças individuais na capacidade de memória de trabalho em aprendizes iniciantes de L2. *Letrônica, 7*(1), 113–129. https://doi.org/10.15448/1984-4301.2014.1.16839

Feio, M. J., Hughes, R. M., Serra, S. R., Nichols, S. J., Kefford, B. J., Lintermans, M., ... & Sharma, S. (2023). Fish and macroinvertebrate assemblages reveal extensive degradation of the world's rivers. *Global Change Biology*, *29*(2), 355-374. https://doi.org/10.1111/gcb.16439

Foulke, E. (1967). The influence of a reader's voice and style of reading on comprehension of time-compressed speech. *Journal of Visual Impairment & Blindness*, *61*(3), 65-68. https://doi.org/10.1177/0145482X6706100301

Foulke, E. (1968). Listening Comprehension as a Function of Word Rate. *The Journal of Communication, 18*(3), 198-206.

Foulke, E., & Sticht, T. G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin*, *72*(1), 50–62. https://doi.org/10.1037/h0027575

Friedman, H. L., Orr, D. B., & Graae, C. N. (1967). Effects of listening aids and self-pacing on the comprehension of time-compressed speech. *Proceedings of the Annual Convention of the American Psychological Association, 2,* 279–280.

Gaspar, N., & da Costa Pinto, A. (2001). Amplitude de memória a curto prazo e operatória para dígitos e palavras com a adaptação da tarefa de Daneman e Carpenter (1980). https://hdl.handle.net/10216/18503

Goldhaber, G. M. (1970). Listener comprehension of compressed speech as a function of the academic grade level of the subjects. *Journal of Communication*, *20*(2), 167-173.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items (3rd ed.)*. Routledge. https://www.routledge.com/Developing-and-Validating-Multiple-choice-Test-Items/Haladyna/p/book/9781138967472

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior*, *24*(3), 786-797. https://doi.org/10.1016/j.chb.2007.02.009

Jacobson, B. P., Dorneich, M. C., & Potter, L. A. (2018). Impact of lecture video acceleration in a flipped introductory engineering course. *The International Journal of Engineering Education*, *34*(6), 1863-1875.

James, P. J., Krishnan, S., & Aydelott, J. (2014). Working memory predicts semantic comprehension in dichotic listening in older adults. *Cognition, 133*(1), 32–42. https://doi.org/10.1016/j.cognition.2014.05.014

Jiang, H., & Farquharson, K. (2018). Are working memory and behavioral attention equally important for both reading and listening comprehension? A developmental comparison. *Reading and Writing, 31*(7), 1449–1477. https://doi.org/10.1007/s11145-018-9840-y

Joëls, M., Pu, Z., Wiegert, O., Oitzl, M. S., & Krugers, H. J. (2006). Learning under stress: how does it work?. *Trends in Cognitive Sciences*, *10*(4), 152-158. https://doi.org/10.1016/j.tics.2006.02.002

John B. Voor & Rev. Joseph M. Miller (1965) The effect of practice upon the comprehension of time-compressed speech. *Speech Monographs, 32*(4), 452-454. https://doi.org/10.1080/03637756509375469

Kıyak, Y. S., Budakoğlu, I. İ., Masters, K., & Coşkun, Ö. (2023). The effect of watching lecture videos at 2× speed on memory retention performance of medical students: An experimental study. *Medical Teacher*, 1–5. https://doi.org/10.1080/0142159x.2023.2189537

Lang, D., Chen, G., Mirzaei, K., & Paepcke, A. (2020). Is faster better? A study of video playback speed. In *Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 260-269). https://doi.org/10.1145/3375462.3375466

Malmberg, J., Haataja, E., & Järvelä, S. (2022). Exploring the connection between task difficulty, task perceptions, physiological arousal and learning outcomes in collaborative learning situations. *Metacognition and Learning, 17*(3), 793-811. https://doi.org/10.1007/s11409-022-09320-z

Mayer, R. E. (2001). *Multimedia Learning.* Cambridge University Press. https://doi.org/10.1017/CBO9781139164603

Mayer, R. E. (2009). *Multimedia Learning*. Cambridge University Press.

