

# As consoantes plosivas surdas no contexto da voz cantada: análise do *Voice Onset Time* (VOT) e sua relação com a inteligibilidade

## Voiceless plosive consonants in the context of the singing voice: analysis of Voice Onset Time - VOT and its relationship with intelligibility



**Cristina de Souza Gusmão**

Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil  
tina\_gusmao@tahoo.com.br



**Mônica Pedrosa de Pádua†**

Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil  
monicapedrosa1@gmail.com



**Fabio Wanderley Janhan Sousa**

Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil  
fabio.janhan@ufpe.br

**Resumo:** As consoantes desempenham um papel crucial na interpretação e na estruturação das palavras em textos cantados. O estudo teve como objetivo analisar se o ambiente de performance, sejam em auditórios de tamanho variado como Câmara ou Grande Teatro, influencia a duração do Voice Onset Time (VOT) das consoantes /p/, /t/ e /k/, bem como das sílabas melódicas “por”, “tor” e “quis” em uma canção. Além disso, investigou-se o VOT de duas cantoras, uma considerada menos e outra mais inteligível durante uma performance destinada à Câmara. Participaram do estudo 10 sopranos que interpretaram a música “Quando eu Morrer” (2022) em ambientes simulados de Câmara e Grande Teatro. Posteriormente, foram selecionados trechos das performances que continham as consoantes /p/, /t/, /k/ e suas respectivas sílabas melódicas para a medição do VOT. Para o segundo objetivo, o mesmo procedimento foi realizado, porém comparando apenas

os valores de VOT da performance destinada à Câmara entre a cantora menos e mais inteligível. Os resultados indicaram que não houve diferença estatisticamente significativa nos valores de VOT das consoantes surdas e sílabas entre as duas performances observadas. Contudo, ao comparar os valores de VOT entre as duas intérpretes (menos e mais inteligível), constatou-se que a cantora mais inteligível apresentou um VOT menor. Portanto, os resultados sugerem que uma duração reduzida para as consoantes surdas oclusivas pode favorecer a inteligibilidade do texto cantado.

**Palavras-chave:** Canção de câmara brasileira; *Voice Onset Time*; Inteligibilidade do texto cantado; Consoantes plosivas surdas.

**Abstract:** Consonants play a crucial role in the interpretation and structuring of words in sung texts. The study aimed to analyze whether the performance environment, whether in auditoriums of varying sizes such as Chamber or Grand Theatre, influences the duration of the Voice Onset Time (VOT) of the consonants /p/, /t/ and /k/, as well as of the melodic syllables “por”, “tor” and “quis” in a song. Furthermore, the VOT of two singers was investigated, one considered less and the other more intelligible during a performance intended for the Chamber. The study included 10 sopranos who performed the song “When I Die (2022)” in simulated Chamber and Grand Theater environments. Subsequently, excerpts from the performances containing the consonants /p/, /t/, /k/ and their respective melodic syllables were selected to measure VOT. For the second objective, the same procedure was carried out, but only comparing the VOT values of the performance destined for the Chamber between the least and most intelligible singer. The results indicated that there was no statistically significant difference in the VOT values of voiceless consonants and syllables between the two observed performances. However, when comparing the VOT values between the two performers (less and more intelligible), it was found that the more intelligible singer had a lower VOT.

Therefore, the results suggest that a reduced duration for voiceless stop consonants can favor the intelligibility of the sung text.

**Keywords:** Brazilian chamber song; Voice Onset Time; Intelligibility of the sung text; Voiceless plosive consonants.

Submetido em: 11 de junho de 2024

Aceito em: 1 de outubro de 2024

Publicado em: janeiro de 2025

## 1. Introdução

Estudos que envolvem a fala como objeto de pesquisa utilizam-se de bases físico-acústicas para compreender como os sons são produzidos, quais são as características acústicas de cada som e, ainda, como essas características podem variar entre diferentes falantes e contextos linguísticos. Essa base permite uma compreensão mais precisa e detalhada desse fenômeno complexo tanto em seus aspectos vocálicos ou consonantais quanto nos aspectos fonéticos - fenômenos de coarticulação e efeitos da prosódia nos segmentos (Camargo, 2002).

Para que haja a produção da fala, faz-se necessário uma combinação de fontes do som que podem ser: “nenhuma (silêncio), somente fonte de voz (sons ressoantes), somente fonte(s) de ruído (sons obstruintes surdos) e, finalmente, fonte de voz mesclada a fontes de ruído (sons obstruintes sonoros)” (Camargo, 2002 p. 25). Todas essas categorias apontadas, exceto a segunda, são consideradas como sons obstruintes. Esses sons são caracterizados, segundo Fant:

[...] por uma oscilação acústica sobreposta ao fluxo da respiração, causada por um objeto iniciando fricção, plosão ou interrupção da corrente aérea e podem ocorrer conjuntamente com a fonte sonora de vibração de pregas vocais (Fant, 1970). Tais fontes geram sons de natureza aperiódica no aparelho fonador, mais especificamente na região supraglótica e são representadas pelas fontes de ruído transiente e contínuo, características respectivamente das consoantes plosivas e fricativas e na sua combinação no caso das africadas. As fontes de ruído referem-se aos eventos acústicos primários no trato vocal, responsáveis pela geração de sons cochichados, aspirados, fricativos e plosivos (Fant, 1973 *apud* Camargo, 2002, p. 25).

As alterações na corrente de ar decorrentes da ação dos articuladores revelam uma turbulência no aparelho fonador, responsável por produzir uma variedade de sons, incluindo os fricativos, que são caracterizados pela fricção, e os plosivos, caracterizados por um ruído transiente (Camargo, 2002).

Em relação às consoantes no português brasileiro, elas podem ser divididas, segundo o modo de articulação, em oclusivas e constrictivas. Com relação ao ponto articulatorio, ou seja, o lugar onde a corrente de ar é articulada, podemos classificá-las como bilabiais (produzidas por meio do contato dos lábios superior e inferior), labiodentais (quando há o encontro do lábio inferior com os dentes superiores), linguodentais (produzidas pelo contato da língua com os dentes superiores), alveolares (produzidas pelo contato da língua tocando a região alveolar dos dentes superiores), palatais (produzidas pelo contato do dorso da língua com o palato duro) e, por fim, velares (produzidas pelo contato da parte superior da língua com o palato mole).

