

**PERCEÇÃO DE FALA COMPRIMIDA:  
UM ESTUDO EXPERIMENTAL**

PERCEPTION OF TIME-COMPRESSED SPEECH:  
AN EXPERIMENTAL STUDY

René Alain Santana de Almeida | [Lattes](#) | [reneasalm@gmail.com](mailto:reneasalm@gmail.com)  
Universidade Federal do Recôncavo da Bahia

Miguel Oliveira Jr | [CV](#) | [miguel@fale.ufal.br](mailto:miguel@fale.ufal.br)  
Universidade Federal de Alagoas

Ayane Nazarela Santos de Almeida | [Lattes](#) | [ayanensa@gmail.com](mailto:ayanensa@gmail.com)  
Universidade Federal do Recôncavo da Bahia

Oyedeji Musiliyu | [Lattes](#) | [bodeses@yahoo.fr](mailto:bodeses@yahoo.fr)  
Canada Revenue Agency

**Resumo:** Este artigo objetiva identificar quais valores de taxa de elocução são considerados aceitáveis e a partir de que ponto a taxa de elocução torna o conteúdo do enunciado ininteligível no Português Brasileiro (Pb). Para tanto, relata um estudo experimental sobre o impacto da fala comprimida na aceitabilidade e inteligibilidade de enunciados do Pb. Para os experimentos, frases curtas de áudio contendo mensagens de alerta foram utilizadas como estímulos. Essas frases foram gravadas em uma taxa de elocução natural e, em seguida, manipuladas digitalmente para taxas mais rápidas de forma escalar (de 9 a 19 sílabas por segundo). Testes de inteligibilidade e aceitabilidade foram então conduzidos com participantes cegos e videntes. Os resultados indicam que a fala comprimida tem um impacto significativo na aceitabilidade e na inteligibilidade das elocuições para ambos os grupos de participantes e que, embora os participantes cegos tendam a taxas de aceitabilidade ligeiramente mais altas em todas as condições de velocidade de fala, os participantes com visão tiveram um desempenho melhor no experimento de inteligibilidade, o que contradiz uma tendência frequentemente relatada na literatura.

**Palavras-chave:** Taxa de elocução. Aceitabilidade. Inteligibilidade.

**Abstract:** This paper aims to identify which speech rate values are considered acceptable and from which point the speech rate makes the utterance content unintelligible in Brazilian Portuguese (BP). Therefore, it reports an experimental study on the impact of time-compressed speech on the acceptability and intelligibility of utterances in BP. 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.

**Keywords:** Speech rate. Acceptability. Intelligibility.

## 1 Introdução

Ao regulamentar as propagandas de medicamentos na televisão, a Agência Nacional de Vigilância Sanitária (Anvisa) alerta que, dependendo da composição química do medicamento, o comercial deve obrigatoriamente ser acompanhado de advertências sobre os riscos relacionados a sua administração. Além disso, determina que quando as advertências escritas não forem exibidas por um período de tempo adequado à leitura, elas devem vir também na forma falada (ANVISA, 2015). Uma óbvia importância para tal regulamentação é que, ao segui-la, mensagens vitais chegam a um público que não consegue ler, como analfabetos, deficientes visuais e cegos.

O regulamento, no entanto, não estabelece nenhuma estipulação de taxa de elocução para tais avisos, não define qual seria a velocidade adequada, nem em palavras por minuto, nem em sílabas por segundo. Uma análise preliminar de alguns anúncios (propagandas) recentes de medicamentos farmacêuticos comuns em transmissões de Tv da mídia de massa indica que a taxa de elocução tende a ser 100% mais rápida nas advertências, se comparada à velocidade em que todo o anúncio é enunciado.

A velocidade média de fala de um falante nativo fluente do Português Brasileiro (Pb), de acordo com a literatura (Martins; Andrade, 2008; Oliveira Jr., 2000) varia de 3,2 a 5,5 sílabas por segundo (sil/s), dependendo de diversos fatores, como por exemplo, a ida-

de, o contexto e o conteúdo do enunciado. Não existe, no entanto, até onde sabemos, nenhum ponto de corte em termos de taxa de elocução que esteja associado à inteligibilidade e à aceitabilidade no Pb, ou seja, não há registro conhecido acerca de graus de aceitabilidade/inteligibilidade de taxas de elocução no Pb.

Levando em consideração todos os fatos acima mencionados, bem como a hipótese de que a velocidade utilizada durante um texto falado desempenha um papel crucial na aceitabilidade e inteligibilidade de enunciados, este estudo tem como objetivo explorar, por meio de um experimento perceptual, o impacto da velocidade de fala manipulada na aceitabilidade e inteligibilidade de enunciados no Pb em cegos e videntes, a fim de identificar que valores de taxa de elocução são considerados aceitáveis e a partir de que ponto a taxa de elocução torna o conteúdo do enunciado ininteligível no Pb.

