

UNIVERSIDADE DE BRASÍLIA
GRADUAÇÃO EM FONOAUDIOLOGIA

THAYSSA MILLÊIDE SANTOS SALES

**PUBLICAÇÕES NACIONAIS SOBRE IDENTIFICAÇÃO DE FALANTES,
DISFARCE NA VOZ E NA FALA E O USO DE FILTRO TELEFÔNICO**

Brasília

2019

THAYSSA MILLÊIDE SANTOS SALES

**PUBLICAÇÕES NACIONAIS SOBRE IDENTIFICAÇÃO DE FALANTES,
DISFARCE NA VOZ E NA FALA E O USO DE FILTRO TELEFÔNICO**

Trabalho de conclusão de curso
apresentado como requisito
parcial para a obtenção de
menção na disciplina TCC 2.

Orientadora: Prof.^a Dr.^a Aveliny
Mantovan Lima

Brasília

2019

RESUMO

A fonética forense é uma ciência que abrange várias atividades relacionadas aos aspectos da fala e dos sons em geral e se dedica à análise perceptiva e acústica de voz e de fala para fins criminalísticos, registradas em algum tipo de mídia, especialmente, para identificação de falantes. Este estudo tem por objetivo fazer um levantamento bibliográfico na literatura nacional sobre a presença de estudos relacionados à identificação de falantes em situação de disfarce e filtro telefônico. Foram utilizados os termos de busca: imitação da voz, imitação da fala, disfarce da voz, disfarce da fala, forense, reconhecimento de falantes, identificação de falantes, análise forense e filtro telefônico, tanto sozinhos ou combinados entre si, no buscador Google Acadêmico e na base de dados Scielo. Foram encontrados, após a aplicação dos critérios de inclusão, 14 artigos, sendo que 9 deles envolvem a temática de análise de produção da fala e apenas 2 estiveram mais relacionados à percepção da fala. Isso demonstra a necessidade de produção de pesquisa nacional sobre o núcleo temático percepção da fala em perícia forense para melhor compreensão sobre o desempenho de testemunhas auriculares.

Palavras-chave: identificação de falantes; disfarce; filtro telefônico, literatura nacional.

ABSTRACT

The forensic phonetics is a science that encompasses various activities related to the aspects of speech and sounds in general and is devoted to the perceptual and acoustic analysis of voice and speech, for criminalistic purposes, recorded in some kind of media, especially for Identification of speakers. This study aims to make a bibliographical survey in the national literature on the presence of studies related to the performance of auricular witnesses in disguise and telephone filter situations. The search terms were used: imitation of voice, imitation of speech, voice disguise, speech disguise, forensic, speaker recognition, identification of speakers, forensic analysis and phone filter, alone or combined with each other, in the Google search engine Academic and in the Scielo database. After the analysis of duplicates, application of inclusion criteria and disagreement analysis were found, only 14 articles and of these 9 involve the production theme and only 1 involves perception. This demonstrates the need to produce national studies on this thematic nucleus, to better understand the performance of ear-witnesses in forensic research.

Keywords: identification of speakers; Cover Telephone filter; National literature.

LISTA DE TABELAS

TABELA 1 – Termos de busca

TABELA 2 – Artigos selecionados

LISTA DE IMAGENS

IMAGEM 1 - fluxograma

SUMÁRIO

INTRODUÇÃO	7
OBJETIVO	9
METODOLOGIA	10
TABELA 1. Termos de busca	10
Imagem 1. Fluxograma	12
TABELA 2. Artigos selecionados	12
DISCUSSÃO	16
<i>Estudos sobre Percepção da Fala e desempenho de testemunhas auriculares</i>	17
<i>Estudos sobre Produção de fala e identificação de falantes</i>	19
CONCLUSÃO	22
REFERÊNCIAS	23

INTRODUÇÃO

A Fonética Forense é uma ciência que abrange várias atividades relacionadas aos aspectos da fala e dos sons em geral, em todos os misteres criminalísticos (BRAID, 2003). Essa ciência se dedica à análise de voz e de fala registradas em algum tipo de mídia, especialmente, para identificação de falantes.

Vieira (2017) define a Fonética Forense como uma subárea da Linguística Forense relacionada à perícia de Identificação de Falantes. E afirma que a Fonética Forense é a ciência que utiliza o arcabouço teórico da Fonética para produção de provas que serão utilizadas para fins judiciais. A Fonética Forense analisa uma série de parâmetros acústicos com o objetivo de obter o máximo de informações acerca de uma voz analisada, podendo compará-la a uma determinada gravação questionada a fim de se identificar o sujeito.

De acordo com Hollien (2002) existem alguns tipos de trabalhos periciais com as amostras de fala e voz, a saber: (a) verificação do falante: o falante é cooperativo, ou seja, quer ser reconhecido, e a verificação se faz entre membros de um corpus pré-existente para confrontação. Um exemplo poderia ser o acesso a uma conta bancária via telefone por comando de voz; (b) reconhecimento do falante: trata-se de um conceito geral, que abrange os outros dois termos – verificação e identificação. Alguns autores consideram que a denominação Reconhecimento de Falantes não traduz o trabalho de identificação, uma vez que o termo reconhecer, para o autor, significa “conhecer de novo o que se tinha conhecido noutra tempo” (GAINOTTI, 2013; SILVA et al., 2012; ARAÚJO e PASQUALI, 2010); (c) identificação do falante: é um trabalho mais difícil, pois geralmente envolve algum tipo de distorção, seja do canal de transmissão, seja do próprio falante (nervosismo, por exemplo), ou ainda, pelo uso de algum tipo de disfarce. Em um contexto de crime, o falante certamente não terá a mesma atitude cooperativa de um cliente bancário com a intenção de ter sua voz identificada.

Tendo em vista que o recorte do presente estudo é a busca por trabalhos sobre a identificação de falantes é importante que se defina o termo e explore como alguns autores tratam as distorções possíveis desse aspecto pericial. Sendo assim, Redher (2015) define a identificação de falantes como os resultados de uma comparação amostral. Essa comparação pode não ser tão simples, caso as amostras envolvam alguma distorção, como, por exemplo, o disfarce e o filtro telefônico.

A situação de disfarce acontece quando o indivíduo não quer ser identificado, ou seja, o disfarce é uma tentativa de manter sua identidade vocal oculta. O discurso disfarçado é tipicamente encontrado, nos casos forenses, em situações em que o criminoso pensa que ele está sendo gravado. (FIGUEIREDO E BRITTO, 1996).