Além disso, as consoantes podem ser classificadas como orais ou nasais, surdas ou sonoras. As consoantes surdas ou desvozeadas não têm vibração das pregas vocais durante sua produção e as sonoras ou vozeadas são consoantes que têm vibração das pregas vocais. Para que haja a produção dos sons surdos, faz-se necessário uma obstrução no trato vocal. Tal obstrução associada à passagem do ar, gera um ruído que depende da abordagem da cavidade à frente da constrição e do ponto onde o estreitamento no trato vocal acontece.

Neste contexto, há um intervalo de tempo, medido em milissegundos, que existe entre a soltura de uma consoante plosiva e o início da vibração vocálica do segmento após a consoante, importante para “revelar o controle no tempo das ações motoras no aparelho fonador” (Borden *et al.*, 1994 *apud* Camargo, 2002, p. 28). A esse intervalo de tempo dá-se o nome *Voice Onset Time* – VOT.

O *VOT*, tempo de início da voz, é definido como o intervalo entre a liberação de uma constrição oral e o início da vibração das pregas vocais (McCrea; Morris, 2005) ou seja, a “posição das pregas vocais abduzidas e a gradativa adução para caracterizar a sonoridade da vogal seguinte” (Camargo, 2002, p. 27).

O *Voice Onset time (VOT)* é utilizado exclusivamente nas consoantes oclusivas, por serem as únicas a realizar uma oclusão total no trato vocal (Klein, 1999). Esse parâmetro pode ser classificado como *VOT* positivo, que é quando a explosão do ar antecede a vibração das pregas vocais, como acontece, por exemplo, nos sons plosivos surdos do Português Brasileiro (PB) /p/, /t/ e /k/. O *VOT* zero, padrão no qual o início da vibração das pregas vocais ocorre de forma simultânea à soltura do segmento plosivo, e o *VOT* negativo, padrão no qual o início da vibração das pregas vocais antecede à soltura do segmento vocálico (Kampff, 2018; Lara; Battisti, 2014; Schwartzhaupt; Alves, 2014).

As consoantes oclusivas, também chamadas de plosivas, são, do ponto de vista articulatorio, sons da fala caracterizados por uma obstrução total e momentânea da corrente de ar iniciada nos pulmões, em algum ponto da cavidade oral. No português, esta obstrução pode ocorrer nos lábios, nos dentes ou no véu palatino, resultando, respectivamente, em oclusivas bilabiais, dentais e velares, a saber, /p, b/, /d, t/ e /k, g/ respectivamente surdas e sonoras (Klein, 1999, p. 6). De acordo com Cho e Ladefoged (1999), “quanto mais rápido o movimento articulatorio, menor o valor de *VOT*. Além da velocidade do movimento articulatorio, aspectos aerodinâmicos e diferenças na massa dos articuladores podem influenciar as medidas” (Cho; Ladefoged, 1999 *apud* Lara; Battisti, 2014, p. 40).

É de suma importância ressaltar a necessidade de uma articulação correta das consoantes para garantir a compreensão do texto cantado. De acordo com DiCarlo (1994), quando as consoantes não são devidamente articuladas devido à prolongação excessiva

das vogais, o entendimento do texto acaba sendo prejudicado. Vale salientar que a habilidade de articular, tanto na fala como no canto, vai além dos aspectos técnicos vocais, englobando também a capacidade do cantor em possuir um articulador eficiente.

Estudos envolvendo a duração das consoantes plosivas surdas já são comuns na linguística e na fonoaudiologia (Kupske; Oliveira, 2020; Schwartzaupt; Alves, 2014; Melo; Martins, 2014; Constantini; Rocha; Mourão, 2013; Barroco *et al.*, 2007; Klein, 1999) principalmente no que diz respeito aos aspectos relacionados à fala e à comparação entre falantes. Entretanto, os estudos que abordam a relação do VOT no canto ainda são restritos. Alguns autores (Nolan; Sykes, 2015; DiCarlo, 2007; Nelson; Tiffany, 1968) demonstraram em seus estudos que tanto as frequências agudas quanto a má articulação das consoantes comprometem a compreensão do texto cantado.

Na busca de referências sobre essa temática, encontramos alguns estudos (McCrea; Morris, 2005 e 2007; Medeiros, 2002; Swartz, 1992). Dentre estes, o mais relevante para este estudo, visto que aborda também o contexto de voz cantada, é o estudo de Medeiros (2002), que, ao comparar a duração das consoantes surdas no padrão de fala e canto, observou que no canto as vogais são prolongadas e as consoantes são produzidas em um tempo mínimo. Descreve ainda que mesmo “encurtadas em termos absolutos e relativos, as consoantes são responsáveis pela manutenção do texto da canção” (Medeiros, 2002, p.163). Já o estudo de McCrea e Morris (2007) comparou os valores de VOT entre cantoras treinadas e não treinadas em tarefas de fala e canto. Os autores constataram que não houve diferença no valor de VOT entre as cantoras treinadas e não treinadas. Já em relação à tarefa de fala e canto, concluíram que a consoante /p/ foi mais longa na tarefa de fala do que no canto. Além disso, na tarefa de canto, o valor da consoante /p/ foi menor entre as cantoras treinadas (McCrea; Morris, 2005).

Após a explanação dos aspectos consonantais do português brasileiro, este estudo tem como objetivo verificar primeiramente se o ambiente de performance destinado à Câmara e Grande teatro são capazes de influenciar na duração do VOT das consoantes oclusivas surdas /p/, /t/ e /k/ e suas sílabas melódicas “por”, “tor” e “quis”. Em seguida, verificar se a duração das consoantes /p/, /t/ e /k/ se diferem entre duas cantoras, uma considerada menos e outra mais inteligível. Por último, este artigo fez parte de um estudo mais amplo, inserido no contexto de um doutorado, intitulado “A inteligibilidade do português brasileiro cantado na canção de câmara brasileira: uma investigação analítica perceptivo-auditiva e acústica”.

## 2. Metodologia

Este estudo foi submetido e aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais sob o número 4.890.170. Todos os participantes admitidos no estudo assinaram o Termo de Consentimento Livre e Esclarecido, conforme resolução 196-96 CNS, a fim de que tenham conhecimento dos objetivos do estudo, bem como da metodologia utilizada. A pesquisa foi realizada respeitando as condições éticas de seu emprego e preservando as identidades dos sujeitos.