Já existem alguns estudos sobre o impacto da manipulação da taxa de elocução na inteligibilidade e aceitabilidade para idiomas como alemão (Trouvain, 2007; Moos e Trouvain, 2007; Dietrich, Hertrich e Ackermann, 2013a, 2013b), holandês (Janse, 2003), japonês (Asakawa *et al.*, 2003), persa (Boroujeni *et al.*, 2017), mandarim (Feng *et al.*, 2019) e inglês (Loiotile *et al.*, 2020). A maioria desses estudos considera os indivíduos cegos ou com deficiência visual como participantes da pesquisa devido ao fato de que muitas dessas pessoas fazem uso da fala comprimida em uma série de aplicativos que são projetados especificamente para elas.

Até a elaboração do presente estudo, nenhuma pesquisa conhecida havia sido realizada sobre o impacto da manipulação da taxa de elocução na inteligibilidade e aceitabilidade para o português brasileiro que considere cegos como sujeitos da pesquisa. Para garantir que as regulamentações impostas ao uso da língua, essenciais à segurança das pessoas com necessidades especiais, como a que mencionamos acima, sejam de fato úteis, os órgãos de vigilância devem levar em consideração as particularidades da própria língua. Nesse sentido, o presente estudo também se destaca como uma contribuição para a elaboração de normas mais precisas que tratem da compressão da fala em advertências que são obrigatoriamente veiculadas em comerciais.

## **2 Aceitabilidade e Inteligibilidade de fala comprimida**

O estabelecimento de métodos de compressão artificial da fala ocorreu já em 1953, primeiro por meios mecânicos (Garvey, 1953) e depois por meios eletrônicos (Fairbanks; Everitt; Jaeger, 1974). Desde então, várias aplicações para fala comprimida foram desenvolvidas e são amplamente utilizadas em diferentes áreas. Na educação, por

exemplo, é utilizado como auxiliar de estudo; na tecnologia da fala, como meio de leitura do conteúdo de uma tela de computador; na telefonia, para permitir a revisão rápida de mensagens em sistemas de alto tráfego; na propaganda, como forma de permitir que mais informações sejam inseridas em um anúncio e reduzir seu custo.

Como acontece com qualquer sistema de comunicação, processamento ou síntese de fala, os aplicativos que utilizam fala comprimida precisam ser avaliados adequadamente. Isso geralmente é feito por meio de testes de inteligibilidade e/ou aceitabilidade envolvendo ouvintes humanos.

O conceito geral de inteligibilidade refere-se à extensão em que a fala pode ser adequadamente compreendida pelos ouvintes. De acordo com Foulke (2015) e Schmidt-Nielsen (2015), a habilidade de repetir com precisão uma palavra, um sintagma ou uma frase curta é frequentemente considerada como um índice de inteligibilidade. Os testes de inteligibilidade consistem em um procedimento avaliativo muito importante para o desenvolvimento da tecnologia da fala e estão em uso desde 1910 (Foulke, 2015).

Um procedimento típico para testar a inteligibilidade da fala comprimida é aquele em que as palavras são apresentadas, uma de cada vez, a um ouvinte. A tarefa do ouvinte é reproduzi-las oralmente ou por escrito, e sua pontuação é mensurada a partir da quantidade de palavras identificadas corretamente. Esse procedimento às vezes é chamado de teste de articulação (Foulke, 2015). Um procedimento semelhante foi adotado por Ghitza e Greenberg (2009) ao observar a inteligibilidade da fala comprimida com inserções periódicas e aperiódicas de silêncio. A inteligibilidade de enunciados de som natural comprimidos no tempo foi medida usando um paradigma que altera o ritmo da fala em uma faixa de frequências. É importante notar, entretanto, que a inteligibilidade da fala comprimida pode ser afetada por fatores relacionados ao ouvinte, incluindo idade, habilidade auditiva, fluência da linguagem e familiaridade com a fala sintetizada (Sutton *et al.*, 1995; Janse; Werff; Quené, 2007).

A aceitabilidade, por outro lado, lida com as opiniões subjetivas de como um enunciado falado é percebido. Em contraste com os métodos de medição da inteligibilidade da fala, que podem ser pontuados objetivamente, as avaliações da aceitabilidade da fala são baseadas no julgamento subjetivo do ouvinte. Os avaliadores ouvem amostras de fala e avaliam a qualidade da fala usando uma escala numérica ou rótulos categóricos (uma escala do tipo Likert), que podem ser posteriormente convertidos em números (Schmidt-Nielsen, 2015). Um procedimento muito comum usado para avaliar a aceitabilidade de fala é o *Mean Opinion Score* (MOS), que é expresso em uma escala de cinco pontos definida por rótulos de categoria, sendo 1 o pior e 5 o melhor.

Devido ao uso difundido da fala comprimida em aplicativos de computador (como leitores de tela), que geralmente são projetados para pessoas cegas ou com deficiência visual, vários estudos foram conduzidos sobre a aceitabilidade e inteligibilidade da fala comprimida com participantes cegos. O que a maioria desses estudos relata é que, em geral, os cegos apresentam melhor desempenho na compreensão da fala comprimida do que os indivíduos com visão (videntes), do grupo controle.