Künzel (2000) especifica o disfarce como alterações de voz, fala e modo de fala, sendo elas: falseto, voz crepitante, sussurro, imitação de sotaque estrangeiro e segurar as narinas enquanto fala. E afirma que estes disfarces são os tipos mais comuns. Além dos disfarces “naturais” como os citados por acima, existem também os dispositivos eletrônicos, utilizados

como ferramenta de disfarce, como observado por Gfroerer (1994) representando 10% dos casos por ele estudados.

Figueiredo e Britto (1996) apresentam em seu estudo outro tipo de disfarce, o do lápis na boca. Esse tipo de disfarce ocorre quando o indivíduo fala após colocar um lápis firmemente entre os dentes anteriores, paralelos aos lábios. Passeti (2012) afirma que a utilização de um instrumento na boca, como um lápis, tende a dificultar estratégias de articulação por parte do falante. O lápis limita a movimentação dos lábios e da língua, além de bloquear a mandíbula. Alterações no comprimento entre as cavidades anterior e posterior também são causadas pela restrição de movimento da laringe.

Quando o indivíduo utiliza esse tipo de disfarce geralmente produz-se um conjunto de pistas que, como um todo, não se encaixa em nenhum quadro dialetal conhecido. Segundo os autores o discurso disfarçado caracteriza-se por uma redução do espaço da vogal. Perceptivelmente, o efeito mais evidente são as mudanças ouvidas para as vogais altas anteriores e altas posteriores. Esta mudança pode dar enganosas sugestões para os indivíduos que fazem a identificação dos falantes. No Brasil, esse problema é de particular importância, uma vez que as diferenças regionais estão relacionadas com a altura da vogal (FIGUEIREDO e BRITTO, 1996).

Ao consultar o dicionário Webster, Eriksson (2010) define a imitação como o ato de copiar ou reproduzir algo. E ao dar continuidade a sua busca diz que a definição mais relevante no contexto atual para o termo “imitação” é: "a assunção do comportamento observado em outros indivíduos". Em seu estudo ainda exemplifica com a imitação do idioma estrangeiro, que pode ser usada por pelo menos dois motivos: esconder a própria voz, ou para fingir que o idioma utilizado é a língua nativa do falante. E afirma que um dialeto imitado não precisa ser particularmente convincente para mascarar eficientemente a variedade nativa do falante. No entanto, considera que, no caso de personificações de interesse forense, o imitador tem como objetivo produzir uma acumulação máxima da cópia do discurso de um indivíduo específico e que os critérios para avaliar o sucesso de tais personificações são de dois tipos: na medida em que o imitador sucede em enganar os ouvintes críticos que estão bem familiarizados com a voz do falante do alvo e quão próximo o imitador pode chegar a determinados alvos acústicos, definidos pela voz alvo. Os alvos acústicos mais frequentemente considerados são o nível de frequência fundamental e taxa de fala.

Quanto à identificação de falantes por parte dos ouvintes, Schlichting e Sullivan (1997) estudaram o efeito de incluir personificações de um falante familiar em uma formação de reconhecimento de voz usando gravações de um conhecido político sueco e personificações de seu discurso por profissionais e imitadores amadores. A voz normal do imitador profissional também foi usada. Com este experimento descobriram que os ouvintes podiam distinguir com segurança a voz real e a voz imitada quando ambos foram apresentados na mesma amostra.

Segundo Eriksson (2014) os disfarces vocais, na medida em que são usados, podem ser um problema sério para a identificação de falantes. Em um dos extremos do espectro de possibilidades, encontramos a manipulação eletrônica ou até mesmo a comunicação via

síntese de fala, que tornaria a identificação de falantes impossível na prática. No mundo do trabalho forense real, no entanto, o emprego dos disfarces vocais tende a ser muito pouco sofisticado.

Eriksson (2014) diz que mesmo que os tipos de disfarce usados, na maioria das vezes, não sejam sofisticados, o disfarce pode, ainda assim, ter um efeito negativo considerável na identificação de falantes. Estudos realizados por Reich e Duke (1979) e Orchard e Yarmey (1995) resultaram em menos identificações corretas quando usado algum tipo de disfarce, ou ainda, quando usado o filtro telefônico.

O filtro telefônico é outro tipo de distorção aplicada à amostra a ser analisada pericialmente. De acordo com Öhman (2010) a transmissão telefônica influencia a qualidade do som negativamente, mas o efeito sobre a identificação do falante foi examinado em apenas alguns estudos. E, diz ainda, que a transmissão do telefone celular afeta a qualidade do som, em parte, da mesma forma que telefones fixos fazem (por exemplo, largura de banda limitada, perdas de transmissão e geralmente o efeito qualidade pobre do microfone), mas não necessariamente de maneira idêntica. A codificação digital é, no entanto, susceptível de ser responsável pela maior parte da diferença de qualidade. Brümmer e Strasheim (2009) relataram não ter conhecimento de nenhum estudo onde o efeito dessas diferenças para reconhecimento ou memória auditiva de voz foi pesquisado, mas demonstrou ter um efeito considerável no reconhecimento automático de falantes.

Em seu estudo Rothman (1977) solicita que os ouvintes ouçam seis amostras de voz pareadas, usando gravações diretas originais ou versões filtradas de passagem de banda que simulam a qualidade do telefone. A tarefa dos ouvintes foi decidir se o falante era o mesmo nas duas amostras. A descoberta foi que as vozes simuladas da qualidade do telefone eram mais difíceis de serem identificadas do que as vozes da gravação original.

De acordo com Künzel (2001) com a ascensão do padrão global para telefonia móvel (GSM), novos tipos de distorções ocorreram. Tais distorções podem degradar seriamente a transmissão da fala e também afetam negativamente a identificação do falante, tanto para análise pericial (forense) quanto para o seu uso comercial.

Com a intenção de melhor compreender o panorama nacional sobre a temática acima exposta, o presente estudo procurou fazer um levantamento dos trabalhos produzidos no país até o momento. Dessa forma, pretende-se também conhecer os resultados já encontrados, bem como apontar caminhos para novos estudos na área.

OBJETIVO

Este estudo tem por objetivo investigar, na literatura brasileira, a presença de pesquisas relacionadas à identificação de falantes em situação de disfarce e filtro telefônico.

METODOLOGIA

Para este estudo foi realizado um levantamento bibliográfico no buscador Google Acadêmico e na base de dados Scielo. A busca foi realizada mediante o uso dos seguintes termos, de acordo com a tabela a seguir.