Mo, C. Y., Wang, C., Dai, J., & Jin, P. (2022). Video Playback Speed Influence on Learning Effect From the Perspective of Personalized Adaptive Learning: A Study Based on Cognitive Load Theory. *Frontiers in Psychology*, *13*, 839982. https://doi.org/10.3389/fpsyg.2022.839982

Molloy, P. J. (1997). The role of individual differences in working memory in reading and listening comprehension in intermediate grade students. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 58*(6-A), 2072.https://doi.org/10.14288/1.0087941

Monroe, A. H., & Ehninger, D. (1974). *Principles and types of speech compression.* Glenview, IL: Scott Foreman.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, *87*(2), 319–334. https://doi.org/10.1037/0022-0663.87.2.319

Murphy, D. H., Hoover, K. M., Agadzhanyan, K., Kuehn, J. C., & Castel, A. D. (2022). Learning in double time: The effect of lecture video speed on immediate and delayed comprehension. *Applied Cognitive Psychology*, *36*(1), 69–82. https://doi.org/10.1002/acp.3899

Murphy, D. H., Hoover, K. M., & Castel, A. D. (2023). The effect of video playback speed on learning and mind-wandering in younger and older adults. *Memory*, *31*(6), 802–817. https://doi.org/10.1080/09658211.2023.2198326

Nadig, A. (2013). Listening Comprehension. *Encyclopedia of Autism Spectrum Disorders*, *1743*.

Nagahama, T., & Morita, Y. (2017). Effect analysis of playback speed for lecture video including instructor images. *International Journal for Educational Media and Technology*, *11*(1), 50-58. Retrieved from https://ijemt.org/index.php/journal/article/view/217

Ness, I., Opdal, K., & Sandnes, F. E. (2021). On the convenience of speeding up lecture recordings: Increased playback speed reduces learning. In *Innovative Technologies and Learning: 4th International Conference, ICITL 2021, Virtual Event, November 29–December 1, 2021, Proceedings 4 (pp. 461-469). Springer International Publishing.*

Orr, D. B., & Friedman, H. L. (1968). Effect of massed practice on the comprehension of time-compressed speech. *Journal of Educational Psychology, 59*(1, Pt.1), 6–11. https://doi.org/10.1037/h0025366

Orr, D. B., Friedman, H. L., & Williams, J. C. C. (1965). Trainability of listening comprehension of speeded discourse. *Journal of Educational Psychology, 56*(3), 148–156. https://doi.org/10.1037/h0021987

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Pastore, R. S. (2009). The effects of diagrams and time-compressed instruction on learning and learners' perceptions of cognitive load. *Educational Technology Research and Development, 58*(5), 485–505. https://doi.org/10.1007/s11423-009-9145-6

Pastore, R. (2012). The effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. *Computers & Education, 58*(1), 641-651. https://doi.org/10.1016/j.compedu.2011.09.018

Pastore, R., & Ritzhaupt, A. D. (2015). Using time-compression to make multimedia learning more efficient: Current research and practice. *TechTrends*, *59*(2), 66-74. https://doi.org/10.1007/s11528-015-0841-2

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students'learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Rączy, K., & Orzechowski, J. (2021). When working memory is in a mood: Combined effects of induced affect and processing of emotional words. *Current Psychology*, *40*(6), 2843-2852. https://doi.org/10.1007/s12144-019-00208-x

Reid, R. H. (1968). Grammatical Complexity and Comprehension of Compressed Speech. *Journal of Communication*, *18*(3), 236–242. https://doi.org/10.1111/j.1460-2466.1968.tb00073.x

Ritzhaupt, A. D., & Barron, A. (2008). Effects of time-compressed narration and representational adjunct images on cued-recall, content recognition, and learner satisfaction. *Journal of Educational Computing Research, 39*(2), 161–184. https://doi.org/10.2190/EC.39.2.d

Ritzhaupt, A. D., Pastore, R., & Davis, R. (2015). Effects of captions and time-compressed video on learner performance and satisfaction. *Computers in Human Behavior*, *45*, 222–227. https://doi.org/10.1016/j.chb.2014.12.020

Rodero, E. (2019). Do Your Ads Talk Too Fast To Your Audio Audience? *Journal of Advertising Research*, JAR–2019–038. https://doi.org/10.2501/jar-2019-038