Em um estudo com falantes nativos de alemão, por exemplo, Trouvain (2007) demonstrou que os cegos podem entender a fala sintetizada a uma taxa de elocução de até 17 sil/s, enquanto que os videntes só compreendem a fala sintetizada a uma taxa de elocução máxima de 9 sil/s. Em outro estudo, também sobre o alemão, Moos e Trouvain (2007) investigaram até que ponto a fala poderia ser comprimida no tempo e ainda compreendida por pessoas cegas que têm prática diária com síntese de fala, em comparação com videntes sem esse treinamento. Os resultados mostraram que ouvintes cegos conseguiam entender a fala sintetizada muito melhor do que ouvintes com visão. De acordo com pesquisas prévias (Moos; Trouvain, 2007; Moos *et al.*, 2008, por exemplo), o fato de pessoas cegas serem capazes de compreender a fala em uma taxa de elocução de 22 sil/s pode estar relacionado a uma habilidade particular desses indivíduos em usar partes de seu sistema visual para melhorar o processamento da fala.

Para elucidar ainda mais o mecanismo neural subjacente a esta habilidade em particular, Dietrich, Hertrich e Ackermann (2013a; 2013b) conduziram um experimento com participantes cegos, utilizando imagem de ressonância magnética funcional (fmri) enquanto ouviam enunciados com fala comprimida. Os resultados mostraram uma ativação hemodinâmica significativa do córtex visual primário do hemisfério direito, o que indica que as estruturas visuais centrais são recrutadas durante a compreensão auditiva da fala comprimida, indicando que os cegos possuem uma acuidade auditiva elevada. Em outro estudo, Feng *et al.* (2019) mostraram que ouvintes cegos tiveram um processamento pré-atentivo de tons lexicais mais eficiente no hemisfério direito e que exibiram maior sensibilidade no processamento de consoantes e/ou vogais no estágio pré-atentivo em ambos os hemisférios em comparação com indivíduos videntes. Os resultados dessas pesquisas sugerem que cegos aperfeiçoam o processamento da fala utilizando partes do seu sistema visual.

A percepção da fala tem sido frequentemente considerada de natureza audiovisual, e o sistema visual parece estar conectado ao sistema auditivo de várias maneiras. Boroujeni *et al.* (2017) conduziram um experimento com 16 indivíduos adultos com

visão e 16 cegos, a fim de comparar o limiar de marcação de fronteira acústica de indivíduos videntes e cegos por meio de aspectos espectrais, utilizando um teste psicoacústico de delimitação do fluxo auditivo. Os resultados mostraram que ouvintes cegos exibiram habilidades de marcação de fronteiras significativamente superiores aos videntes.

Loiotile *et al.* (2020) investigaram se ouvintes cegos possuem habilidades superiores de processamento da fala e, em caso afirmativo, se essas habilidades estão relacionadas a vantagens na memória de curto prazo. Os resultados evidenciaram que ouvintes cegos, em comparação com videntes, exibiram maior desempenho nas tarefas de compreensão de sentenças e de memória de curto prazo. Tal constatação nos faz supor que cegos são mais proficientes em testes de inteligibilidade do que videntes.

Os achados das pesquisas prévias apontam para o consenso de que pessoas com deficiência visual provavelmente compensam a falta de informação visual por um aumento da entrada de outras modalidades sensoriais, incluindo o aumento da informação auditiva. Assim, parece plausível que pessoas com deficiência visual sejam capazes de usar partes de seu sistema visual para melhorar o processamento da fala, seja por uma melhor acuidade auditiva, seja por uma melhor memória de curto prazo, em relação aos ouvintes videntes. Entretanto, vale destacar que a maioria dos participantes cegos, desses estudos prévios, utilizava algum tipo de aplicativo e/ou sistema de síntese de voz com fala ultrarrápida no seu cotidiano.

### **3 Procedimentos metodológicos**

Um total de 50 textos de advertência, provenientes de várias fontes, foi pré-selecionado para este estudo. Os critérios utilizados para a seleção deste corpus preliminar foram (i) validade ecológica: os textos deveriam ser mensagens reais de advertência que são utilizadas para diferentes propósitos e (ii) tamanho: os textos deveriam ter em média 12 palavras, com uma variação entre 11 e 13 palavras. Para evitar um efeito de familiaridade no experimento de inteligibilidade, realizamos um teste de familiaridade realizado com doze estudantes especialistas em linguística. Eles foram instruídos a atribuir a cada uma das advertências uma nota de 1 (menos familiar) a 5 (mais familiar). A partir desse teste de familiaridade, um total de 36 sentenças consideradas menos familiares foi selecionado para os testes de aceitabilidade e inteligibilidade. As frases restantes foram usadas em uma sessão de treinamento realizada antes do experimento.