TABELA 1. Termos de busca

	Termos de busca	Quantidade encontrada	Duplicados
Google Acadêmico	Imitação da voz AND forense	11.300	10
	Disfarce da voz AND forense	2.340	10
	Imitação da fala AND forense	11.100	4
	Disfarce da fala AND forense	3.130	4
	Reconhecimento da voz AND identificação de falantes	27.900	8
	Análise forense AND identificação de falantes	6.090	6
	Imitação da voz AND filtro telefônico	4.880	0
	Disfarce da voz AND filtro telefônico	953	0
Scielo	Termos de busca	Quantidade encontrada	Duplicados
	Imitação AND forense	0	0
	Disfarce AND forense	1	1
	Fala AND forense	2	1

Disfarce da fala AND forense	0	0
Reconhecimento AND forense	2	2
Voz AND forense	1	1
Filtro telefônico AND forense	0	0
Disfarce da voz AND filtro telefônico	0	0

Tabela 1. Resultado da primeira busca especificando a quantidade de material encontrado e duplicado.

A tabela 1 ilustra o resultado do primeiro levantamento de material de acordo com a combinação dos termos escolhidos. A busca foi realizada a partir da leitura das 5 primeiras páginas de resultado do buscador e da base de dados. Em seguida, realizou-se a filtragem de textos duplicados nessas mesmas 5 primeiras páginas.

Logo após, os textos restantes foram analisados através dos critérios de inclusão, a saber:

- a) artigo original de pesquisa;
- b) artigo de revisão de literatura;
- c) sem limite de ano publicação;
- d) língua de publicação português brasileiro ou português europeu; e
- e) a voz humana / fala fosse o material utilizado para análise.

Foi realizada a análise do título e do resumo dos artigos encontrados. A aplicação dos critérios se deu de maneira *ad hoc* entre duas pesquisadoras. Uma terceira pessoa foi consultada para os casos de discordâncias.

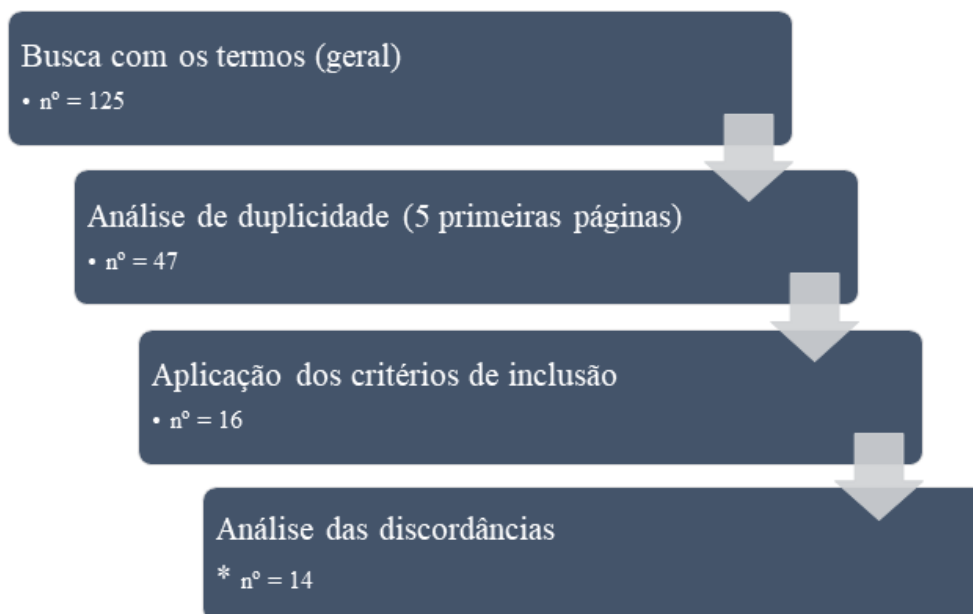


Imagem 1. fluxograma do processo de seleção

Após a seleção restaram 14 artigos originais de pesquisa (12 encontrados no Google Acadêmico e 2 encontrados no Scielo) e deste material realizou-se a análise crítica, em que foram avaliados de acordo com os seguintes assuntos, a saber: (A) perfil dos falantes: sexo, idade, escolaridade; (B) corpus de análise: fala espontânea, fala de laboratório, repetição; condições ou situações da amostra de fala/corpus; (C) uso de disfarce utilizado na voz: imitação, filtro ou outro; (D) se o estudo tratava-se de produção ou de percepção da fala; (E) análise de parâmetros acústicos; e, por fim, (F) se o estudo é referente à amostra ou ao perito. A tabela 2 descreve os estudos selecionados.

TABELA 2. Artigos selecionados

ARTIGOS SELECIONADOS				
Autores e Ano	Temática	Objetivos	Resultados	Conclusão
Passetti e Barbosa (2018)	<ul style="list-style-type: none"> Fala Imitação Produção 	Avaliar quais parâmetros acústicos são modificados ou não na fala com disfarce.	Os locutores conseguiram desviar da própria fala por meio de alterações na entonação, no esforço vocal, em alongamentos vocálicos e por alterações no tempo global do discurso.	As modificações estavam associadas ao desejo de alcançar o registro vocal do locutor-alvo, no caso da imitação, ou às restrições impostas pelo uso do lápis na boca.
Martins et al (2014)	<ul style="list-style-type: none"> Fala Produção Pré-vozeamento 	Estudar o pré-vozeamento em fonética forense.	Os resultados do corpus analisado demonstram que não	O pré-vozeamento apresenta valores típicos muito

