Desviando da própria fala: implicações para a verificação de locutor por falantes e não-falantes do português brasileiro

Renata Regina Passetti

Mestranda em Linguística. Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Brasil.
re.passetti@gmail.com

Plínio Almeida Barbosa

Doutor em Signal-Image-Parole/Option Parole. Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Brasil.
pabarbosa.unicampbr@gmail.com

Anders Eriksson

Doutor em Linguística. Departamento de Linguística, Universidade de Estocolmo, Suécia.
anders.eriksson.gu@gmail.com

Resumo: Este trabalho tem como objetivo avaliar a taxa de reconhecimento de locutor entre grupos de falantes e não-falantes do português brasileiro e investigar qual o tipo de informação (acústica e/ou lexical) empregada durante a tarefa de verificação de locutor, além de tecer considerações sobre possíveis pistas acústicas que estariam interferindo na decisão destes falantes. Os resultados obtidos pela análise comparativa da duração de unidades VV e do grupo acentual e, posteriormente, pela análise da mediana de f0, entre pares de amostras de fala (locutor-alvo e distrator falsamente escolhido), apresentam uma possível explicação para a escolha incorreta dos ouvintes.

Palavras-chave: Verificação de locutor; percepção da fala; disfarce vocal

Abstract: This research aims at analyzing some of the perceptual features used by speakers and non-speakers of Brazilian Portuguese for speaker verification in Brazilian Portuguese speech samples. The focus of the work was to evaluate the identification rate and to analyze which type of information (acoustic and/or lexical) was used by listeners during the speaker verification task. The results of a comparative analysis of duration-related acoustic salience and speech rate in VV units/s and further by the analysis of the speakers f0 median could be one explanation for listeners' wrong choices, since their choice could be relying on these rhythmic cues.

Keywords: speaker verification; speech perception; vocal disguise

Introdução

O desvio e manipulação da identidade fonético-fonológica do locutor para a realização de disfarces vocais provocam modificações prosódicas e acústicas nos padrões habituais da fala, com o intuito de ocultar a identidade do locutor em situações de interação com ouvintes leigos. O exame das estratégias perceptuais de ouvintes interessa aos estudos de verificação de locutor na medida em que contribui de maneira significativa para a compreensão dos processos de raciocínio em testemunhas auditivas.

As amostras de fala desempenham um papel fundamental no processo de verificação de locutor por testemunhas auditivas. Eriksson (2005) afirma que os fatores que primeiramente influenciam a precisão na verificação de locutor são a duração da amostra de fala e a qualidade acústica da amostra. Em estudos iniciais de Pollack, Pickett e Sumby (1954), analisou-se o efeito de diversas variáveis no desempenho da verificação de locutor. Os autores observaram que, até aproximadamente 1,2 segundo, a precisão da identificação de falantes aumentava conforme o tamanho da amostra, para palavras monossilábicas. Para as amostras longas, a variação fonética foi assumida como o fator mais importante para a identificação de um locutor em um conjunto fechado. Os autores concluíram que a duração apenas se mostra importante na medida em que admite uma amostragem estatística maior ou menor do repertório de fala do locutor.

Bricker e Pruzansky (1966), ao analisarem os efeitos na verificação de locutor de estímulos com variações fonêmicas e duracionais, descobriram que a taxa de identificação somente aumentava com a duração se os estímulos mais longos também contivessem uma maior variação fonêmica. O aumento na precisão da identificação também estava diretamente relacionado ao número de fonemas na amostra de fala, mesmo quando a duração era controlada.

As escolhas linguísticas que fazemos são elementos centrais na nossa concepção como membros de grupos sociais e também como indivíduos independentes e distintos de todos os outros. A partir dessa visão de singularidade linguística, Watt (2010) propôs uma análise das maneiras de identificação de um indivíduo pelos seus padrões de fala através da consideração de dois tipos diferentes de identificação de locutor: a "técnica" e a feita por "leigos". O autor define esta como uma tarefa realizada com base no cotidiano dos falantes e que não necessita de treinamento linguístico e fonético. Em uma identificação de locutor informal, é plausível supor que os mecanismos cognitivos que permitem o reconhecimento de vozes conhecidas são inconscientemente ativados, mesmo quando o ouvinte é exposto a uma voz desconhecida. Watt ainda argumenta que a quantidade de atenção empregada pelo ouvinte no sinal de fala parece afetar a exatidão de uma subsequente verificação de locutor. Entretanto, algumas limitações podem afetar a verificação de locutor por leigos, como quando características linguísticas e dialetais da amostra de fala são manipuladas pelo uso de disfarces vocais, como o que usaremos neste trabalho.