As coletas de dados foram realizadas em dois momentos distintos e serão apresentadas a seguir. A coleta de dados 1 descreve a metodologia de pesquisa do primeiro objetivo deste estudo e a coleta de dados 2 apresenta o processo metodológico do segundo objetivo.<sup>1</sup>

### 2.1 Coleta de dados 1

Para este primeiro momento, a pesquisa contou com a colaboração de dez cantoras líricas profissionais classificadas como soprano e com mais de cinco anos de experiência. O convite

<sup>1</sup> Canção inédita do professor de canto e compositor Mauro Chantal (1971), com poema de Suzanna de Campos (1907-1987), composta especialmente para a tese de doutorado intitulada: A inteligibilidade do português brasileiro cantado na canção de câmara brasileira: uma investigação analítica perceptivo-auditiva e acústica, da primeira autora. Os dados deste artigo compõem a tese.

foi feito por telefone e por meio de aplicativos como Instagram e WhatsApp. As cantoras que aceitaram ao convite foram orientadas quanto aos objetivos e procedimentos realizados.

Para a coleta de dados, as cantoras foram orientadas a gravar uma canção inédita intitulada Quando eu morrer (2022) gravada em um estúdio profissional na cidade de Belo Horizonte, Minas Gerais. A escolha pelo estúdio se deu devido às melhores condições de captação do som de forma a privilegiar uma análise eficiente e eliminar qualquer interferência de ruído proveniente de um ambiente real de concerto. Todos os sopranos deste estudo receberam a partitura e o áudio do acompanhamento do piano com trinta dias de antecedência. As cantoras foram solicitadas a interpretar a canção em dois momentos: no primeiro, imaginando uma performance destinada à Câmara; e, no segundo momento, destinada a um Grande teatro.

Para a captação do áudio, utilizamos uma interface RME UFX+ e um microfone capacitivo de cápsula grande modelo Rode NT2-A com padrão cardióide e sem qualquer atuação do filtro interno. Todas as cantoras foram orientadas a cantar de pé, de frente a um pedestal com uma distância de 20 cm entre o corpo e o microfone. O microfone foi posicionado apontando para o queixo da cantora para redução da captação de sinais provenientes do deslocamento do ar. Para cada cantora, ajustamos a altura do pedestal e do microfone, sendo posicionados sempre no mesmo ponto da sala mantendo sua distância relativa das paredes e do tratamento acústico.

A fim de testar o tempo da duração das consoantes entre os ambientes de performance, foram selecionados os compassos 1 ao 9 da canção Quando eu morrer, que continha, na mesma frase, as três consoantes surdas plosivas, /p/, /t/ e /k/ e suas respectivas sílabas melódicas “por, “tor” e “quis”, como podemos verificar na Figura 1, a seguir.

Figura 1 – excerto para análise das consoantes plosivas surdas representada pelos compassos 1 a 9

1

*Quando eu morrer...*

Mauro Chantal  
Poema de Suzanna de Campos (1894-1945)  
Sabará, 21 de janeiro de 2022.  
Para Cristina Gusmão. *delicado*

mp  
p

Quan - do eu mor - rer, há s de sen - tir sau - da - de Do nos - so lin - do a - mor, meu so - nho

5  
tris - te, Que gutan - to quis tor - nar fe - li - ci - da - de, Mas foi em vão... por - que e - la não e -

9  
xis - te.

Fonte: Gusmão (2024).

Descrição da imagem: partitura tradicional dos nove primeiros compassos da canção "Quando eu morrer" para voz e piano do compositor Mauro Chantal com poema de Suzanna Campos.

Na partitura, essas consoantes estão representadas na ordem /k/, /t/ e /p/, respectivamente, e as durações silábicas por "quis", "tor" e "por" nas notas Fá3, Fá4 e novamente na nota Fá4. Adotamos aqui a padronização do Dó central como sendo o Dó3 do piano, segundo o sistema francês. A primeira sílaba, "quis", encontra-se em posição tônica, já as demais, "tor" e "por", em posição pretônica. Todas as consoantes e sílabas foram retiradas dos contextos de voz cantada. A Figura 2, a seguir, nos mostra as sílabas analisadas destacadas em laranja nos compassos 6 e 7, retiradas das palavras "quis", "tornar" e "porque".

**Figura 2** – Consoantes plosivas surdas [p], [t] e [k] e suas respectivas sílabas “por”, “tor” e “quis” representada pelo retângulo em laranja no pentagrama

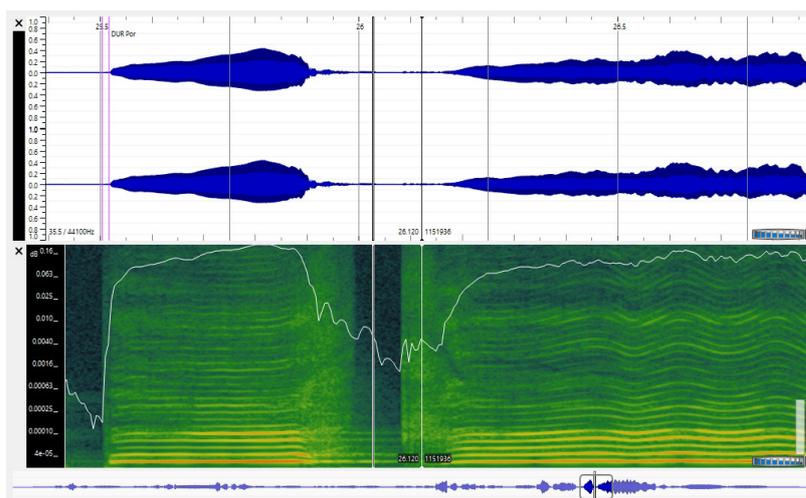


Fonte: Gusmão (2024).

Descrição da imagem: partitura tradicional do trecho analisado indicando as três notas musicais com suas respectivas sílabas de onde foram extraídas as consoantes plosivas “por”, “tor” e “quis”.

Para esta avaliação foi utilizado o programa de análise acústica *Sonic Visualiser*, que nos forneceu a espectrografia do trecho avaliado. As análises dos excertos foram realizadas individualmente para cada uma das 10 cantoras, tanto para a *performance* destinada à Câmara quanto para um Grande teatro. Posteriormente, foi selecionada a consoante avaliada e a duração do *VOT* até a vocalização da vogal subsequente. Em seguida, medimos o tempo da duração da sílaba melódica. A Figura 3 ilustra um recorte da seleção do tempo da consoante e da sílaba melódica “por”.