			são só os valores médios de PreVoz de cada falante que podem ser distintivos; também a variabilidade desses valores – entre consoantes ou mesmo entre valores médios gerais – se revela potencialmente discriminadora de cada falante.	reduzidos, que vão dos 0 aos 0,017s; o pré-vozeamento varia de consoante para consoante mas cada falante apresenta valores médios característicos e a qualidade da vogal não determina a duração do seu pré-vozeamento;
Vieira (2017)	<ul style="list-style-type: none"> • Fonética e Sociolinguística. • Variação da fala. • Perito 	Demonstrar as relações entre a perícia de Identificação de Falantes e os estudos sociofonéticos, que envolvem métodos e fundamentos da Fonética e da Sociolinguística.	A sociolinguística, traz para os estudos em Identificação de Falantes uma gama de conhecimentos quanto à origem, nível de escolaridade e faixa etária, além de amparar os estudos sobre variação da fala, tão necessários para a realização desse trabalho.	Demonstrou a necessidade de que o perito em Identificação de Falantes tenha conhecimentos sólidos em Linguística para a correta descrição com fundamentação sociofonética das amostras de fala utilizadas em suas análises periciais.
Kremer e Gomes (2014)	<ul style="list-style-type: none"> • Fala • Filtro telefônico • Disfarce • Percepção 	Análise da qualidade da voz comparando a voz normal com a disfarçada, com foco na frequência fundamental.	Na análise perceptiva grande parte dos disfarces se apresentou eficiente - das 30 vozes analisadas, quatorze não foram reconhecidas no disfarce.	O experimento mostrou que, quando se realiza uma maior mudança de $\square 0$, combinada com algum outro tipo de ajuste, obtém-se uma maior eficiência no disfarce.
Gomes; Carneiro e Dresch (2016)	<ul style="list-style-type: none"> • Comparação de falantes • Disfarce • Percepção • Produção 	Introduzir um grupo de alunos de graduação em Letras nas atividades de análise perceptiva e acústica em contexto forense. Comparar de vozes masculinas e femininas.	Informantes do sexo feminino apresentam valores de frequência de F0, F1 e F2 mais altas que os do sexo masculino; as mulheres apresentam espaço vocálico maior que os homens.	Os ajustes mais frequentes para o disfarce de voz foram: a mandíbula aberta e a protrusão labial, como ajustes supralaríngeos, o falsete e a voz áspera, como ajustes de fonação.
Machado e Barbosa (2014)	<ul style="list-style-type: none"> • Fala gravada • Filtro telefônico 	Investigar a eficácia de um conjunto de medidas acústicas no que concerne ao reconhecimento da fala de um indivíduo em um grupo de dez	Os parâmetros acústicos que mais se mostraram robustos em relação à mudança de canal de transmissão foram: média da duração das vogais, taxa de elocução, ΔC (o desvio padrão da	Nenhum dos parâmetros acústicos estudados foi definidor para a identificação precisa do falante. Porém, foi demonstrado que

		falantes do português brasileiro.	duração de intervalos consonânticos) e taxa de movimento de segundo formante. A frequência fundamental também obteve um resultado esperado ao ser afetada pelo telefone celular.	os parâmetros acústicos que mais se caracterizam como robustos pela literatura internacional para a identificação interfalante apresentaram valor significativo para o trabalho. São eles: média da duração das vogais, taxa de elocução, ΔC e taxa de movimento de segundo formante.
Figueiredo (1993)	<ul style="list-style-type: none"> • Fala • Produção • Percepção 	Abordar alguns aspectos relacionados à performance do ELT (espectro de longo termo) como indicador de identidade do falante ainda não examinado em trabalhos anteriores.	A informação contida no ELT não se distribui homogeneamente ao longo de todo espectro. Transformações que privilegiassem as regiões informativas, enfatizando, por exemplo, pontos de inflexão mais acentuada (picos e vales) poderiam criar ELTs ajustados mais eficientes para a identificação de falantes.	O ELT pode ser bastante sensível a certas características do meio e do canal de transmissão. Algumas condições do falante também podem alterar consideravelmente a configuração do ELT, tais como disfarce, presença de stress psicológico, rouquidão etc
Silva (2016)	<ul style="list-style-type: none"> • Fala • Produção 	Realizar a comparação de dois métodos de identificação de locutores, por meio de seus registros de áudio (identificação e comparação de falantes.)	O processo de identificação dos locutores nos 28 arquivos foi efetiva em todos os ensaios, para o algoritmo-01. Para o algoritmo-02 o mesmo processo foi realizado, mas com o recorte por classe acústica, da mesma forma cada locutor foi identificado corretamente, a acurácia também foi computada como a razão entre o número de quadros em que o locutor foi identificado corretamente dentro do número total de quadros (frames).	Os resultados encontrados, do ponto de vista de comparação de locutores, podem ser considerados qualitativamente como promissores. Sobre o custo computacional, o algoritmo-02 é mais parcimonioso, pois o modelo de locutor do algoritmo-02 é mais complexo.
Barbosa; Borges e	<ul style="list-style-type: none"> • Identificação de falantes • Arquivologia 	Demonstrar como o conhecimento arquivístico pode	Proposta de um modelo de descrição dos padrões fonéticos,	A prática da interdisciplinaridade de entre as duas

Braid (2017)		ser utilizado para descrever os padrões fonéticos fornecidos pelo foneticista, auxiliando na construção de sistemas de identificação de falantes.	organizado em forma de matriz, com informações relevantes para o foneticista, para ser utilizado em sistemas de identificação de falantes.	áreas, Arquivologia e Fonética, pode contribuir de maneira importante nos resultados de organização de dados e informações de natureza fonética (padrões fonéticos da fala de indivíduos diretamente extraídos por foneticistas), o que demonstra a ampla abrangência do conhecimento arquivísticos na resolução de problemas hodiernos.
Silva, G.A (2016)	<ul style="list-style-type: none"> • Fala espontânea • Produção • Variação na fala 	Proposta de uma base forense de dados de fala para o estado de São Paulo.	O projeto promoverá o robustecimento da prova em crimes nos quais são utilizadas conversações telefônica – tais como o tráfico de drogas, extorsões, ameaças – e em evidências coletadas por meio de gravações ambientais	A construção da base de dados proposta demandaria um esforço multi-institucional e multidisciplinar.
Arantes (2014)	<ul style="list-style-type: none"> • Análise acústica • Fala 	Investigar a média, a mediana e o valor de base cumulativos para estimar quanto tempo leva para variabilidade atingir estabilidade.	1-A variância após o ponto de estabilização foi reduzida por um fator de 40 vezes no caso da média e mediana e 120 vezes no caso do valor de base. 2-A diferença mediana no valor estimado de F0 no ponto de estabilização nos dois textos é de 2%. 3-O deslocamento mediano do ponto de estabilização é 0,4s no caso da falante feminina e 5,6s no caso do falante masculino.	As diferenças nas estimativas de F0 estão em torno de 5% para a falante feminina e 12% para o falante masculino. Os resultados sugerem que as estimativas de F0 atingem a estabilidade mais cedo do que sugere a literatura.
Azzariti; Deusdará e Rocha (2014)	<ul style="list-style-type: none"> • Atuação do perito no RJ quanto à identificação de falantes. 	Análise de questionamentos dirigidos aos peritos.	Entoações podem marcar diferentes situações de fala: uma conversa dita normal, uma fala submetida à pressão ou a forças de coação, uma fala	O fato de muitos dos quesitos poderem ser respondidos com certo grau de segurança, algumas das