Ainda no campo de estudos sobre verificação de locutor, a influência de sotaques e línguas estrangeiras tem sido objeto estudos de diversas pesquisas. Schiller e Köster (1996), com o intuito de revisar determinados aspectos de uma investigação experimental sobre a influência do conhecimento da língua materna na prática de verificação de locutor, utilizaram gravações de seis falantes nativos de alemão para testar falantes nativos de inglês americano sem conhecimento de alemão, falantes de inglês americano com algum conhecimento de alemão e falantes nativos de alemão. Os resultados indicaram que a não familiaridade com a língua-alvo afeta a habilidade de reconhecimento de locutor. Sujeitos sem conhecimento de alemão obtiveram, significativamente, mais erros de

identificação e sujeitos com algum conhecimento de alemão obtiveram resultados semelhantes aos dos falantes nativos de alemão. Os autores argumentam que a prática de reconhecimento de locutor parece não envolver apenas características puramente fonéticas, mas também a incorporação de informação linguística.

Köster e Schiller (1997) replicaram este experimento (SCHILLER e KÖSTER, 1996) em ouvintes nativos de espanhol e chinês. Os resultados mostraram que ouvintes espanhóis e chineses com algum conhecimento de alemão obtiveram melhores taxas de reconhecimento de locutor do que ouvintes espanhóis e chineses sem conhecimento da língua-alvo. Porém, quando comparados com falantes nativos de alemão e ouvintes nativos de inglês americano com algum conhecimento de alemão, os ouvintes espanhóis e chineses com algum conhecimento de alemão obtiveram resultados mensuravelmente piores. A partir desses estudos, concluiu-se que, em testes de reconhecimento de locutores, o conhecimento prévio da língua-alvo pelos sujeitos ouvintes aumenta a confiabilidade dos resultados de reconhecimentos.

O desempenho de reconhecimento de locutor em estudos envolvendo língua e sotaques estrangeiros na verificação de locutor parece estar mais propenso a ser afetado pela familiaridade do ouvinte com a língua. O desempenho também se mostra sensível à duração da amostra de fala, como atestado anteriormente por Pollack *et al.* (*id.*) e Bricker e Pruzansky (*id.*). Com base nos estudos desenvolvidos acerca da verificação de locutor envolvendo amostras de falas de línguas e sotaques distintos dos ouvintes, o principal objetivo desta pesquisa foi avaliar a taxa de reconhecimento de locutor em amostras de fala do português brasileiro apresentadas a ouvintes nativos de sueco e a ouvintes nativos de português brasileiro que vivem na Suécia e analisar qual o tipo de informação (acústica e/ou lexical) empregada durante a

tarefa de verificação, além de tecer considerações sobre as possíveis pistas acústicas que estariam interferindo na decisão de falantes e não-falantes do português brasileiro.

1. Metodologia

1.1 Corpus

O *corpus* utilizado neste trabalho foi composto pela gravação de 7 locutores, 3 homens e 4 mulheres. O grupo apresentava idade média de 21 anos, todos eram estudantes de graduação da Universidade Estadual de Campinas, exibiam um sotaque do sudeste brasileiro (interior do estado de São Paulo) e não apresentavam problemas auditivos ou fonoarticulatórios. As gravações consistiam em diferentes sessões de leitura de um discurso do apresentador de televisão Sílvio Santos. Os locutores foram gravados em dois momentos diferentes: (1) utilizado o disfarce vocal "objeto na boca", pelo uso de um lápis posicionado firmemente entre os dentes frontais e (2) com o estilo de fala habitual, com voz e estilo de fala naturais.

As gravações foram realizadas utilizando um microfone Shure, modelo SM58A e uma placa MBox 2, ligada ao programa computacional Pro Tools. Os equipamentos utilizados pertencem ao Grupo de Estudos de Prosódia da Fala. As gravações fazem parte do banco de dados do LIAAC, tendo sido autorizado o uso das amostras para o estudo aqui apresentado, dentro dos termos da resolução 196 do CNS, que rege pesquisas com seres humanos.

As gravações foram utilizadas na elaboração de 6 filas de reconhecimento, três com amostras de fala masculinas e três com amostras de fala femininas. O programa PRAAT (BOERSMA & WEENINK, 2012) foi utilizado para a preparação das filas de reconhecimento. As filas de reconhecimento foram organizadas de acordo com a Fig. 1:

Voz disfarçada + *beep* + Voz sem disfarce 1 + Silêncio + Voz sem disfarce 2 + Silêncio + Voz sem disfarce 3 + Silêncio + Voz sem disfarce 4 + Silêncio + Voz sem disfarce 5 + Silêncio + Voz sem disfarce 6.