Todos os estímulos utilizados neste estudo foram produzidos por um falante nativo do português brasileiro do sexo masculino, com velocidade de fala normal (em torno de



5 sil/s). Deve-se observar aqui que os enunciados utilizados neste estudo não apresentaram nenhuma ocorrência de pausas - definida aqui como um “período de silêncio” maior que 150ms (Kowal; Wise; O’connell, 1983). As gravações tiveram então o parâmetro de duração manipulado no Praat (Boersma; Weenink, 2017) com o script Vocal Toolkit (Corretge, 2012). O script utiliza o método de síntese Psola, que preserva grande parte das características físicas do sinal acústico de fala, mesmo em altas taxas de compressão temporal.

Seis taxas de elocução diferentes foram utilizadas para os experimentos perceptuais: 9, 11, 13, 15, 17 e 19 sílabas por segundo. Uma dessas taxas (13 sil/s) é comumente usada nos anúncios que analisamos.

Vale destacar que, embora uma variedade de unidades de medida seja utilizada em pesquisas sobre taxa de elocução, a unidade de medida mais apropriada, de acordo com muitos autores, é a das sílabas por unidade de tempo (Blaauw, 1995; Fon, 1999; Donzel, 1999). Entretanto, tal unidade de medida pode apresentar algumas falhas, a principal, talvez seja o fato de não levar em consideração os processos que poderiam resultar na omissão de sílaba, fenômeno frequentemente encontrado na fala rápida, como assimilação e exclusão silábica (O’connell; Cowal, 1972).

O presente estudo optou pela unidade amplamente utilizada em pesquisas temporais sobre fala para fins de comparação. Os autores reconhecem os problemas relacionados a essa opção, mas presume-se que eles não sejam significativos o suficiente para invalidar a interpretação dos dados. A taxa de elocução será analisada neste estudo, portanto, utilizando-se a unidade de sílabas por segundo (sil/s).

Os estímulos foram apresentados aleatoriamente aos participantes dos experimentos por meio de um computador *notebook*. Os participantes tiveram acesso apenas à versão em áudio dos estímulos e, durante todo o procedimento, utilizaram um *headset* com fones de ouvido e microfone para que pudessem ouvir os estímulos auditivos e pudessem coletar suas respostas. Os experimentos foram realizados em uma área tranquila a fim de que ruídos externos não interferissem no experimento. Instruções foram fornecidas antes do início da tarefa bem como uma seção de treinamento foi realizada antes do experimento.

Para o experimento de aceitabilidade, um conjunto de dezoito mensagens de advertência foi apresentado aos participantes, uma de cada vez. Para cada estímulo, os participantes deveriam falar suas respostas usando uma escala numérica de 1 a 5, que representa

os seguintes conceitos: 1 “terrível”, 2 “ruim”, 3 “regular”, 4 “bom” e 5 “excelente”. Este é o mesmo procedimento adotado por Dagenais, Brown e Moore (2006), por exemplo, e é baseado no protocolo Mos. Experimentos que empregam a medida MOS geralmente consideram valores acima de 4,0 como uma indicação de boa qualidade, enquanto valores abaixo de 3,5 são considerados inaceitáveis (Carne, 2004).

Para o experimento de inteligibilidade, outro conjunto de dezoito mensagens de advertência, diferente do anterior, foi apresentado, uma de cada vez, aos participantes, que foram instruídos a repetir o que ouviram enquanto sua resposta verbal era gravada. Após o término do experimento, as transcrições das vozes gravadas foram feitas e a taxa de memorização foi definida como o número de palavras corretas em comparação com o número total de palavras na frase. Para este estudo, uma sentença foi considerada inteligível quando a taxa de memorização foi superior a 90%, seguindo os critérios adotados em Asakawa *et al.* (2003).

É importante ressaltar que, de acordo com estudos que consideraram a quantidade de itens que podem ser armazenados na memória de trabalho, um indivíduo pode recordar até 9 palavras isoladas em uma lista. No entanto, quando as palavras fazem parte de uma frase, a capacidade da memória operacional aumenta para até 16 palavras (Baddeley e Hitch, 2000). Conforme mencionado anteriormente, os estímulos utilizados no presente estudo são compostos por frases que contêm até treze palavras. Isso foi feito para evitar qualquer interferência da capacidade da memória de trabalho no experimento de inteligibilidade.

Trinta cegos (10 mulheres, 20 homens, média da idade = 39,36, desvio padrão da idade = 13,64) e trinta videntes (17 mulheres, 13 homens, média da idade = 25,66, desvio padrão da idade = 6) participaram deste estudo. Todos os sujeitos são falantes nativos do português brasileiro, alfabetizados e maiores de idade. Indivíduos com alterações auditivas, motoras e/ou cognitivas, conforme autodeclaradas, não foram incluídos no estudo.