**Figura 3** – Espectrografia vocal: representação da consoante surda /p/ com a duração do *VOT* demarcada pelos traços paralelos em vermelho na vertical e a duração da sílaba “por” representada pelo traço vermelho na horizontal



Fonte: Gusmão (2024).

Descrição da imagem: na metade superior observamos a forma de onda do sinal gravado e na metade inferior a análise espectrográfica, em ambos observamos uma linha vermelha vertical que indica o início de cada consoante e o início da vogal, permitindo-nos medir a duração da consoante.

A fim de verificar se houve diferenças entre as durações das consoantes e das sílabas entre as performances foram realizadas as análises estatísticas<sup>2</sup>, posteriormente à verificação de normalidade dos dados por meio do teste de Shapiro-Wilk<sup>3</sup>, que comparou a média dos valores das consoantes oclusivas /p/, /t/ e /k/ das dez cantoras em performance destinada à Câmara com a média dos valores das mesmas consoantes, mas em performance destinada ao Grande teatro. A mesma comparação foi realizada para cada uma das sílabas melódicas “por”, “tor” e “quis”. Neste estudo, adotamos um nível de significância  $\alpha = 5\%$  em todos os testes estatísticos utilizando *software* livre R<sup>4</sup>. Isso significa que consideramos um resultado estatisticamente significativo se o p-valor obtido nos testes estatísticos for menor que 0,05.

## 2.2 Coleta de dados 2

Para o experimento dois foi utilizado o mesmo excerto representado pela Figura 1 deste artigo. Primeiramente, os áudios das 10 sopranos interpretando a performance destinada à Câmara foi submetido à escuta de três docentes universitários de canto, atuantes há mais de sete anos na docência. Cada professor ouviu os excertos e selecionou, de acordo com suas percepções, qual a cantora entre as 10 observadas apresentou um texto menos inteligível e qual foi o mais inteligível. A escolha da cantora menos inteligível foi unânime entre os participantes. Já a cantora mais inteligível foi escolhida por 66,66% dos professores. Após essa verificação, foram separados somente dois áudios, o áudio da cantora menos inteligível, nomeado como cantora 1, e o da cantora mais inteligível, nomeado como cantora 2.

2 Foram verificadas quatro variantes: ANOVA, *t de student*, *Tukey* e *shapiro-Wilk*. A normalidade dos dados é uma suposição e verificação necessária para a realização de métodos estatísticos paramétricos, como a Análise de Variância (ANOVA) e os de comparação de médias *t de Student* e *Tukey*. O teste de *Shapiro-Wilk* é uma opção de método estatístico utilizado para verificar a normalidade de um conjunto de dados. O teste *t de Student* é uma ferramenta estatística usada para comparar médias entre dois grupos independentes que seguem uma distribuição normal de probabilidade. ANOVA é uma técnica estatística utilizada para verificar se existe diferença entre médias de três ou mais grupos independentes. O teste de *Tukey* é um teste de comparação múltipla usado para identificar quais grupos têm médias significativamente diferentes entre si, geralmente realizado após uma ANOVA. Os testes permitem determinar se as diferenças observadas são estatisticamente significativas ou se podem ter ocorrido por acaso, a partir da definição de um nível de significância  $\alpha$ .

3 O teste de *Shapiro-Wilk* é uma opção de método estatístico utilizado para verificar a normalidade de um conjunto de dados.

4 No *software* R é possível observar a porcentagem da variação da resposta, que é explicada por um modelo linear. Disponível em: [https://www.r-studio.com/pt/Data\\_Recovery\\_Download.shtml](https://www.r-studio.com/pt/Data_Recovery_Download.shtml).

Em seguida, comparou-se os valores de VOT das consoantes /p/, /t/ e /k/ da cantora menos e mais inteligível a fim de verificar se havia diferença na duração do VOT e sua relação com a inteligibilidade.

### 3. Resultados

De posse dos valores de VOT das consoantes oclusivas /p/, /t/ e /k/ e das sílabas melódicas “por”, “tor” e “quis”, foram submetidos à análise estatística<sup>5</sup>, cujos resultados foram descritos na Tabela 1.

**Tabela 1 – Médias de VOT das consoantes oclusivas /p/, /t/ e /k/ e médias do Tempo de Duração das sílabas melódicas “por”, “tor” e “quis” obtidas a partir de cada performance por meio do teste *t de student***

|                      | Performance                    |               | p-valor |
|----------------------|--------------------------------|---------------|---------|
|                      | Câmara                         | Grande teatro |         |
| Consoantes oclusivas | Média de VOT (ms)              |               |         |
| P                    | 29,38                          | 26,00         | 0.5577  |
| T                    | 24,15                          | 38,34         | 0.0859  |
| K                    | 70,79                          | 76,97         | 0.4268  |
| Sílabas melódicas    | Média de Tempo de Duração (ms) |               |         |
| Por                  | 494,00                         | 473,10        | 0.4819  |
| Tor                  | 545,50                         | 560,60        | 0.6327  |
| Quis                 | 534,50                         | 544,70        | 0.7942  |

Para haver diferença significativa entre as médias pelo teste t, o p-valor deve ser menor do que 0,05.

Fonte: Gusmão (2024).

Descrição da imagem: A imagem acima é uma tabela contendo as consoantes oclusivas /p/, /t/ e /k/ e as sílabas melódicas “por”, “tor” e “quis” distribuídas na vertical do lado esquerdo da tabela. Na coluna do meio estão distribuídos os valores da média de VOT para performance de câmara, sendo estes 29.38, 24.15, 70.79, 494, 545.5 e 534.5 milissegundos, respectivamente, e, ao lado, para performance em grande teatro, os tempos de 26, 38.34, 76.97, 473.1, 560.6 e 544.7. Na coluna da extrema direita temos os valores de -p, sendo estes 0.5577, 0.0859, 0.4268, 0.4819, 0.6327 e 0,7942.