			induzida, de algum modo, por terceiros.	respostas à quesitação se limitam a indicar tendências, probabilidades, sem que se chegue a qualquer termo definitivo acerca da matéria em debate.
Gonçalves (2017)	<ul style="list-style-type: none"> Fala espontânea. 	Verificar a taxa de elocução (TE) e de articulação (TA), bem como a adequação do uso dessas medidas temporais na perícia de Comparação de Locutores.	Ao considerar-se, ainda, que a TA mostrou-se a taxa intra-individualmente mais estável, tem-se que ela é, segundo os dados de amostras forenses analisados, o tipo de taxa mais indicado para utilização na Comparação de Locutores.	Obteve-se coeficiente de correlação intraclasses satisfatório associado à TA, diferença significativa quanto à ciência da gravação na TE e correlação significativa entre <i>gap</i> temporal e TA e entre tamanho do intervalo de fala e TE.
Rehder et al (2014)	<ul style="list-style-type: none"> Transcrição e textualização de áudios de interceptação telefônica. 	Investigar coincidências e divergências entre transcrição e textualização de áudios, a fim de verificar se há evidências estatísticas que possam servir de subsídio quanto ao melhor procedimento a ser aplicado.	O número médio de palavras utilizadas na transcrição dos arquivos de áudio foi 27% maior que o número de palavras utilizadas na textualização, $p=0,52$.	As semelhanças entre os processos de transcrição e textualização foram com relação ao número mediano de palavras e a manutenção do número mediano de focos de conteúdo. Considerando os dados obtidos neste estudo, o processo de textualização mostrou ser o mais indicado na degravação de áudios.

DISCUSSÃO

Conforme observado na tabela 2, os estudos encontrados e selecionados para esse levantamento seguem dois grandes núcleos temáticos: estudos que envolvem os peritos (VIEIRA, 2017; AZZARITTI, DEUSDARÁ E ROCHA, 2014) e estudos que envolvem análise de amostra de fala e voz (MARTINS ET AL., 2014; MACHADO E BARBOSA,

2014; FIGUEIREDO, 1993; BARBOSA, BORGES E BRAID, 2017; ARANTES, 2014; GONÇALVES, 2017; REHDER ET AL., 2014). Dentre os estudos que investigaram amostras de fala, observam-se outros dois grandes eixos temáticos: percepção de fala (KREMER E GOMES, 2014; GOMES, CARNEIRO E DRESCH, 2016) e produção da fala (PASSETTI E BARBOSA, 2018; GOMES, CARNEIRO E DRESCH, 2016); MACHADO E BARBOSA, 2014; SILVA, 2016a); SILVA, 2016b).

Sobre os dois grandes núcleos temáticos, pode-se observar que os assuntos abordados estão em sua maioria voltados para análise acústica das amostras de fala, por exemplo, a análise de variação de f_0 vinculada ao uso de disfarce (KREMER e GOMES, 2014) ou sua diferenciação entre o sexo masculino e feminino associada a ajustes vocais, tal como em Gomes et al. (2016). E mesmo aqueles que possuem por temática a percepção auditiva, foi observado que os trabalhos não investigam o desempenho dos ouvintes na identificação de falantes. Além desses temas surgiu um estudo que tinha como proposta a criação de uma base de dados de amostras de fala para o Estado de São Paulo (SILVA, 2016b), com o objetivo de auxiliar nas investigações de crimes que utilizam conversações telefônicas e gravações ambientais.

Nos estudos que envolvem o papel do perito, muito foi argumentado sobre sua importância no cenário brasileiro, desde o histórico de atuação, que teve início em 1993. Vieira (2017) expôs a necessidade de que o perito considere em seu trabalho a variação sociofonética na identificação de falantes. Além disso, concluíram que é preciso que se tenha nas entidades públicas, sejam institutos de criminalística ou secretarias de segurança, peritos-linguistas, que, de acordo com os autores, são os profissionais capacitados e chamados a contribuir com a resolução de um litígio, conforme definido por Azzaritti et al. (2014).

Com relação aos estudos sobre amostras de fala, observou-se que os assuntos de investigação mais comuns foram os referentes à análise acústica de parâmetros como f_0 (MACHADO E BARBOSA, 2014; ARANTES, 2014), taxa de elocução (GONÇALVES, 2017), espectro de longo termo (FIGUEIREDO, 1993) e pré-vozeamento (MARTINS ET AL, 2014). Além disso, foi debatida a necessidade de criação de um método arquivístico para melhor organização dos padrões de fala extraídos de indivíduos por foneticistas com a finalidade de facilitar o acesso a essas informações no futuro, como exposto por Barbosa et al. (2017). Nesse núcleo também foram estudadas as semelhanças e diferenças entre os métodos de transcrição e textualização de áudios (REHDER ET AL, 2014). Ambos os métodos tiveram como resultado a igualdade no número final de palavras, no entanto, considerou-se o método de textualização mais adequado devido ao menor tempo gasto em sua realização. Segundo Rehder et al. (2014), a média do tempo em minutos necessários para a realização da transcrição foi o dobro do tempo necessário para a textualização.

Uma discussão mais aprofundada sobre os dois sub-eixos da temática produção e percepção da fala em relação à identificação do sujeito, disfarce de fala e voz e uso do filtro telefônico, foco de interesse do presente trabalho, será realizada a seguir.

De acordo com a análise da tabela 2, é possível confirmar a escassez de estudos sobre percepção da fala em análise forense, especialmente os voltados para o desempenho de testemunhas auriculares em situação de disfarce de fala e voz, bem como sobre o uso de filtro telefônico. Foram encontrados apenas dois estudos que se aproximaram dessa temática.

No primeiro, Kremer e Gomes (2014) realizaram experimentos que investigaram a diferença entre fala normal e disfarçada de vozes femininas, através de análise de produção da fala e também de análise perceptiva da fala. Foram analisadas amostras de fala de 50 participantes. Os participantes leram um texto de 75 palavras, simulando uma ligação telefônica. A leitura foi feita, primeiramente, com a voz normal e, em seguida, com voz disfarçada. O ajuste vocal utilizado para o disfarce foi de escolha livre de cada participante. Os autores compararam em ambas as condições (normal e disfarce) os valores de f_0 em configurações variadas do trato vocal (por exemplo, mandíbula aberta, protrusão labial e língua recuada), e também compararam a qualidade vocal por meio da classificação dos modos de fonação (por exemplo, falseto, voz áspera e sussurro).