Os dados utilizados no presente estudo, avaliado e aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Alagoas (CAAE: 34571014.3.0000.5013), foram resultantes de um estudo experimental na Universidade Federal de Alagoas e na Escola Estadual Cyro Accioly, em Maceió. Todos os participantes leram e assinaram o Termo de Consentimento Livre e Esclarecido. Sujeitos cegos tiveram acesso a uma versão do termo em Braille, impresso na Escola Cyro Accioly.



#### 4 Resultados e discussão

Na presente seção, apresentamos os resultados e discussões acerca dos testes experimentais de aceitabilidade e inteligibilidade realizados com ouvintes cegos e videntes, falantes nativos do Pb.

##### Teste de aceitabilidade

As médias (e desvios-padrão) das pontuações resultantes do teste de aceitabilidade (1 - 5) para os dois grupos de participantes nas seis taxas de elocução estão demonstradas na Tabela 1.

**Tabela 1** – Médias (e desvio padrão) das pontuações de aceitabilidade

| Taxas de elocução do estímulo | Participantes |            |
|-------------------------------|---------------|------------|
|                               | Cegos         | Videntes   |
| 9 sil/s                       | 4.6 (0.54)    | 4.4 (0.68) |
| 11 sil/s                      | 3.9 (1.03)    | 3.6 (1.04) |
| 13 sil/s                      | 3.2 (1.04)    | 2.6 (0.87) |
| 15 sil/s                      | 2.3 (1.01)    | 1.9 (0.86) |
| 17 sil/s                      | 1.8 (0.87)    | 1.5 (0.55) |
| 19 sil/s                      | 1.4 (0.45)    | 1.2 (0.50) |

Fonte: elaborada pelos autores

Como pode ser visto na Tabela 1, quanto maior a taxa de elocução, menos aceitável o enunciado é para os dois grupos de participantes. Em geral, os enunciados com velocidade de fala de 9 sil/s foram bem avaliados, enquanto aqueles com velocidade de fala de 11 sil/s foram considerados, em média, como “regular” ou “bom”, tanto por cegos quanto pelos participantes com visão. Enunciados com velocidade de fala maior ou igual a 13 sil/s foram considerados, em média, “regulares” ou “ruins” para os cegos e “ruins” ou “péssimos” para os videntes. Deve ser ressaltado aqui, mais uma vez, que esta é justamente a taxa de elocução comumente empregada para a mensagem de advertência na maioria dos anúncios de Tv e rádio que analisamos.

Os valores da Tabela 1 também indicam que os cegos tendem, em média, a maiores taxas de aceitabilidade do que os videntes em todos os casos de enunciados comprimidos no tempo. No entanto, essa diferença não foi estatisticamente significativa em todas

as taxas de elocução, conforme observado por meio de um teste *t* pareado realizado para comparar médias entre grupos de participantes em cada condição de taxa de elocução.

Os resultados dos testes *t* pareados mostram uma diferença significativa entre as pontuações de aceitabilidade apresentadas por participantes cegos e videntes nas taxas de elocução de 13 sil/s, [ $t(89) = 3,15, p < 0,05$ ]; 15 sil/s, [ $t(89) = 2,81, p < 0,05$ ] e 17 sil/s, [ $t(89) = 2,78, p < 0,05$ ]. Nas outras taxas de elocução, existem algumas diferenças entre os dois grupos de participantes, mas não são significativas.

O teste de Análise de Variância (ANOVA) com medidas repetidas indicou que as classificações de aceitabilidade foram significativamente diferentes entre as condições (taxas de elocução) para os indivíduos cegos,  $F(5, 534) = 102,6, p < 0,05$ , e também para os indivíduos com visão,  $F(5, 534) = 143,7, p < 0,05$ . No entanto, a interação observada entre taxa de elocução e grupo de participante não foi significativa,  $F(5.290) = 1,2, p > 0,05$ .

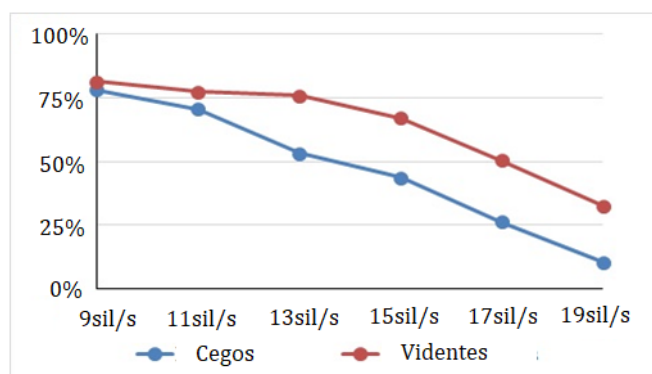
Os resultados do teste de aceitabilidade demonstram que, embora os cegos tenham julgado como mais aceitáveis os estímulos auditivos apresentados do que os videntes, os testes estatísticos mostraram não haver uma interação significativa entre os fatores taxa de elocução e grupo de participante. Tal constatação nos conduz a não poder afirmar que cegos consideram fala comprimida mais aceitável do que videntes.