Fig.1: organização das amostras de fala nas filas de reconhecimento

A seleção dos distratores para cada amostra da fila de reconhecimento seguiu critérios de acordo com as características dos locutores-alvo (cf. BROEDERS e AMELSVOORT, 1999): sexo, idade (perceptível), características linguísticas (mesmo dialeto), nível educacional perceptível e as mesmas características de pronúncia (perceptíveis) enquanto liam os trechos. As amostras de fala contidas nas filas de reconhecimento seguiram critérios duracionais estabelecidos em estudos anteriores (cf. POLLACK *et al.*, 1954; BROEDERS e AMELSVOORT, *ibid.*). A amostra de referência (voz disfarçada) possuía 40 segundos e cada amostra de fala não-disfarçada continha 6,5 – 8 segundos de discurso interrupto. Um *beep* após a amostra de referência indicava o início das amostras de fala não-disfarçadas. Estas, por sua vez, estavam separadas uma das outras por um período de silêncio de 4 segundos.

1.2 Ouvintes

Os ouvintes que participaram do teste de percepção foram selecionados de dois grupos: (1) ouvintes nativos de sueco, sem nenhum conhecimento em português brasileiro (n = 20) e (2) ouvintes nativos de português brasileiro, que moravam na Suécia há, pelo menos, 1 ano (n = 10). Ambos os grupos não possuíam treinamento fonético e todos participaram do experimento voluntariamente.

O primeiro grupo era formado por 10 mulheres, com idade entre 25 e 36 anos (m_{idade} = 33; sd_{idade} = 4,2) e 10 homens, com idade entre 26 e 38 anos (m_{idade} = 27,7; sd_{idade} = 4,7). Todos eram falantes nativos de sueco e livres

de problemas auditivos. O segundo grupo era formado por 3 mulheres, com idade entre 32 e 55 anos (m_{idade} = 42,6; sd_{idade} = 11,5) e 7 homens, com idade entre 24 e 48 anos (m_{idade} = 32,7; sd_{idade} = 8,1). Todos eram falantes de português brasileiro, habitantes da Suécia e também livres de problemas auditivos. A tabela 1 contém informações sobre os ouvintes incluídos neste estudo:

Grupo	*Número testado*	*Número incluído*	*Média de idade*	*sd idade*	*Faixa etária*	*Homens*	*Mulheres*
Ouvintes nativos de sueco	21	20[1]	30,5	5,2	25 - 38	10	10
Ouvintes nativos de português brasileiro	10	10	35,7	10,5	24 - 55	7	3

Tabela 1: Informações sobre os ouvintes participantes deste estudo

[1] Um dos ouvintes suecos não foi incluído no teste de percepção devido a problemas de compreensão durante o teste.

[2] Veja o texto informativo apresentado aos ouvintes na seção Apêndice.

1.3 Procedimentos experimentais

O teste de percepção foi apresentado por meio de uma apresentação de *slides*, iniciada pela simulação de um crime[2], no qual esses ouvintes haviam sido supostamente vítimas, seguido de instruções sobre a condução das filas de reconhecimento e a tarefa de verificação de locutor. Essas informações eram precedidas pelas seis filas de reconhecimento. Cada fila de reconhecimento era iniciada por uma amostra de fala do português brasileiro com o locutor-alvo utilizando o disfarce vocal do tipo "objeto na boca"; seguida da apresentação de outras seis vozes, aleatoriamente organizadas, compostas pela voz habitual do locutor-alvo juntamente com outras cinco vozes distratoras também no estilo habitual de elocução, conforme apresentado anteriormente na fig.1.

A apresentação do teste de percepção foi elaborada nos moldes propostos por Broeders e Amelsvoort (*ibid.*), e informava que a tarefa dos ouvintes era, primeiramente, ouvir a amostra de fala disfarçada e, em seguida, tentar reconhecer qual dentre as seis amostras de fala sem

disfarce correspondia à voz do locutor-alvo. As gravações utilizadas nas amostras de fala das filas de reconhecimento consistiam em diferentes sessões do mesmo texto enunciado por diferentes locutores e diferentes amostras de fala de um mesmo locutor (denominadas "versões"). Dessa forma, um locutor que havia desempenhado o papel de "locutor-alvo" em uma fila de reconhecimento, poderia ser distrator em outra fila com amostras de fala de locutores do mesmo sexo.