De acordo com a Tabela 1, pode-se verificar que tanto as médias de VOT (ms) das consoantes oclusivas quanto as médias do tempo de duração (ms) das sílabas melódicas não apresentaram diferença significativa entre as performances. Esse dado sugere que as

<sup>5</sup> Após a análise estatística, o resultado do teste de *Shapiro-Wilk* para as variáveis VOT (ms) e Tempo de Duração das sílabas em milissegundos (ms) mostraram que as variáveis seguem uma distribuição normal de probabilidade a um nível de significância  $\alpha = 5\%$  e, portanto, pôde ser realizado o teste *t de Student* de comparação de médias.

cantoras deste estudo não apresentaram diferenças significativas em relação ao tempo da articulação das consoantes e da duração das sílabas melódicas ao cantar com uma performance destinada à Câmara e ao Grande teatro. Como não houve diferença entre os parâmetros de VOT na duração das consoantes oclusivas nem das sílabas entre os ambientes de performance, decidimos continuar com a análise das 10 cantoras, mas, agora, a fim de verificar se havia diferença entre a duração das consoantes e das sílabas melódicas entre si, independente do ambiente de performance. Para isso, foi realizada outra análise, comparando as consoantes surdas /p/, /t/ e /k/ e as sílabas “por”, “tor” e “quis”, considerando-as como um único ambiente de performance.

Dessa forma, os 10 valores referentes ao VOT das consoantes oclusivas obtidos a partir da performance destinada à Câmara foram unificados aos dez valores de VOT obtidos a partir da performance destinada ao Grande teatro, resultando, portanto, em um conjunto de dados formado por 20 observações de VOT para cada consoante oclusiva.

Após realizar uma análise preliminar dos dados de VOT de cada consoante oclusiva e do tempo de duração de cada sílaba melódica, verificou-se que alguns conjuntos de dados não atenderam às pressuposições necessárias para a aplicação direta da estatística ANOVA, sendo necessário realizar a transformação dos dados para atender às pressuposições da ferramenta estatística.

Por outro lado, os conjuntos de dados formados pelos 20 valores de tempo de duração de cada sílaba melódica atenderam às pressuposições da ANOVA, viabilizando sua aplicação direta sem necessidade de transformação de dados. Os resultados da ANOVA para os dados transformados de VOT e para os dados de sílaba melódica estão apresentados na Tabela 2.

**Tabela 2 – Resultados da ANOVA para as variáveis VOT (ms) das consoantes oclusivas e tempos de duração das sílabas melódicas (ms)**

| Fonte de variação | GL | SQ     | MQ    | F     | p-valor |
|-------------------|----|--------|-------|-------|---------|
| Consoante         |    |        |       |       |         |
| Oclusivas         | 2  | 2,574  | 1,287 | 39,31 | ***     |
| Resíduos          | 57 | 1,866  | 0,033 |       |         |
| Total             | 60 |        |       |       |         |
| Sílabas melódicas | 2  | 54352  | 27176 | 5,161 | **      |
| Resíduos          | 57 | 300167 | 5266  |       |         |
| Total             | 60 |        |       |       |         |

\*p-valor <0,05. \*\* p-valor<0,01. \*\*\*p-valor <0,001. SQ: Soma dos Quadrados, GL: Graus de Liberdade, MQ: Média dos Quadrados, F: Estatística F, p-valor: Valor-p associado ao teste F.

Fonte: Gusmão (2024).

Descrição da imagem: A tabela acima mostra os dados da estatística Anova sendo apresentados os valores para consoante oclusiva de 2 para grau de liberdade, 2,574 para soma dos quadrados, 1,287 para média dos quadrados, 39,31 para estatística F e valor -p inferior a 0,001. Para os resíduos temos os valores de 57 para o grau de liberdade, 1,8666 para soma dos quadrados e 0,033 média dos quadrados. Para as sílabas melódicas temos os valores de 2 para grau de liberdade, 54352 para a soma dos quadrados, 27176 para a média dos quadrados, 5,161 para a estatística F e valor -p inferior a 0,01. Para os resíduos temos os valores de 57 para o grau de liberdade, 300167 para a soma dos quadrados e 5266 para a média dos quadrados.

Após a análise, pôde-se observar na Tabela 2 que o resultado na ANOVA indica que houve diferença significativa entre os VOT das consoantes oclusivas e entre os tempos de duração das sílabas melódicas. Ou seja, pelo menos uma consoante oclusiva difere das demais. Semelhantemente, o tempo da duração de pelo menos uma das sílabas melódicas também se difere das demais. Em decorrência desse resultado, foi realizado o Teste de Tukey, que identificou quais das consoantes diferem entre si e quais das sílabas melódicas diferem entre si quanto ao tempo de duração. O resultado do teste de Tukey está apresentado na Tabela 3.

**Tabela 3 – Médias do VOT (ms) das consoantes oclusivas /p/, /t/ e /k/ e do tempo de duração (ms) das sílabas melódicas “por”, “tor” e “quis”**

| Consoantes Oclusivas | VOT (ms)* | Sílabas Melódicas | Tempo de Duração (ms)* |
|----------------------|-----------|-------------------|------------------------|
| P                    | 19,9 a    | Por               | 483,55 a               |
| T                    | 21,5 a    | Quis              | 539,60 b               |

| K                                                                                                                                                                                                                                                                                                                                                                                                | 50,1 b | Tor | 553,05 b |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|----------|
| *Na coluna, médias seguidas de mesma letra não diferem estatisticamente entre si, pelo teste de Tukey a 5% de probabilidade. As médias foram obtidas a partir da <i>performance</i> das 20 cantoras (10 <i>performances</i> destinadas à Câmara e 10 <i>performances</i> destinadas ao Grande teatro). As médias de <i>VOT</i> das consoantes oclusivas são provenientes de dados transformados. |        |     |          |