### **Teste de inteligibilidade**

As médias das pontuações do teste de inteligibilidade (porcentagem de repetição correta) para os dois grupos de participantes nas seis taxas de elocução estão apresentadas na Figura 1. A tendência observada no teste de aceitabilidade também está presente aqui: quanto maior a taxa de elocução, menos inteligível o enunciado é considerado, para ambos os grupos de participantes.

A Figura 1 também destaca que nenhum grupo de participantes atingiu 90% no teste de inteligibilidade, nem mesmo com velocidade de fala de 9 sil/s. Se 90% deve ser tomado como nota mínima em um teste perceptual para que um enunciado seja considerado inteligível, conforme propõe a literatura (Asakawa *et al.*, 2003), os resultados do presente estudo sugerem que a partir de 9 sil/s, qualquer enunciado de fala comprimida será considerada ininteligível no Pb, independentemente da condição do ouvinte: cego ou vidente.

**Figura 1** – Resultados do experimento de inteligibilidade



Fonte: elaborada pelos autores

Além disso, os resultados desse experimento perceptual apontam para uma direção diferente daquela comumente relatada em estudos anteriores cujos achados mostram que pessoas cegas sistematicamente têm melhor desempenho na compreensão da fala comprimida que sujeitos com visão em experimentos perceptuais (Trouvain, 2007; Moos e Trouvain, 2007, por exemplo). A Figura 1 indica claramente que esse não é o caso no presente estudo: sujeitos com visão tiveram melhor desempenho no experimento de inteligibilidade do que sujeitos cegos, em todas as condições de taxa de elocução.

Os resultados dos testes *t* pareados mostram uma diferença estatisticamente significativa entre as pontuações de inteligibilidade apresentadas por participantes cegos e videntes nas taxas de elocução de 13 sil/s, [ $t(89) = 5,15, p < 0,05$ ]; 15 sil/s, [ $t(89) = 4,81, p < 0,05$ ]; 17 sil/s, [ $t(89) = 5,45, p < 0,05$ ] e 19 sil/s,  $t(89) = 6,0, p < 0,05$ .

Um teste ANOVA revelou uma diferença significativa entre os sujeitos,  $F(1, 58) = 20,7, p < 0,05$ , e entre as taxas de elocução,  $F(5, 290) = 271,7, p < 0,05$ . Além disso, foi observada uma interação significativa entre a taxa e o sujeito,  $F(5, 290) = 3,6, p < 0,05$ . Testes *post-hoc* revelaram que os resultados do teste de inteligibilidade diferem significativamente a partir de 15 sil/s para os participantes com visão e a partir de 13 sil/s para os cegos ( $p < 0,05$ ).

Como resultado do efeito de interação, todos os participantes de cada grupo que conseguiram repetir, em média, mais de 90% das palavras do enunciado foram considerados separadamente para cada condição de taxa de elocução. A Tabela 2 mostra o número de participantes em cada grupo com melhor desempenho, em média, entre todos os participantes no experimento de inteligibilidade.

**Tabela 2** – Participantes que repetiram mais que 90% das palavras no teste de inteligibilidade

| Taxas de elocução | Participantes |      |          |      |
|-------------------|---------------|------|----------|------|
|                   | Cegos         |      | Videntes |      |
|                   | 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         | -             | -    | -        | -    |

Fonte: elaborada pelos autores

Os resultados da Tabela 2 indicam: i) uma redução na inteligibilidade à medida que a velocidade de fala aumenta, ii) um maior número de participantes videntes, em comparação com sujeitos cegos, que tiveram melhor desempenho no experimento de inteligibilidade, em todas as condições de taxa de elocução, iii) a representatividade muito baixa, em termos de porcentagem, de participantes cegos e videntes que conseguiram realmente repetir mais de 90% das palavras após submetidos a estímulos auditivos com fala comprimida e iv) uma ausência de participantes que conseguiram repetir mais de 90% das palavras nos enunciados associados às maiores taxas de elocução (17sil/s e 19sil/s).

### 5 Considerações finais

Os resultados relatados aqui demonstram claramente que a fala comprimida tem um impacto significativo tanto na aceitabilidade quanto na inteligibilidade de enunciados no Pb. Verificou-se que a inteligibilidade diminui significativamente com o aumento da compressão do tempo, ou seja, com o aumento da taxa de elocução. Esses achados estão de acordo com os de King e Behnke (1989) e Beatty, Behnke e Froelich (1980). Em média, nenhum grupo de participantes atingiu mais de 82% de pontuação de inteligibilidade em qualquer uma das taxas de elocução manipuladas que foram utilizadas no experimento, o que é considerado uma pontuação baixa pelos padrões dos testes de inteligibilidade.