O teste de percepção foi aplicado no estúdio de fonética do Departamento de Filosofia, Linguística e Teoria da Ciência da Universidade de Gotemburgo, Suécia. A duração total do teste era de, aproximadamente, 15 minutos.

2. Resultados e Discussão

A análise do desempenho dos ouvintes no teste de percepção considerou três etapas. Primeiramente, as respostas dos ouvintes foram avaliadas de acordo com 4 diferentes tipos de análises: (1) análise da taxa de reconhecimento do locutor-alvo pela porcentagem de identificações falsas e corretas em cada fila de reconhecimento; (2) análise das falsas identificações por distrator; (3) análise das falsas identificações pela posição da amostra de fala na fila de reconhecimento e (4) análise das falsas identificações pelas amostras de fala de um mesmo distrator (versão). A segunda etapa de análise consistiu em um estudo acerca da duração do grupo acentual e da taxa de elocução em unidades Vogal-Vogal (unidades VV) por segundo entre as amostras de fala do locutor-alvo e do distrator erroneamente escolhido com maior frequência para cada uma das filas de reconhecimento. Esta análise tinha como objetivo investigar se estes parâmetros poderiam ser considerados como pistas acústicas, nas quais os ouvintes estariam apoiando suas falsas escolhas.

Por fim, foi conduzida uma análise da mediana da frequência fundamental (doravante *f0*) entre os pares de locutores (alvo e erroneamente escolhido) para cada fila de reconhecimento, cujo objetivo era aprimorar os resultados estabelecidos para a duração das unidades VV e para a duração do grupo acentual e também auxiliar na compreensão das falsas identificações feitas por ambos os grupos de ouvintes.

Como explicado anteriormente, o material de apresentação tinha, ao todo, 6 filas de reconhecimento, 3 delas com amostras de fala com vozes masculinas e outras 3 com amostras de fala com vozes femininas. Havia 7 distratores diferentes (m = 3 e f = 4). As filas de reconhecimento foram aleatoriamente organizadas e, em uma única fila de reconhecimento, poderia haver mais que uma amostra de fala de um mesmo distrator. Para diferentes amostras de fala de um mesmo locutor (distrator) deu-se o nome de "versão".

As filas de reconhecimento foram organizadas de acordo com o seguinte critério: (a) código do locutor, (b) versão do locutor (distrator), (c) posição. A tabela 2 mostra um exemplo da organização das filas de reconhecimento.

Fila de reconhecimento nº 3							
Amostra de referência	*Posição*	1	2	3	4	5	6
VOZ DISFARÇADA	*Código do locutor*	4	5	5	4	ALVO	6
	Versão do locutor (distrator)	4.1	5.1	5.2	4.2		6.1

Tabela 2. Organização da fila de reconhecimento nº 3

De acordo com a Tabela 2, se algum ouvinte escolhesse erroneamente a segunda posição na folha de resposta, ao invés da quinta posição (locutor-alvo), nesta fila de reconhecimento (número 3), sua escolha seria

anotada como "5.1.2": código do locutor 5, versão do locutor (distrator) 1 e posição 2.

2.1 Análise da taxa de reconhecimento do locutor-alvo pela porcentagem de identificações falsas e corretas em cada fila de reconhecimento

A taxa de reconhecimento do locutor-alvo para cada grupo de ouvintes foi obtida pelo contagem das identificações corretas e falsas do locutor-alvo em cada fila de reconhecimento. A tabela 3 mostra as taxas de reconhecimento do locutor-alvo pela porcentagem de identificações corretas em relação ao número total de ouvintes de cada grupo.

Taxa de reconhecimento do locutor-alvo (%)		
Fila de reconhecimento	*Suecos (n = 20)*	*Brasileiros (n = 10)*
1	15	30
2	70	80
3	40	40
4	30	40
5	15	50
6	45	30

Tabela 3. Porcentagem de identificação do locutor-alvo por fila de reconhecimento em ambos os grupos de ouvintes.

A fila de reconhecimento 2 foi identificada corretamente com maior frequência para ambos os grupos de ouvintes. O melhor desempenho dos ouvintes para esta fila de reconhecimento pode ser explicado pela posição do locutor-alvo. Este se encontrava na primeira posição, logo após a amostra de referência (voz disfarçada), o que pode ter auxiliado na decisão dos ouvintes, visto que houve maior retenção de características perceptuais da voz do locutor alvo na memória acústica desses ouvintes.