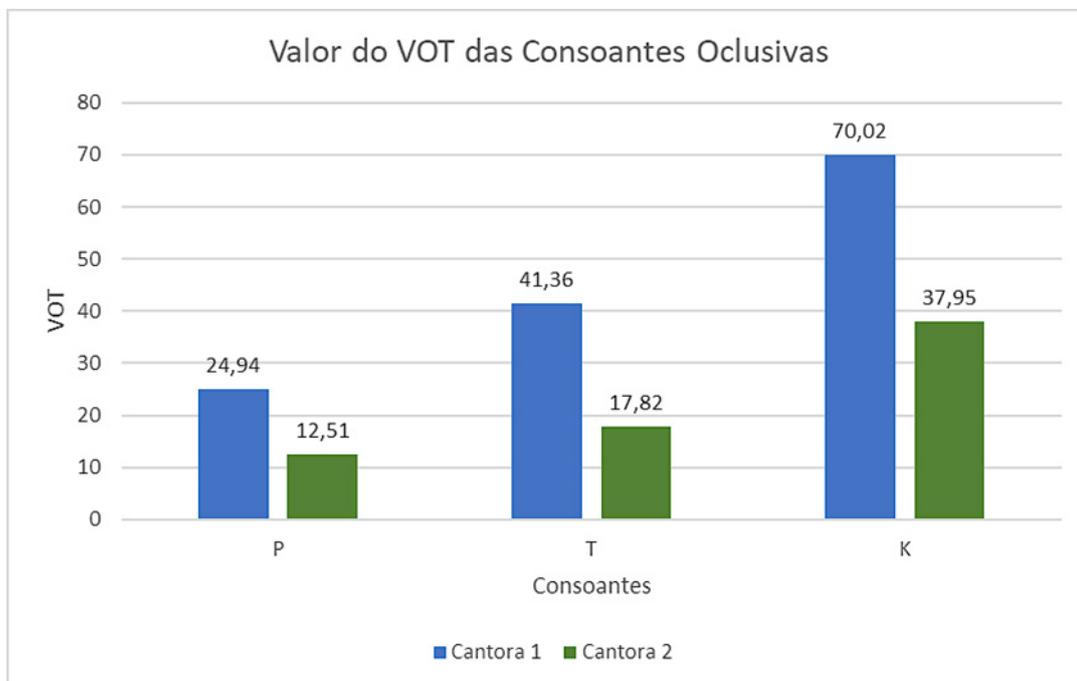
Fonte: Gusmão (2024).

Descrição da imagem: a tabela apresentada mostra no eixo vertical as três consoantes oclusivas estudadas /p/, /t/ e /k/ e, no eixo horizontal, as médias em milissegundos do *VOT*, 19.9, 21.5 e 50.1, as sílabas melódicas analisadas "por", "quis" e "tor" e o tempo de duração da sílaba 483.55, 539.60 e 553.05, respectivamente.

O resultado apresentado na Tabela 3 indica que as médias de *VOT* das consoantes /p/ e /t/ não diferem estatisticamente entre si, enquanto a média do *VOT* da consoante /k/ difere e foi maior que as das consoantes /p/ e /t/ (p-valor < 0,05). Ou seja, pode-se considerar que, em média, as cantoras deste estudo apresentaram um tempo de duração maior para a consoante /k/ quando comparado com as consoantes /p/ e /t/. A análise dos resultados do teste de *Tukey* para o tempo de duração das sílabas melódicas indicou que o tempo de duração da sílaba "por" é significativamente diferente (p-valor < 0,05) e menor que a duração das sílabas "tor" e "quis" (Tabela 3). Pode-se observar, também, que o tempo de duração dessas últimas sílabas melódicas não foi significativamente diferente, uma vez que médias seguidas de mesma letra não diferem entre si.

Essa análise nos mostrou que o ambiente onde a *performance* acontece não influenciou estatisticamente na duração do tempo de fonação das consoantes surdas /p/, /t/ e /k/ e nem de suas respectivas sílabas "por", "tor" e "quis", concluindo, portanto, nosso primeiro objetivo. Em relação aos resultados do experimento 2, identificou-se que a duração das consoantes /p/, /t/ e /k/ difere entre as duas cantoras. O Gráfico 1, a seguir, demonstra a comparação dos valores descritos em milissegundos (ms).

**Gráfico 1 – Comparação dos valores de VOT das consoantes oclusivas surdas /p/, /t/ e /k/ da cantora 1, representado pela cor azul e, da cantora 2, representado pela cor verde**



Fonte: Gusmão (2024).

Descrição da imagem: O gráfico de colunas acima apresenta os valores do VOT em milissegundos das consoantes oclusivas /p/, /t/ e /k/ de duas cantoras. A cantora 1, representada por colunas em azul, e, a cantora 2, representada por colunas verdes. As consoantes foram organizadas na ordem P, T e K no eixo horizontal. Para a cantora 1, os valores do VOT medidos foram 24,94; 41,36 e 70,02, respectivamente. Enquanto que para a cantora 2 os valores registados foram 12,51; 17,82 e 37,95, respectivamente.

Diante dos resultados encontrados, percebe-se diferença entre os valores de VOT entre as cantoras 1 e 2, visto que a cantora 2, considerada a mais inteligível, apresentou o valor do VOT menor do que a cantora 1.

## 4. Discussão

Na primeira análise do Voice Onset Time (VOT), que comparou a duração do VOT e das sílabas melódicas com ambientes distintos de performance, podemos verificar que não houve diferenças significativas quando comparados entre si. Ao comparar os valores de VOT, sem separá-los pelas performances, foi possível perceber que a duração da consoante /k/ é maior do que as demais consoantes analisadas /p/ e /t/. Esse resultado concorda

com os achados de Klein (1999) e Schwartzaupt e Alves (2014), que mostraram que no português brasileiro falado a consoante surda /k/ é mais longa do que as demais oclusivas /p/ e /t/. Mesmo sendo analisados em performance de fala nos estudos anteriores e em performance de canto como este, os dados foram próximos, mostrando que independentemente do comando, fala ou canto, no português do Brasil, a consoante /k/ é mais longa. Ao comparar os nossos resultados com o estudo de McCrea e Morris (2005) verificamos que a consoante velar /k/ também apresentou um VOT maior, tanto na fala quanto no canto norte-americano.

Isso demonstra que a forma como a cantora articula é pessoal e sofre influência de sua técnica vocal individual, e que, se a articulação for ruim ou imprecisa, esse fator seguirá para qualquer ambiente de performance em que a cantora se apresente. Em relação aos valores das sílabas melódicas, a sílaba “por” foi estatisticamente mais curta do que as demais sílabas, “tor” e “quis”. Acreditamos que isso tenha ocorrido pela agilidade em produzir o som bilabial de forma precisa e eficaz, visto que esta precedia de uma respiração.