Nesse mesmo estudo (KREMER e GOMES, 2014), em relação à análise perceptiva, três avaliadores deveriam reconhecer ou não o disfarce nas vozes apresentadas, em duas sessões diferentes. Primeiro, a avaliação foi feita individualmente, e depois os avaliadores chegaram a um acordo quanto ao nível atribuído: se o disfarce foi eficiente ou não foi eficiente. Os avaliadores concluíram que grande parte dos disfarces se apresentou eficiente - das 30 vozes analisadas, 14 não foram reconhecidas no disfarce. Em um segundo momento, analisaram as estratégias de disfarces escolhidas por cada uma das participantes. O resultado dos ajustes envolveram, principalmente, a mudança da frequência fundamental “ f_0 ” em combinação com outro ajuste vocal que poderia envolver a utilização de algum tipo de fonação diferente do modal.

O segundo estudo (GOMES ET AL., 2016) também combinou experimentos de produção e percepção de fala para investigar a comparação de vozes femininas e masculinas em situação de disfarce vocal. Foi criado um texto de 75 palavras para simular uma situação de sequestro. Cinquenta participantes, divididos em grupos de dez de acordo com gênero e faixa etária, foram gravados lendo o texto em voz normal e, em seguida, disfarçando a sua voz. Para a análise acústica foi realizada a extração dos valores de frequência fundamental (f_0), primeiro (F1) e segundo (F2) formantes. Obtiveram como resultado que os informantes do sexo feminino apresentam valores de frequência de f_0 , F1 e F2 mais altas que os do sexo masculino e que as mulheres apresentam espaço vocálico maior que os homens. Além disso, os ajustes comumente utilizados pelos falantes para o disfarce vocal foram a mandíbula aberta e a protrusão labial, e os ajustes de fonação escolhidos foram o falseto e a voz áspera. Os autores destacaram o fato de os participantes do estudo (alunos de graduação em letras) terem sido capazes de atuar como juízes para a realização da análise perceptiva da qualidade vocal normal e disfarçada, identificando os ajustes realizados pelos falantes. Estiveram aptos também a analisar os parâmetros extraídos, estabelecendo relações entre a avaliação perceptiva e os achados acústicos, relacionando-os.

Cabe ressaltar que em ambos os estudos explorados nessa seção (KREMER E GOMES, 2014; GOMES ET AL., 2016) a tarefa dos participantes no experimento de percepção da fala foi a identificação da distorção (ou ajuste) e não do falante. É importante lembrar que a carência de estudos sobre percepção da fala em pesquisa forense pode impactar na produção de evidências sobre o desempenho de testemunhas auriculares. Não foram encontrados estudos sobre percepção em situação de uso de filtro telefônico ou utilizando mídia audiogravada em aparelhos telefônicos fixos e móveis (celulares).

Estudos sobre Produção de fala e identificação de falantes

No estudo de Passetti e Barbosa (2018) avaliou-se quais parâmetros acústicos foram modificados e quais permaneceram inalterados na fala de indivíduos durante a utilização de dois disfarces vocais: sendo a utilização de um lápis posicionado atrás dos dentes incisivos e outro buscando imitar a fala do apresentador de televisão Sílvio Santos. Participaram deste estudo de três locutores do sexo masculino que leram a transcrição de um discurso deste apresentador em três situações: (1) utilizando seus padrões de fala habituais; (2) imitando o apresentador Sílvio Santos; e (3) utilizando um lápis na boca. Quanto aos parâmetros foram analisadas as frequências dos dois primeiros formantes; frequência fundamental, por meio dos descritores estatísticos mediana e valor de base; ênfase espectral; duração das vogais orais do português brasileiro e das pausas silenciosas; e a duração entre dois picos consecutivos da frequência fundamental.

Os resultados mostraram que os locutores conseguiram desviar da própria fala por meio de alterações na entonação, no esforço vocal, em alongamentos vocálicos e por alterações no tempo global do discurso, além disso, as alterações nas frequências formânticas levaram a reconfiguração do espaço vocálico. E por fim concluíram que as modificações encontradas na análise acústica das vozes estavam associadas ao desejo de alcançar o registro vocal do locutor-alvo, no caso da imitação, ou às restrições impostas pelo uso do lápis na boca.

No que diz respeito à identificação de falante, há um dado importante neste estudo quanto a duração de interpicos de f_0 , que não apresentou mudanças significativas para nenhum dos disfarces analisados e que a preservação da hipótese nula se mostra importante para as tarefas forenses, pois permite reconhecer o que permanece, ou seja, o que não modifica na fala de sujeitos mesmo em situações de disfarces vocais (PASSETTI e BARBOSA, 2018).

Em seu estudo Machado e Barbosa (2014) coletaram gravações de seis participantes do estado de São Paulo (três da capital, um de Jundiaí, um de Campinas e um de Cordeirópolis); dois sujeitos da Bahia, ambos de Salvador; um sujeito de Santarém no estado do Pará; e, por fim, um de Pelotas no Rio Grande do Sul. Os sujeitos tinham uma faixa etária de 18 a 28 anos, com nível de educação mínimo de ensino superior incompleto (completando a Graduação) e moraram a maior parte da vida (mais do que a metade) em suas respectivas cidades natais.

Para realização dessa tarefa, os autores analisaram os seguintes parâmetros acústicos, de todas as vogais do português brasileiro, de cada falante: frequência dos dois primeiros formantes, frequência fundamental média, taxa de elocução, frequência baseline, ênfase espectral, a dinamicidade dos formantes e o desvio padrão de duração de intervalos consonânticos.

As amostras de todos os indivíduos foram gravadas em dois canais de gravação, gravação direta e gravação por telefone celular; essa última simula a dificuldade encontrada pelos peritos ao analisar gravações de baixa qualidade, tal como pode ser observado, por exemplo, em uma interceptação telefônica, cujo áudio apresenta ruído e deterioração. Além disso, o indivíduo escolhido, ao qual nos referimos como “criminoso”, teve também sua fala gravada em ambiente acusticamente tratado para uma análise comparativa mais robusta.