Nossos achados indicam ainda que quanto maior a taxa de elocução de um enunciado, menos provável será para os ouvintes considerá-lo aceitável. Especificamente, os enunciados com taxa de elocução de 11 sil/s foram avaliados, em média, 3,9 pelo grupo de participantes cegos e 3,6 pelo grupo de videntes. Enunciados com velocidade de fala

de 13 sil/s, muitas vezes associados a mensagens de advertência em anúncios de Tv e rádio brasileiros, foram avaliados, em média, 3,2 pelo grupo de participantes cegos e 2,6 pelo outro grupo. Como apontado antes, os protocolos de avaliação que usam a medida MOS muitas vezes consideram valores acima de 4,0 como adequados e abaixo de 3,5 inaceitáveis. Portanto, podemos concluir que nem 11 sil/s nem 13 sil/s podem ser consideradas taxas de elocução adequadas para material de fala comprimida, sendo esta última taxa também inaceitável.

As descobertas também sugerem que a avaliação da inteligibilidade e da aceitabilidade são muito semelhantes para indivíduos com visão e cegos, apesar do fato de que pessoas cegas são supostamente capazes de usar partes de seu sistema visual para melhorar o processamento da fala (Moos *et al.*, 2008; Moos; Trouvain, 2007). Embora seja verdade que os cegos tendem a apresentar, em média, taxas de aceitabilidade ligeiramente maiores para enunciados comprimidos no tempo, como os resultados deste estudo indicaram e mostraram ser estatisticamente significativos em três taxas de elocução (13 sil/s, 15 sil/s e 17 sil/s), os videntes tiveram melhor desempenho no experimento de inteligibilidade, em todas as condições de taxa de elocução, com diferenças estatisticamente significativas em quatro delas (13 sil/s, 15 sil/s, 17 sil/s e 19 sil/s).

Este último achado contradiz o que é geralmente observado na literatura para outras línguas (Trouvain, 2007; Moos; Trouvain, 2007; Dietrich, Hertrich e Ackermann, 2013a, 2013b; Boroujeni *et al.*, 2017; Feng *et al.*, 2019; Loiotile *et al.*, 2020). Talvez essa divergência se deva ao fato de que a maioria desses estudos teve como participantes, cegos que estavam acostumados a ouvir fala ultrarrápida em seus aplicativos, diferente do perfil dos nossos participantes cegos que não possuíam tal hábito, embora também utilizassem aplicativos com fala sintetizada, ela não era comprimida. Isso nos leva a concluir que talvez a exposição constante à fala ultrarrápida seja o que explica verdadeiramente o fato de os cegos compreenderem, melhor do que os videntes, enunciados com taxas de elocução elevadas e não a falta de um dos sentidos, a visão.

Pesquisas futuras devem examinar o impacto da fala comprimida em outros grupos de participantes que também se utilizam da audição para compreender as advertências em anúncios, por não conseguirem ler, como os analfabetos, por exemplo. Além disso, pretendemos investigar se e de que forma a possibilidade de alteração da taxa de elocução em aplicativos e aparelhos de *streaming* interfere na aceitabilidade e inteligibilidade de enunciados no Pb. Ou seja, pesquisar se quem está exposto constantemente a taxas de elocução elevadas, ainda que não seja cego, possui uma maior aceitabilidade e inteligibilidade de enunciados ultrarrápidos.

Os resultados do presente estudo sugerem claramente ser essencial que a fala comprimida seja utilizada com cautela em aplicações que exijam compreensão total da mensagem que ela transmite. Em particular, os resultados aqui apresentados justificam a necessidade de uma legislação mais precisa no que diz respeito à taxa de elocução das advertências nos comerciais de medicamentos, uma vez que essas advertências são de importância vital e se destinam a pessoas que não conseguem ler as advertências escritas.

## Referências

ANVISA. Regulation about pharmaceutical drug advertising. Disponível em: [http://portal.anvisa.gov.br/wps/wcm/connect/b12a03004745973d9f9adf3fbc4c6735/rdc\\_9608\\_comentada.pdf?MOD=AJPERES](http://portal.anvisa.gov.br/wps/wcm/connect/b12a03004745973d9f9adf3fbc4c6735/rdc_9608_comentada.pdf?MOD=AJPERES). Acesso em: 04 fev. 2015.

ASAKAWA, C. *et al.* Maximum listening speeds for the blind. *Proceedings of the 2003 International Conference on Auditory Display, (ICAD03)*, Boston, p. 276-279, 2003.

BADDELEY, A. D.; HITCH, G. J. Development of working memory: should the Pascual Leone and the Baddeley and Hitch models be merged? *Journal of Experimental Child Psychology*, v. 77, n. 2, p. 128-137, 2000.

BEATTY, M. J.; BEHNKE, R. R.; FROELICH, D. L. Effects of achievement incentive and presentation rate on listening comprehension. *Quarterly Journal of Speech*, v. 66, n. 2, p. 193-200, 1980.

BLAAUW, E. *On the perceptual classification of spontaneous and read speech*. Research Institute for Language and Speech, Utrecht University, 1995.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer* (Versão 6.0.36) [Computer program]. 2017. Disponível em: <http://www.praat.org/>.

BOROUJENI, F. M. *et al.* Comparison of auditory stream segregation in sighted and early blind individuals. *Neuroscience letters*, v. 638, p. 218-221, 2017.