Em relação ao experimento 2, observou-se que no gráfico 1 o valor de VOT da cantora 1 foi praticamente o dobro do valor de VOT da cantora 2. Conforme Cho e Ladefoged (1999 apud Lara; Battisti, 2014), quanto mais rápido o movimento articulatório, menor o valor de VOT. Além disso, a velocidade do movimento articulatório, os aspectos aerodinâmicos e as diferenças na massa dos articuladores podem influenciar as medidas de VOT (Lara; Battisti, 2014). Dessarte, acreditamos que talvez isso se deva a um ponto articulatório mais preciso e uma maior tonicidade dos músculos articulatorios, visto que tanto para as consoantes /k/ e /t/ há “movimentos de língua, a primeira envolvendo o dorso e a segunda, a ponta da língua” (Medeiros, 2002, p. 75).

Segundo Stevens (2000, apud Medeiros, 2002, p. 52), “a consoante (cuja principal característica é a descontinuidade acústica) é importante para que o ouvinte identifique o texto no canto e deve ser realizada num tempo mínimo suficiente para ser

percebida como tal". Isso demonstra a importância da articulação precisa da consoante para melhor compreensão do texto cantado. Já que acreditamos que a duração menor e precisa das consoantes oclusivas surdas sofra influência dos fatores *anatomofisiológicos* da motricidade orofacial, assim como da técnica vocal.

Músculos orofaciais com pouca tonicidade ou articulações com pouca mobilidade podem afetar a precisão articulatória. Na produção das consoantes, o ponto articulatório e a força muscular dos articuladores vão afetar para que o som saia mais ou menos compreensível. Desvios fonéticos<sup>6</sup> podem apresentar contribuições importantes na inteligibilidade da fala, como, por exemplo, o ceceo anterior que pode ser identificado quando se tem uma incorreta produção dos sons fricativos línguo-alveolares, sobretudo [s] e [z]. Nesse contexto, considera-se que o ceceo anterior ocorre "quando o sopro respiratório é emitido centralmente, mas a língua encontra-se mal posicionada nos planos vertical e ânteroposterior, em geral entre os dentes anteriores" (Leite *et al.*, 2008, p. 31). Por isso, há uma associação muito forte entre ceceo e força de língua. Essa informação serve para nos alertar que alterações dos aspectos da motricidade orofacial podem influenciar nos parâmetros da fala e, conseqüentemente, no canto, podendo, portanto, interferir na inteligibilidade.

No canto, uma posição de língua inapropriada para fonemas específicos ou mesmo tensão em base de língua pode proceder em resultados vocais ineficazes (Miller; Schutte, 2019). Além disso, destacamos a importância de uma produção vocal em um corpo e em uma biomecânica laríngea saudáveis. A interação do corpo, da voz e da mente com o discurso cantado possibilita uma função integral de todo o corpo.

Exercícios cujo objetivo é a eliminação de tensões corporais, cervicais e orofaciais auxiliam em uma produção vocal mais saudável. A prática de vocalizações é necessária para a inteligibilidade das vogais e consoantes no canto, sem perder, contudo, o legato. Santos (2011) propõe utilizar trechos da própria canção como

6 Segundo Leite *et al.* (2008, p. 30), "o desvio fonético tem como principais causas as alterações de estruturas ósseas e/ou musculares, envolvidas na articulação e nas alterações de produção da fala".

recurso para o treino. Exercícios que estimulem a produção de sons curtos (staccato) e longos (legato) são necessários para que o cantor entenda a precisão do movimento dos articuladores durante a produção do texto cantado.

## 5. Considerações Finais

Este estudo buscou elucidar a relação entre a duração das consoantes oclusivas surdas e o ambiente de performance, bem como a influência do tempo de Voice Onset Time (VOT) na inteligibilidade do texto cantado. Os resultados indicaram que o ambiente de performance não influenciou a duração do VOT, sugerindo que, no contexto deste estudo, o ambiente de performance não teve impacto no tempo de VOT das consoantes e das sílabas melódicas. Já em relação ao valor do VOT das consoantes entre as cantoras, foi observado que o VOT foi menor em milissegundos para a cantora considerada mais inteligível, revelando uma estreita relação entre uma menor duração de VOT e a inteligibilidade do texto cantado.

Esse tema oferece possibilidades de aprofundamento em pesquisas futuras, que podem incluir um maior número de participantes e abranger também as consoantes oclusivas sonoras, a fim de estabelecer um parâmetro de comparação que busque analisar a influência das consoantes na compreensão do texto cantado. Concluímos, portanto, que este estudo contribui para a pedagogia vocal, tendo em vista que a fonética acústica é um tema cada vez mais abordado no desenvolvimento técnico-vocal e, sobretudo, para a inteligibilidade do texto cantado.

## Referências

BARROCO, Mário André L. *et al.* Análise temporal das oclusivas orais do português europeu: um estudo de caso de normalidade e perturbação fonológica. **Revista CEFAC**, São Paulo, v. 9, n. 2, p. 154-163, abr./jun. 2007. Disponível em: <https://www.scielo.br/j/rcefac/a/VX6wyRxXszB8BvGBWjhWk3f/?lang=pt>. Acesso em: 22 out. 2024.

CAMARGO, Zuleica A. de. **Análise da qualidade vocal de um grupo de indivíduos disfônicos**: uma abordagem interpretativa e integrada de dados de natureza acústica, perceptiva e eletroglotográfica. 2002. 312 f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – Centro de Pesquisas, Recursos e Informação em Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2002.

CHO, Taehong; LADEFOGED, Peter. Variation and universals in VOT: evidence from 18 languages. **Journal of Phonetics**, Los Angeles, v. 27, n. 2, p. 207-229, 1999. Disponível em: [https://site.hanyang.ac.kr/documents/24916/113960/Cho\\_Ladefoged\\_1999\\_JPhon.pdf](https://site.hanyang.ac.kr/documents/24916/113960/Cho_Ladefoged_1999_JPhon.pdf). Acesso em: 22 out. 2024.

CONSTANTINI, Ana Carolina; ROCHA, Camila B.; MOURÃO, Lucia F. Uso do *Voice Onset Time* associado à diadococinesia oral e laringea na avaliação da disartria na Esclerose lateral amiotrófica. **Distúrbios da Comunicação**, São Paulo, v. 25, n. 2, p. 219-228, ago. 2013. Disponível em: <https://revistas.pucsp.br/index.php/dic/article/view/16475>. Acesso em: 22 out. 2024.