A média percentual de identificações corretas dentre todas as filas de reconhecimento não mostrou diferenças significantes entre os dois grupos de falantes. Embora os ouvintes brasileiros apresentassem uma

tendência em alcançar um desempenho melhor, os resultados revelaram que o teste foi difícil para ambos os grupos de ouvintes, atestando nível baixo de concordância entre os ouvintes de um mesmo grupo ($kappa_{suecos} < 0,12$ e $kappa_{brasileiros} < 0,02$).

2.2 Análise das falsas identificações por distrator

As taxas de falsas identificações por distrator também foram calculadas. O distrator que obteve o maior número de falsas identificações, para os ouvintes suecos, foi o locutor número 1, com 17 falsas identificações ao todo (representando 22,4% sobre a quantidade total de falsas identificações diante de todos os distratores). Os ouvintes brasileiros erroneamente escolheram, com uma maior frequência, o locutor número 4, com 8 falsas identificações ao todo (representando 24,2% sobre a quantidade total de falsas identificações diante de todos os distratores). A tabela 4 apresenta, detalhadamente, as taxas de falsas identificações para ambos os grupos de ouvintes e as porcentagens correspondentes às possíveis falsas escolhas por distrator considerando o número de versões de cada distrator e a quantidade de ouvintes em cada grupo.[3]

[3] Para este cálculo, foi considerado o número de ouvintes testados em cada grupo, multiplicado pela quantidade total de versões de um mesmo distrator. Por exemplo, para o grupo de ouvintes suecos, para o locutor (distrator) 1, existiam 5 versões desse distrator e 20 ouvintes, o que resultava em um total de 100 possíveis escolhas errôneas deste distrator. Como houve 17 escolhas incorretas, a escolha desse distrator representa 17% de todas as escolhas possíveis.

Falsas identificações por distrator							
Distrator (nº)	1	2	3	4	5	6	7
Ouvintes suecos	17 (17,0)	13 (16,2)	4 (5,0)	15 (18,7)	14 (35,0)	-	13 (32,5)
Ouvintes brasileiros	7 (14,0)	3 (7,5)	5 (12,5)	8 (20,0)	7 (35,0)	1 (10,0)	2 (10,0)

Tabela 4. Taxas de falsas identificações por distrator e porcentagem correspondente, em parênteses.

2.3 Análise das falsas identificações pela posição da amostra de fala na fila de reconhecimento

A posição na fila de reconhecimento que foi escolhida com maior frequência, para ambos os grupos

de ouvintes, foi a posição número 3. Os ouvintes suecos escolheram erroneamente esta posição 20 vezes (representando 25,6% sobre a quantidade total de posições erroneamente escolhidas), enquanto os ouvintes brasileiros escolheram erroneamente essa posição 12 vezes (representando 36,4% sobre a quantidade total de posições erroneamente escolhidas). A tabela 5 apresenta, detalhadamente, as taxas de falsas identificações pela posição na fila de reconhecimento para ouvintes brasileiros e suecos e as porcentagens correspondentes às possíveis falsas escolhas por posição considerando a quantidade total de posições e de ouvintes participantes em cada grupo.

Falsas identificações pela posição na fila de reconhecimento						
Distrator (posição)	1	2	3	4	5	6
Ouvintes suecos	6 (5,0)	15 (12,5)	20 (16,7)	16 (13,3)	11 (9,2)	10 (8,3)
Ouvintes brasileiros	2 (3,3)	5 (8,3)	12 (20,0)	8 (13,3)	3 (5,0)	3 (5,0)

Tabela 5. Taxas de falsas identificações pela posição na fila de reconhecimento e porcentagem correspondente, em parênteses.

2.4 Análise das falsas identificações pelas amostras de fala de um mesmo distrator (versão)

Os dados permitiram também analisar a quantidade total de falsas identificações do locutor-alvo por meio do cálculo de escolhas dos ouvintes pela versão do locutor (distrator). Como apresentado anteriormente, para os ouvintes suecos, o locutor número 1 obteve a maior frequência entre os distratores erroneamente escolhidos, com sua versão 1.3 falsamente escolhida mais vezes entre todas as outras versões deste locutor (29,4%). Os ouvintes brasileiros escolheram erroneamente o locutor número 4 com maior frequência, sendo sua versão 4.3 falsamente escolhida mais vezes entre as outras versões deste mesmo

locutor (87,5%). A tabela 6 mostra as escolhas de ambos os grupos de ouvintes para cada versão de locutor (distrator) e, também, a porcentagem correspondente para cada versão de locutor (distrator).