CARNE, E. B. *A professional's guide to data communication in a TCP/IP world*. Boston: Artech House, 2004.

CORRETGE, R. *Praat vocal toolkit* [Computer Software]. 2012. Disponível em: < <http://www.praatvocaltoolkit.com/>.

DAGENAIS, P.; BROWN, G.; MOORE, R. Speech rate effects upon intelligibility and acceptability of dysarthric speech. *Clinical Linguistics & Phonetics*, v. 20, n. 2-3, p. 141-148, 2006.



DIETRICH, S.; HERTRICH, I.; ACKERMANN, H. Training of ultra-fast speech comprehension induces functional reorganization of the central-visual system in late-blind humans. *Frontiers in Human Neuroscience*, v. 7, article 701, 2013a.

DIETRICH, S.; HERTRICH, I.; ACKERMANN, H. Ultra-fast speech comprehension in blind subjects engages primary visual cortex, fusiform gyrus, and pulvinar - a functional magnetic resonance imaging (fMRI) study. *BMC Neuroscience*, v. 14, article 74, 2013b.

DONZEL, M. van. *Prosodic Aspects of Information Structure in Discourse*. Thesis (PhD). Faculteit der Geesteswetenschappen, University of Amsterdam, Amsterdam, 1999.

FAIRBANKS, G.; EVERITT, W. L.; JAEGER, R. P. Method for time or frequency compression-expansion of speech. In: DUKER, Sam (Ed.), *Time-compressed speech*, v. 1, p. 172-180. Metuchen, N.J.: Scarecrow, 1974.

FENG, J. *et al.* Effect of blindness on mismatch responses to Mandarin lexical tones, consonants, and vowels. *Hearing Research*, v. 371, p. 87-97, 2019.

FON, J. Speech rate as a reflection of variance and invariance in conceptual planning in storytelling. *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences (ICPhS)*, San Francisco, v. 14, n. 1, p. 663-666, 1999.

FOULKE, E. *The comprehension of rapid speech by the blind: part III*, 1969. Disponível em: <http://files.eric.ed.gov/fulltext/ED034346.pdf>. Acesso em: 25 fev. 2015.

GARVEY, W. D. The intelligibility of speeded speech. *Journal of Experimental Psychology*, vol. 45, n. 2, p. 102-108, 1953.

GHITZA, O.; GREENBERG, S. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, v. 66, n. 1-2, p. 113-126, 2009.

JANSE, E. *Production and perception of fast speech*. Utrecht: LOT, 2003.

JANSE, E.; WERFF, M. van; QUENÉ, H. Listening to fast speech: aging and sentence context. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, p. 681-684, 2007.

KING, P. E.; BEHNKE, R. R. The Effect of Time-Compressed Speech on Comprehensive, Interpretive, and Short-Term Listening. *Human Communication Research*, v. 15, n. 3, p. 428-443, 1989.

KOWAL, S.; WIESE, R.; O'CONNELL, D. The use of time in storytelling. *Language and Speech*, vol. 26, n. 4, p. 377-392, 1983.

LOIOTILE, R. *et al.* Enhanced performance on a sentence comprehension task in congenitally blind adults. *Language, cognition and neuroscience*, v. 35, n. 8, p. 1010-1023, 2020.

MARTINS, V.; ANDRADE, C. R. F. de. Perfil evolutivo da fluência da fala de falantes do português brasileiro. *Pró-Fono Revista de Atualização Científica*, vol. 20, n. 1, p. 7-12, 2008.

MOOS, A. *et al.* 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, p. 297-300, 2008.

MOOS, A.; TROUVAIN, J. Comprehension of ultra-fast speech-blind vs. “normally hearing” persons. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, p. 677-680, 2007.

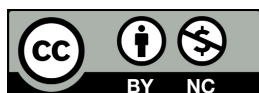
O'CONNELL, D.; KOWAL, S. Cross-linguistic pause and rate phenomena in adults and adolescents. *Journal of Psycholinguistic Research*, v. 1, n. 2, p. 155-164, 1972.

OLIVEIRA JR, M. *Prosodic Features in Spontaneous Narratives*. Thesis (PhD). Simon Fraser University, Vancouver, Canada, 2000.

SCHMIDT-NIELSEN, A. Intelligibility and Acceptability testing for Speech Technology. *Naval Research Laboratory*, 1992. Disponível em: <http://www.dtic.mil/dtic/tr/fulltext/u2/a252015.pdf>. Acesso em: 24 fev. 2015.

SUTTON, B. *et al.* Younger and older adults rate performance when listening to synthetic speech. *Augmentative and Alternative Communication*, v. 11, n. 3, p. 147-153, 1995.

TROUVAIN, J. On the comprehension of extremely fast synthetic speech. *Saarland Working Papers in Linguistics (SWPL)*, v. 1, p. 5-13, 2007.



Data de submissão: 12/12/2022

Data de aceite: 10/03/2023