DiCARLO, Nicolle S. L'intelligibilité de la voix chantée. **Medicine des Arts**, Montauban, n. 10, p. 2-15, 1994. Disponível em: <https://www.medecine-des-arts.com/fr/produit/revue-medecine-des-arts-n-10-1.php>. Acesso em: 02 nov. 2024.

DiCARLO, Nicole S. Effect of multifactorial constraints on intelligibility of opera. **Journal of singing**, Jacksonville, v. 63, n. 4, p. 443-455, 2007.

FANT, Gunnar M. **Speech sounds and features**. Cambridge: MIT Press, 1973.

GUSMÃO, Cristina de Souza. **A inteligibilidade do português brasileiro cantado na canção de câmara brasileira**: uma investigação analítica perceptivo-auditiva e acústica. 2024. 174f. Tese (Doutorado em Música) – Escola de Música da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, 2024. Disponível em: <https://repositorio.ufmg.br/handle/1843/67262>. Acesso em: 22 out. 2024.

KAMPFF, Felipe R. **Desenvolvimento de padrões de Voice Onset Time (VOT) em plosivas iniciais do inglês (L2) por aprendizes do sul do Brasil expostos ou não a treinamento perceptual:** uma verificação longitudinal em quatro etapas. 2018. 139 f. Trabalho de Conclusão de Curso (Graduação em Letras) – Faculdade de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

KLEIN, Simone. **Estudo do VOT no Português brasileiro.** 1999. 134 p. Dissertação (Mestrado em Linguística) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 1999.

KUPSKE, Felipe F.; OLIVEIRA, Michele S. de. O desenvolvimento do padrão de voice onset time das oclusivas surdas iniciais do inglês por aprendizes soteropolitanos: efeitos da instrução explícita. **Ilha do Desterro**, Florianópolis, v. 73, n. 3, p. 185-204, set./dez. 2020. Disponível em: <https://www.scielo.br/j/ides/a/vH3dS93Bm8VHFZzHnjNrCLj/>. Acesso em: 22 out. 2024.

LARA, Cláudia; BATTISTI, Elisa. O *Voice Onset Time* das plosivas do Português brasileiro em contato com o Hunsrückisch e seu desvozeamento variável. **Fórum Linguístico**, Florianópolis, v. 11, n. 1, p. 39-50, jan./mar. 2014. Disponível em: <https://periodicos.ufsc.br/index.php/forum/article/view/1984-8412.2014v11n1p39/26875>. Acesso em: 22 out. 2024.

LEITE, Antonelle F. *et al.* Caracterização do ceceo em pacientes de um Centro Clínico de Fonoaudiologia. **Revista da Sociedade Brasileira de Fonoaudiologia**, São Paulo, v. 13, n. 1, p. 30-36, jan./mar. 2008. Disponível em: <https://www.scielo.br/j/rsbf/a/g9fr5dt9TNT56rws4SgT95r/abstract/?lang=pt>. Aesso em: 22 out. 2024.

McCREA, Christopher R.; MORRIS, Richard J. Comparisons of Voice Onset Time for Trained Male Singers and Male Nonsingers During Speaking and Singing. **Journal of Voice**, Amsterdam, v. 19, n. 3, p. 420-430, set. 2005.

McCREA, Christopher R.; MORRIS, Richard J. Voice Onset Time for female trained and untrained singers during speech and singing. **Journal of Communication Disorders**, Tallahassee, v. 40, n. 5, p. 418-431, set./out. 2007.

MEDEIROS, Beatriz R. **Descrição comparativa de aspectos fonéticos-acústicos selecionados da fala e do canto em português brasileiro**. 2002. 166 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2002.

MELO, Paula L; MARTINS, Priscila S. Medidas do VOT de consoantes oclusivas surdas em *onset* de sílaba tônica não final e em *onset* de sílaba pós-tônica final em dados do dialeto curitibano. **Revista Versalete**, Curitiba, v. 2, n. 3, p. 138-154, 2014. Disponível em: <http://www.revistaversalete.ufpr.br/edicoes/vol2-03/138PaulaMelo.pdf>. Acesso em: 22 out. 2024.

MILLER, Richard; SCHUTTE, Harm Kornelis. The Effect of Tongue Position on Spectra in Singing, **Journal of Singing**, Jacksonville, v. 76, n. 1, p. 25-29, 2019.

NELSON, H. D.; TIFFANY W. R. The Intelligibility of Song: Research Results with a New Intelligibility Test. **Journal of singing**, Jacksonville, v. 25, n. 2, p. 22-33, 1968.

NOLAN, Francis; SYKES, Harriet. Vowel and consonant identification at high pitch: The acoustics of soprano unintelligibility. *In*: INTERNATIONAL CONGRESS OF PHONETIC SCIENCE, 18., 2015, Glasgow. **Anais eletrônicos** [...] Glasgow: International Phonetic Association, 2020. Disponível: <https://www.internationalphoneticassociation.org/user/login?destination=node/%20132>. Acesso em: 28 dez. 2023.

SANTOS, Lenine Alves dos. **O Canto Sem Casaca**: propriedades pedagógicas da canção brasileira e seleção de repertório para o ensino de canto no Brasil. 2011. 479 f. Tese (Doutorado em Música) – Instituto de Artes, Universidade Estadual Paulista Júlio de Mesquita Filho, São Paulo, 2011.

SCHWARTZHAUPT, Bruno M; ALVES, Ubiratã Kickhöfel. A influência do contexto fonético-fonológico nos valores de *Voice Onset Time*: verificação de dados de três sistemas linguísticos. **Fórum linguístico**, Florianópolis, v. 11, n. 1, p. 51-68, jan./mar. 2014. Disponível em: <https://periodicos.ufsc.br/index.php/forum/article/view/1984-8412.2014v11n1p51/26876>. Acesso em: 02 nov. 2024.

SWARTZ, Bradford L. Gender difference in voice onset time. **Perceptual and Motor Skills**, Newbury Park, v. 75, n. 3, p. 983-992, 1992.

## Aprovação de comitê de ética em pesquisa

Este estudo foi submetido e aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais sob o número 4.890.170

## Publisher

Universidade Federal de Goiás. Escola de Música e Artes Cênicas. Programa de Pós-graduação em Música. Publicação no Portal de Periódicos UFG.

As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.