Sandnes, F. E. (2004, September). One-Handed Mobile Text Entry: Evaluation of five-key text entry techniques. In *IFIP Working Conference on Mobile Information Systems* (pp. 331-339). Springer US.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132(6)*, 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Song, K., Chakraborty, A., Dawson, M., Dugan, A., Adkins, B., & Doty, C. (2018). Does the Podcast Video Playback Speed Affect Comprehension for Novel Curriculum Delivery? A Randomized Trial. *The Western Journal of Emergency Medicine*, *19*(1), 101–105. https://doi.org/10.5811/westjem.2017.10.36027

Sorden, S. (2012). The cognitive theory of multimedia learning. Handbook of educational theories. *Charlotte, NC: Information Age Publishing.*

Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods*, *49*, 111-123. https://doi.org/10.3758/s13428-015-0700-2

Stine, E. A. L., Wingfield, A., & Myers, S. D. (1990). Age differences in processing information from television news: The effects of bisensory augmentation. *Journal of Gerontology, 45*(1), 1–8. https://doi.org/10.1093/geronj/45.1.P1

Sweller, J. (1988). Cognitive load during problem-solving: Effects on learning. *Cognitive science*, *12*(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem-solving in mathematics and science. *Journal of Education Psychology, 81*(4), 457–466. https://doi.org/10.1037/0022-0663.81.4.457

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 29–47). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.004

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer New York.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/a:1022193728205

Tobias, S. (1994). Interest, Prior Knowledge, and Learning. *Review of Educational Research, 64*(1), 37–54. https://doi.org/10.3102/00346543064001037

Vemuri, S., DeCamp, P., Bender, W., & Schmandt, C. (2004). Improving speech playback using time-compression and speech recognition. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 295-302). https://doi.org/10.1145/985692.985730

Wilson, K. E., Martin, L., Smilek, D., & Risko, E. F. (2018). The benefits and costs of speed watching video lectures. *Scholarship of Teaching and Learning in Psychology, 4*(4), 243–257. https://doi.org/10.1037/stl0000127

Wood, C. D. (1965). *Comprehension of compressed speech by elementary school children*. Indiana University.

Zemlin, W. R., Daniloff, R. G., & Shriner, T. H. (1968). The Difficulty of listening to time-compressed speech. *Journal of Speech and Hearing Research, 11*(4), 875-881. https://doi.org/10.1044/jshr.1104.875

Universidade do Minho

Conselho de Ética

Comissão de Ética para a Investigação em Ciências Sociais e Humanas (CEICSH)

Identificação do documento: CEICSH 107/2023

Título do projeto: *The Need for Speed: Investigating the Influence of Audio Speed on Listening Comprehension*

Equipa de Investigação: Marta Maria Saraiva Bonjardim (IR), Mestrado em Cognição Humana, Escola de Psicologia, Universidade do Minho; Doutora Ana Paula Carvalho Soares e Doutora Helena Manuela Mendes Oliveira (orientadoras), Escola de Psicologia, Universidade do Minho

PARECER

A Comissão de Ética para a Investigação em Ciências Sociais e Humanas (CEICSH) analisou o processo relativo ao projeto de investigação acima identificado, intitulado *The Need for Speed: Investigating the Influence of Audio Speed on Listening Comprehension.*

Os documentos apresentados revelam que o projeto obedece aos requisitos exigidos para as boas práticas na investigação com humanos, em conformidade com as normas nacionais e internacionais que regulam a investigação em Ciências Sociais e Humanas.

Face ao exposto, a Comissão de Ética para a Investigação em Ciências Sociais e Humanas (CEICSH) nada tem a opor à realização do projeto nos termos apresentados no Formulário de Identificação e Caracterização do Projeto, que se anexa, emitindo o seu parecer favorável, que foi aprovado por unanimidade pelos seus membros.

O Presidente da CEICSH

Assinado por: **EMANUEL PEDRO VIANA BARBAS DE ALBUQUERQUE**
Num. de Identificação: 07089799
Data: 2023.11.02 08:49:20 +0000

(Emanuel Pedro Viana Barbas Albuquerque)

Anexo: Formulário de identificação e caracterização do projeto