A gravação em estúdio da qual o indivíduo participou foi realizada em formato de entrevista conduzida pelos pesquisadores com o objetivo de inserir os mesmos assuntos discutidos na primeira gravação e de deixar o entrevistado o mais à vontade possível para que sua fala fosse fluente e espontânea. Buscando atingir um grau mais próximo de fala espontânea, foi feita uma gravação de cada locutor simulando uma conversa corriqueira, assuntos do dia-a-dia.

Foi observado neste estudo que os parâmetros acústicos que mais se mostraram robustos em relação à mudança de canal de transmissão foram: média da duração das vogais, taxa de elocução, ΔC e taxa de movimento de segundo formante. A taxa de elocução foi um dos parâmetros que se manteve robusto na mudança de canal de transmissão, não tendo variação para o canal telefônico em relação à gravação direta. A frequência fundamental teve um aumento de seu valor de 4% em relação à gravação direta, valor estatisticamente pequeno para a variação, já a frequência do segundo formante teve uma diminuição de 7% em seu valor.

Os autores concluíram que nenhum dos parâmetros acústicos foi definidor para a identificação precisa do “criminoso”, objetivo principal do experimento. Porém, foi possível demonstrar que os parâmetros acústicos que mais se caracterizam como robustos pela literatura internacional para a identificação interfalante, também apresentaram valor significativo para o trabalho, já que ΔC (o desvio padrão da duração de intervalos consonânticos) e a dinamicidade dos formantes foram essenciais para mostrar traços idiossincráticos dos indivíduos e que a taxa de movimento do segundo formante foi o parâmetro acústico que apresentou melhores resultados na pesquisa.

No estudo de Silva (2016a) que tinha por objetivo realizar a comparação de dois métodos de identificação de locutores, por meio de seus registros de áudio, sendo o primeiro método (denominado algoritmo-01), identifica falantes independente de texto, utiliza como características as componentes mel cepstrais e realiza inferência bayesiana modelando os locutores através de modelos de misturas de gaussianas e o segundo método (denominado algoritmo-02) propunha comparar os falantes, através dos seus registros de áudio, também

utilizando como medições as componentes mel cepstrais e inferência bayesiana com de modelos de misturas de gaussianas, entretanto a comparação dos áudio foi realizada em classes de sons acusticamente semelhantes. as amostras de fala foram coletadas em ambiente controlado (com pouca reverberação).

O autor utilizou como protocolo na pesquisa a leitura de 22 (vinte e duas) frases de controle, com objetivo de distribuir as classes de sons anteriormente apresentadas. O procedimento de coleta dos dados buscou catalogar e aleatorizar a idade do falante, sexo, o horário da coleta, condição de saúde, conhecimento prévio de idiomas, entre outros fatores capazes de influenciar a relação competência/desempenho fonológico. O material sonoro dos locutores foi condicionado para realização das análises tendo por base as presentes em Furui (2000), sendo elas: subamostragem para uma frequência de 8kHz, típica de sistemas de comunicação telefônica, e a magnitude foi normalizada para manter o valor RMS do sinal em torno de -1 dB. Além disso, foram calculadas as características do áudio de cada locutor e o agrupamento dos trechos de voz com remoção dos quadros de silêncio e outros artefatos.

Como resultado obteve-se que o processo de identificação dos locutores nos 28 arquivos foi efetiva em todos os ensaios, para o algoritmo-01, cada modelo de locutor λ_i foi comparados com os T quadros das amostras de voz j , A razão entre o número de quadros em que o locutor foi identificado corretamente, em relação ao total de quadros, é a medida de acurácia do algoritmo-01. E que Para o algoritmo-02 o mesmo processo foi realizado, mas com o recorte por classe acústica, da mesma forma cada locutor foi identificado corretamente, a acurácia também foi computada como a razão entre o número de quadros em que o locutor foi identificado corretamente dentro do número total de quadros (frames).

Silva (2016a) concluiu então que os resultados encontrados, do ponto de vista de comparação de locutores, pode ser considerado qualitativamente como promissor. E que apesar das comparações serem realizadas utilizando com uma forma simplificada dos algoritmos de verificação de locutores, o fato das comparações realizadas com agrupamento das classes de sons terem apresentado uma acurácia média superior em aproximadamente 6%, nas mesmas condições de comparação, é no mínimo digna de mais averiguações.

Em seu estudo Silva (2016b) objetivou apresentar, em linhas gerais, o arcabouço bayesiano no qual o resultado é expresso por meio da razão de verossimilhança, doravante LR, fazendo uma revisão acerca de como tal arcabouço vem sendo usado nas abordagens fonético-forense e apresentar, em linhas gerais, uma proposta de criação de uma base de dados de fala para a variedade do Estado de São Paulo.

Aqui o autor considera que O LR é um fator que expressa numericamente o peso da evidência, convertendo uma probabilidade a priori em probabilidade a posteriori. Esse peso é sempre relacionado a duas hipóteses concorrentes. É uma relação que quantifica a hipótese da evidência com respeito a duas proposições alternativas. Dentro do processo legal, a probabilidade a posteriori é, por outro lado, função do juiz. E afirma que a revisão realizada em seu estudo mostra que já é bem consolidado o uso do LR em exames de Fonética Forense,

não importando qual a metodologia utilizada, e que para que os peritos do Brasil possam incorporar esse arcabouço em suas análises, é necessária a construção de uma base de dados de fala.

A metodologia utilizada nessa proposta seria a coleta de amostras de fala de 200 indivíduos do sexo masculino, 100 deveriam representar os falares da capital e da região metropolitana, 50 representar a variedade caipira e 50 deveriam ser provenientes do litoral. E 100 locutores de sexo feminino, a proporção seria 50 falantes da capital e região metropolitana, 25 do interior e 25 do litoral. Sendo que para cada locutor, seria desejável dez minutos de fala líquida em eliciação espontânea, a fim de se obter amostras espontâneas e, ao mesmo tempo, controladas. As gravações deveriam ser feitas em três sessões: uma pela manhã, uma à tarde e outra à noite. Uma vez que a intravariabilidade é um aspecto inerente ao sinal de fala, o aspecto da contemporaneidade das amostras deve ser considerado (SILVA,2016b).

O autor apresentou as vantagens que seriam obtidas com a criação do banco de dados como: maior confiabilidade de uso dos sistemas automáticos e de parâmetros tradicionais, a promoção do robustecimento da prova em crimes nos quais são utilizadas conversações telefônica– tais como o tráfico de drogas, extorsões, ameaças – e em evidências coletadas por meio de gravações ambientais. Este projeto também poderia estimular outras criações de bases de dados de fala, permitindo o compartilhamento entre as instituições, tanto em nível nacional como internacional.