Falsas Identificações por versão		
Versão	*Ouvintes suecos*	*Ouvintes brasileiros*
	Contagem (%)	
1.1	2 (11,7)	-
1.2	3 (17,6)	1 (14,2)
1.3	5 (29,4)	5 (71,4)
1.4	3 (17,6)	-
1.5	4 (23,5)	1 (14,2)
2.1	5 (38,4)	2 (66,6)
2.2	4 (30,7)	1 (33,3)
2.3	3 (23,0)	-
2.4	1 (7,6)	-
3.1	1 (25,0)	-
3.2	1 (25,0)	2 (40,0)
3.3	-	1 (20,0)
3.4	2 (50,0)	2 (40,0)
4.1	4 (26,6)	-
4.2	1 (6,6)	1 (12,5)
4.3	8 (53,3)	7 (87,5)
4.4	2 (13,3)	-
5.1	1 (7,1)	1 (14,2)
5.2	13 (92,8)	6 (85,7)
6.1	-	1
7.1	9 (69,2)	-
7.2	4 (30,7)	2

Tabela 6. Taxas de falsas identificações por versão de locutor (distrator) e porcentagem correspondente, em parênteses.

2.5 Análise da duração relativa do grupo acentual e da taxa de unidades VV/s entre as amostras de fala do locutor-alvo e dos distratores escolhidos

A análise da taxa de reconhecimento de locutor-alvo pela avaliação das falsas identificações do locutor-alvo pôde ser aprimorada por uma análise comparativa

da duração relativa da saliência duracional e da taxa de elocução em unidades VV por segundo. Barbosa (2006) descreve a unidade VV como um componente prosódico, cuja evolução duracional ao longo do enunciado está diretamente associada ao ritmo da fala, revelando sua estrutura rítmica. Os valores das unidades VV em uma passagem permitem avaliar tanto a taxa de elocução, bem como a taxa da duração normalizada das saliências duracionais nessa passagem. Estes são importantes parâmetros a ser considerados em análises de fala inter-locutores, pois permitem avaliar se a diferença dos valores médios entre os distratores escolhidos e o locutor-alvo são suficientes para justificar uma falsa escolha.

A análise comparativa da duração de unidades VV e da duração dos grupos acentuais, delimitados pelos picos locais de duração relativa, foi calculada pelo *script* SGdetector (BARBOSA, *ibid.*), a partir dos intervalos VV marcados em *textgrids* no programa PRAAT. A distância entre dois picos locais de durações normalizadas de unidades VV define um grupo acentual, cuja duração é computada pelo programa.

A análise comparativa da duração de unidades VV e da duração relativa da saliência duracional foi calculada entre as amostras de fala do locutor-alvo e do distrator falsamente escolhido com maior frequência, em cada fila de reconhecimento, para ambos os grupos de ouvintes (suecos e brasileiros). Um teste T de variáveis independentes, com nível de significância de 0,05, foi conduzido com a finalidade de testar a hipótese nula de que as duas amostras comparadas possuíam a mesma média para a duração das unidades VV e para a duração do grupo acentual. A tabela 7 mostra os resultados do teste T para a análise comparativa entre os dois grupos de ouvintes.

Resultado do teste T: probabilidade p				
Fila de reconhecimento	*Ouvintes suecos*		*Ouvintes brasileiros*	
	VV	Dur_{GA}	*VV*	Dur_{GA}
1	1.10^{-3}	0,27	0,04	0,43
2	0,98	0,26	0,04	0,52
3	0,83	0,74	0,83	0,74
4	0,17	0,11	0,53	0,69
5	5.10^{-4}	0,08	1.10^{-4}	0,12
6	0,55	0,11	0,94	0,08

Tabela 7. Resultados do teste T para a análise comparativa da duração das unidades VV (VV) e da duração do grupo acentual (Dur_{GA})

O teste T não apresentou resultados que corroborassem para a aceitação da hipótese nula ($p > 0,05$) para as durações médias das unidades VV na fila de reconhecimento 1, para os ouvintes suecos ($p < 0,001$) e para os ouvintes brasileiros ($p < 0,04$), também na fila de reconhecimento 5, com um p-valor menor que 5.10^{-4} e 1.10^{-4} para ouvintes suecos e brasileiros, respectivamente. O teste T também não mostrou resultados significativos na fila de reconhecimento 2 para os ouvintes brasileiros ($p < 0,04$). Todos os resultados do teste T para a duração do grupo acentual não apresentaram diferenças significantivas entre os valores estabelecidos para os locutores-alvo e os distratores ($p > 0,05$).

Os resultados mostraram que, na maioria das filas de reconhecimento (exceto nas filas de reconhecimento 1, 5 e 2, no caso dos ouvintes brasileiros), o locutor-alvo e o distrator comumente escolhido (falsa identificação) produziram a mesma quantidade de unidades VV por segundo. Todos os locutores comparados (alvo e distratores comummente escolhidos) produziram as mesmas taxas de duração do grupo acentual. A conservação da hipótese nula, em ambos os parâmetros de análise, pode explicar a escolha incorreta dos locutores, já que eles podiam apoiar suas escolhas nestas pistas rítmicas.

A consideração de outro parâmetro – a mediana de *f0* (Hz) – para todas as amostras de fala selecionadas e sua comparação entre os dois locutores escolhidos (alvo e distrator) também auxiliou na compreensão das taxas de falsas identificações pelos ouvintes. Os resultados da mediana de *f0* serão apresentados na seção seguinte.

2.6 Análise da mediana de *f0* (Hz) entre amostras de fala selecionadas

A mediana de *f0* (Hz) foi calculada para cada par de locutores (alvo e distrator erroneamente escolhido com maior frequência) em todas as filas de reconhecimento. A análise deste parâmetro pôde aprimorar os resultados estabelecidos para a duração das unidades VV e para a duração do grupo acentual, e também auxiliar na compreensão das falsas identificações de ambos os grupos de ouvintes. A tabela 8 apresenta as taxa da mediana de *f0* para todos os pares de locutores comparados para ambos os grupos de ouvintes, em todas as filas de reconhecimento.

Taxas da mediana de *f0* (Hz)				
	Ouvintes suecos		Ouvintes brasileiros	
Fila de reconhecimento	Locutor alvo	Distrator (Falsa ID)	Locutor alvo	Distrator (Falsa ID)
1	132	164	132	130
2	140	164	140	118
3	255	267	255	267
4	236	258	236	230
5	223	222	223	266
6	143	127	143	117

Tabela 8. Valores da mediana de *f0*, em Hertz, para todos os pares de locutores (alvo e distratores)

Os valores para a mediana de *f0* mostraram-se similares entre os pares de locutores nas filas de

reconhecimento 3 e 5 para os ouvintes suecos. Um ponto interessante a ser notado, é que, pra esse grupo de falantes, o resultado do teste T para a duração das unidades VV por segundo (taxa de elocução) não foi significante para as amostras de fala da fila de reconhecimento 5, mas uma possível explicação para a escolha errônea desses locutores pode ser dada pela similaridade nos valores estabelecidos para a mediana de *f0* desses locutores, que é praticamente a mesma (223 Hz, para o locutor-alvo e 222 Hz para o distrator erroneamente escolhido). A mesma hipótese pode ser utilizada para algumas comparações entre as escolhas dos ouvintes brasileiros. Os valores da mediana de *f0* apresentaram resultados similares para as amostras de falas entre os locutores nas filas de reconhecimento 1, 3 e 4. Para este grupo de ouvintes, o teste T para análise da taxa de elocução das unidades VV/s não apresentou resultados significantes para as amostras de fala da fila de reconhecimento 1. Por outro lado, a similaridade entre os valores da mediana de *f0* para o locutor-alvo e o distrator comumente escolhido, na fila de reconhecimento 1, pode explicar a falsa identificação desses ouvintes.

3. Conclusão

Apesar de os resultados não terem apresentado diferenças significativas na média percentual de escolhas corretas entre ambos os grupos de ouvintes, o fato dos ouvintes brasileiros apresentarem uma tendência em alcançar um melhor desempenho no teste de percepção, pode concordar com discussões anteriores de Schiller e Köster (1996) que afirmam que a prática de reconhecimento de locutor parece não envolver apenas características puramente fonéticas, mas também a incorporação de informação linguística, desde que esse grupo de ouvintes possua um conhecimento prévio da língua do locutor-alvo.

A análise da taxa de reconhecimento de locutor pela avaliação das falsas identificações do locutor-alvo foi aprimorada por uma análise comparativa da duração do grupo acentual e da taxa de elocução em unidades VV por segundo. Os resultados de um teste T de variáveis independentes (α = 0,05) mostraram que a conservação da hipótese nula em ambos os parâmetros, na maioria das filas de reconhecimento comparadas, poderia explicar a escolha errônea dos ouvintes, visto que eles poderiam estar apoiando suas escolhas nestas pistas rítmicas. A análise da mediana de *f0* ajudou a compreender as taxas de falsas identificações e também a aprimorar os resultados das durações médias das unidades VV (inverso da taxa de elocução em VV/s) e das durações dos grupos acentuais. O resultado mais importante da análise da mediana de *f0* diz respeito aos valores entre os pares de locutores da fila de reconhecimento 5 (para os ouvintes suecos) e na fila de reconhecimento 1 (para os ouvintes brasileiros). Para essas filas de reconhecimento, os valores atestados para este parâmetro mostraram algumas similaridades entre os pares de locutores escolhidos, o que auxiliou na compreensão das falsas identificações dos ouvintes, visto que os resultados apresentados para o teste T para a taxa de elocução em unidades VV/s foram inconclusivos para essas filas de reconhecimento.