CONCLUSÃO

O presente estudo produziu um levantamento de 14 artigos originais de pesquisa publicados em português brasileiro em revistas nacionais sobre um recorte da fonética forense que envolve a identificação de falantes, o uso de disfarce e de filtro telefônico.

Conclui-se que a maioria dos estudos, mais precisamente 5 trabalhos, são sobre produção da fala e apenas 2 estudos encontrados foram sobre percepção da fala. Isso demonstra a necessidade de os pesquisadores nacionais da área de fonética forense produzirem estudos sobre esse sub-núcleo temático, uma vez que há a necessidade de compreendermos melhor o desempenho de testemunhas auriculares.

REFERÊNCIAS

- ARANTES, Pablo. Estimativas de longo termo da frequência fundamental: implicações para a fonética forense. *ReVEL*, vol. 12, n. 23, 2014.
- AZZARITI, M; DEUSDARÁ, B; ROCHA, D. Linguística Forense: saberes implícitos e imagens discursivas do perito em contexto de quesitação. *ReVEL*, vol. 12, n. 23, 2014.
- BARBOSA, T.S; BORGES, J.; BRAID, A.C.M. O conhecimento arquivístico aplicado na descrição de padrões fonéticos. **Informação & Informação**, v. 22, n. 1, p. 242-262, 2017.
- ERIKSSON, A. The disguised voice: Imitating accents or speech styles and impersonating individuals. *Language and Identities*, Edinburg University Press Series, chapter 8, 2010, p. 86.
- ERIKSSON, A. Tutorial sobre fonética forense. *ReVel*, vol.12, nº23, Universidade Federal de São Carlos, 2014. Tradução Pablo Arantes e Suska Gutzeit.
- FIGUEIREDO, R. M. A eficácia de medidas extraídas do espectro de longo termo para a identificação de falantes. *Cadernos de Estudos Lingüísticos*, v. 25, p. 113-127, 13 out. 2012.
- FIGUEIREDO, R.M; BRITTO, H.S. A report on the acoustic effects of one type of disguise. *Forensic Linguistics*, vol. 3, nº1, Departamento de Medicina Legal and UNICAMP, Campinas, Brazil. Routledge, p. 168-175, 1996.
- GOMES, M. L. DE C.; CARNEIRO, D. DE O.; DRESCH, A. A. G. Análise perceptiva e acústica em fonética forense: uma pesquisa em disfarce de voz. *Domínios de Lingu@gem*, v. 10, n. 2, p. 559-589, 27 jun. 2016.
- GOMES, M.L; CARNEIRO, D.O. A fonética forense no Brasil: cenários e atores. *Language and Law*, vol. 1, nº1, Universidade Tecnológica Federal do Paraná, p. 22-36, 2014.
- GONCALVES, Cintia Schivinski. Taxa de elocução e taxa de articulação em corpus utilizado na perícia de Comparação de Locutores. **Let. Hoje**, Porto Alegre , v. 52, n. 1, p. 15-25, Mar. 2017 .
- KREMER, Robinson Luis; GOMES, Maria Lúcia de Castro. A eficiência do disfarce em vozes femininas: uma análise da frequência fundamental. *ReVEL*, vol. 12, n. 23, 2014.
- KÜNZEL, H.J. Beware of the “telephone effect”: the influence of telephone transmission on the measurement of formant frequencies. *Forensics Linguistics*, vol. 8, nº1, University of Birmingham Press, p. 1350-1771, 2001.
- MACHADO, A. P. & BARBOSA, P. A. - Uso de técnicas acústicas para verificação de locutor. . . *Language and Law / Linguagem e Direito*, Vol. 1(2), 2014, p. 100-113
- MARTINS, Fernando; SIMÕES, Deolinda Reis; BRISSOS, Fernando; RODRIGUES, Celeste. A Fonética Forense na produção de prova do ordenamento jurídico português: o parâmetro do pré-vozeamento. *ReVEL*, vol. 12, n. 23, 2014
- MASTHOFF, H. A report on a voice disguise experiment. *Forensics Linguistics*, vol. 3, nº1, University of Germany. Routledge, p. 160-167, 1996.

NOLAN, F; MACDOUGAL, K; HUDSON, T. Effects of the telephone on perceived voice similarity: implications for voice line-ups. Equinox Publishing, vol. 20, nº2, University of Cambridge, p. 229-246, 2013.

ÖHMAN, L; ERIKSSON, A; GRANHAG, P.A. Mobile phone quality vs. Direct quality: How the presentation format affects earwitness identification accuracy. The European Journal of Psychology Applied to Legal Context, Sweden, vol. 2, nº2, p. 161-182, 2010.

PASSETTI, R.R. *Desviando da própria fala: implicações para a verificação de locutor em situações de disfarces vocais*. 2012. 38f. Monografia. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. Campinas, 2012.

PASSETTI, R. R.; BARBOSA, P. A. Análise fonético-acústica de disfarces vocais com aplicações para a fonética forense. **Cadernos de Estudos Lingüísticos**, v. 60, n. 3, p. 567-583, 22 nov. 2018.

REHDER, M.I; CAZUMBÁ, L.F; CAZUMBÁ, M. Identificação de falantes: Uma introdução à fonoaudiologia forense. 1 ed. Rio de Janeiro. Revinter, 2015.

REHDER, Maria Inês Beltrati Cornacchioni et al . Coincidências e divergências entre transcrição e textualização de áudios. **Rev. CEFAC**, São Paulo , v. 16, n. 6, p. 1919-1927, Dez. 2014 .

REICH, A.R; DUKE, J.E. Effects of selected vocal disguises upon speaker identification by listening. Acoustical Society of America, vol. 66, nº4, 1979.

SILVA, A. P. Avaliação da Identificação de Falantes em Classes de Sons Semelhantes. Research Gate, 2016.

SILVA, G.A. Proposta de Construção de um Banco de Dados de Amostras de Fala para Uso Forense em um Arcabouço Bayesiano. **Revista Brasileira de Criminalística**, v. 5, n. 1, p. 35-45, 2016.

VIEIRA, Renata Christina. Contribuições dos Estudos Sociofonéticos para a Identificação de Falantes. Revista Intercâmbio, Especial Expressividade, v. XXXVI: 86-102, 2017. São Paulo: LAEL/PUCSP. ISSN 2237-759X