4. Agradecimentos

Nossos agradecimentos aos colegas do Laboratório de Fonética do Departamento de Linguística, da Universidade de Gotemburgo, pela ajuda na condução do teste de percepção; à Lisa Öhman, pela ajuda com os ouvintes do Departamento de Psicologia, da Universidade de Gotemburgo; a todos os ouvintes que participaram deste experimento e à FAPESP pela concessão e financiamento desta Bolsa de Estágio de Pesquisa no Exterior (Processo/ FAPESP: 2012/10909-3).

Referências

BARBOSA, P. A. *Incursões em torno do ritmo da fala.* Campinas, Brasil: Pontes/FAPESP, 2006, p. 170.

BOERSMA, P. & WEENINK (2012). *Praat: doing phonetics by computer* (Versão 5.3.16) [Programa computacional]. Disponível em: <http://www.fon.hum.uva.nl/praat/>

BRICKER, P.D.; PRUZANSKY, S. *Effects of Stimulus Content and Duration on Talker Identification*. In: The Journal of the Acoustical Society of America, v.40, p. 1441-1450, 1966.

BROEDERS, A.P.A; VAN AMELSVOORT, A.G. *Lineup construction for forensic earwitness identification: a practical approach*. In the Proceedings of the 7th International Congress of Phonetic Sciences. San Francisco, p. 1373-1376, 1999.

ERIKSSON, A. *Tutorial on Forensic Speech Science*. Paper presented at tutorial session on Forensic Speech Science, Interspeech (9th European Conference on Speech Communication and Technology), Lisbon, Portugal, 2005.

KÖSTER, O.; SCHILLER, N.O. *Different influences of the native language of a listener on speaker recognition*. In: Forensic Linguistic, v.4, n.1, p.176-185, 1997.

POLLACK, I.; PICKETT, J.M.; SUMBY, W.H. *On the identification of speakers by voice*. In the Journal of the Acoustical Society of America, v.26, n.3, p.403-412, 1954.

SCHILLER, N.O.; KÖSTER, O. *Evaluation of a foreign speaker in forensic phonetics: a report*. In: Forensic Linguistics, v.3, n.1, p. 176-185, 1996.

[Recebido em 31 de agosto de 2013
e aceito para publicação em 14 de novembro de 2013]

Apêndice

Texto informativo para os ouvintes participantes do experimento

Imaginem a seguinte situação:

Você foi vítima de um crime. Durante um interrogatório policial, você informou à polícia que, no momento do incidente, apenas ouviu a voz de uma pessoa envolvida nesse crime e que essa pessoa parecia utilizar um disfarce vocal que ocultava as verdadeiras características de sua voz.

No decorrer da investigação policial, foi encontrada uma pessoa que pode ter cometido esse crime. Entretanto, não há certeza se esse suspeito é, de fato, o criminoso.

Uma gravação foi feita com a voz desse suspeito. Além disso, outras gravações foram feitas com vozes de outras pessoas semelhantes à voz do suspeito. Essas amostras de vozes adicionais são chamadas "distratoras". As vozes distratoras ***NÃO*** *são suspeitas de terem cometido o crime.*

*Você está prestes a ouvir as vozes gravadas. Em cada gravação, você ouvirá primeiramente uma amostra de voz contendo a f****ala disfarçada do criminoso*** *e, em seguida,* ***vozes sem nenhum disfarce****.*

Essas amostras de fala consistirão na organização das vozes distratoras e da voz do suspeito. ***Sua tarefa será determinar qual voz sem disfarce corresponde à amostra de voz disfarçada do criminoso ouvida anteriormente****.*

Ao reconhecer a voz disfarçada ouvida anteriormente, você deverá marcar, na folha de respostas, um "X" na posição referente a ela.

ATENÇÃO: Para cada apresentação haverá apenas UMA voz sem disfarce correspondente à voz disfarçada ouvida anteriormente.

Ao todo serão ouvidas 6 gravações diferentes, cada uma contendo uma amostra disfarçada, seguida de 5 vozes sem disfarce.