

Fronteiras do vozeamento na identificação do falante

Fernando Martins

Faculdade de Letras da Universidade de Lisboa – Centro de Linguística da Universidade de Lisboa

fmartins@campus.ul.pt

*Celeste Rodrigues*¹

Faculdade de Letras da Universidade de Lisboa – Centro de Linguística da Universidade de Lisboa

celesterodrigues@campus.ul.pt

*Fernando Brissos*²

Centro de Linguística da Universidade de Lisboa

fernandobrissos@campus.ul.pt

Abstract: This paper aims at testing a new parameter of acoustic analysis of speech to be used in forensic phonetic analysis. We state the following hypothesis: *Prevoicing values depend on speaker physiological and motion abilities and are part of his profile*. We describe the results of the acoustic analysis of speech of fourteen European Portuguese speakers and we compare them to a real forensic case. Our results show that Prevoicing is clearly a cue for speaker identity and, consequently, it must be included in acoustic forensic evaluation reports.

Keywords / Palavras-chave: Acoustic Phonetics, Forensic Phonetics, Speaker Identification, Prevoicing / Fonética Acústica, Fonética Forense, Identificação do Falante, Pré-vozeamento

1. Introdução

Os modelos mais recentes de avaliação em Fonética Forense têm demonstrado a necessidade de uma investigação mais aprofundada, no sentido de fornecerem informações mais fiáveis às entidades judiciais. A identificação do falante baseia-se no princípio de que as produções fonéticas incluem as informações suficientes para o reconhecimento do seu autor. Estas informações individuais estão presentes na cadeia de fala, tal como as informações linguísticas que codificam a mensagem transmitida.

A Fonética Acústica ao longo das últimas décadas, em função da utilização de novos programas de análise do sinal sonoro, tem vindo a ser aplicada a cada vez mais domínios, desde logo a área forense. São reconhecidos internacionalmente os bons resultados assim obtidos nas tarefas de identificação e de eliminação forenses. Em Portugal, os estudos em Fonética Acústica Forense têm sido levados a cabo, desde os primeiros tempos, no âmbito do Laboratório de Fonética da Faculdade de Letras da Universidade de Lisboa e têm prosseguido com a realização de diferentes trabalhos sob a sua alçada. À medida que os trabalhos avançam, torna-se claro que a análise cada vez mais fina do sinal acústico poderá trazer mais indícios fortes da autoria da voz.

¹ Financiado por Fundos Nacionais através da Fundação para a Ciência e a Tecnologia (FCT) no âmbito do projecto PEst-OE/LIN/UI0214/2011.

² Financiado pela bolsa de pós-doutoramento da Fundação para a Ciência e a Tecnologia com a referência SFRH/BPD/78479/2011.

As perícias em Fonética Acústica Forense utilizam como referência a grelha de conclusões estabelecida pelo American Board of Recorded Evidence (ABRE) e pela International Association of Forensic Phonetics and Acoustics (IAFPA). Essa grelha define os critérios necessários para a distinção de sete níveis, que vão desde a identificação à eliminação do falante. Os níveis dessa escala, de base qualitativa, são os seguintes:

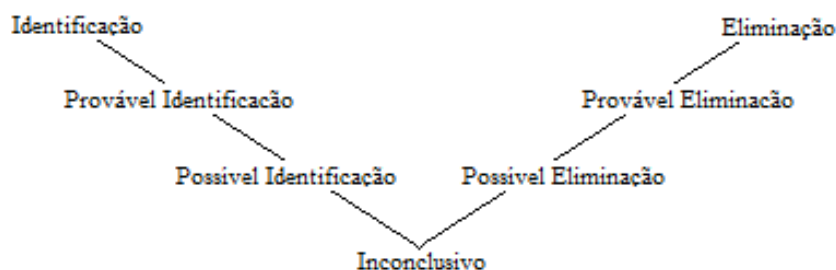


Figura 1: Níveis de confiança nas conclusões em Fonética Forense (ABRE)

Cada nível é definido em função da verificação de determinadas condições, de tal modo que os extremos da escala só serão atingidos caso a análise satisfaça certos critérios. Estes extremos, no topo da escala, correspondem aos níveis de identificação e de eliminação do falante. A escala aproxima-se, assim, de uma escala confiável de conclusões forenses, em termos probabilísticos. Note-se que, no âmbito de um caso forense, é tão importante a demonstração da identidade do falante presente na gravação como a eliminação dessa possibilidade.

Na maioria das situações forenses, trata-se de comparar a gravação de um suspeito (gravação K) com uma gravação não identificada (gravação X, tipicamente produzida em escuta telefónica). Existem diversos factores condicionantes dos resultados: presença de ruído, limitação da banda telefónica (entre 300Hz e 3400Hz), amostra gravada de curta duração, tentativas de disfarce e/ou imitação, suspeito pouco colaborante, etc..

Para se atingir o valor mais elevado da escala, é necessário que se verifiquem as seguintes condições: presença de três formantes bem definidos; pelo menos vinte padrões acústicos de palavras que possam ser comparados; as gravações comparadas com um hiato temporal igual ou inferior a seis anos; ausência de disfarce e/ou imitação nas gravações. A não verificação de qualquer destas condições tem como consequência uma descida na escala, passando-se assim, sucessivamente, para o nível de *provável* (*identificação* ou *eliminação*), depois *possível* (*identificação* / *eliminação*) e finalmente para o nível *inconclusivo*.

Esta escala não oferece, porém, as garantias de rigor que seriam atingidas com escalas de base quantitativa, as quais implicam a análise através de parâmetros objectivamente mensuráveis; esse tipo de escala ainda não está disponível, em formulação adequada, em Fonética Forense. Das tentativas de aplicar uma escala desse tipo à Fonética Forense, ressalta o *modelo Tosi de identificação de voz* (cf. Tosi, 1979). Nesse modelo, a base de análise é o *voiceprint*, em que as variações dos padrões acústicos no espectrograma são associadas a padrões de identificação dos falantes e não a padrões de identificação dos segmentos, tendo sido rejeitado pela comunidade científica (IAFPA - <http://www.iafpa.net/voiceprintsres.htm>).

Desse modo, as evidências obtidas mediante a escala apresentada na figura 1 recebem frequentemente, da parte das instituições judiciais, uma apreciação inferior àquela que poderiam ter se fossem precedidas de estudos quantitativos estatisticamente relevantes para a sua fundamentação.

Os critérios normalmente utilizados nas perícias incluem os seguintes parâmetros de análise:

- (i) Parâmetros de fonte, entre os quais a frequência fundamental (F0) e as harmónicas;
- (ii) Parâmetros de filtro, entre os quais os valores dos formantes e a distribuição espectral da fricção / explosão.

A carta de formantes, que sistematiza em gráfico os valores formânticos das vogais, define um espaço acústico que é correlato do espaço articulatório existente durante a produção dos segmentos. Na maioria das situações, este é um espaço que permite caracterizar os segmentos produzidos com uma fonte laríngea e com filtros na cavidade oral. Fisiologicamente, fonte e filtro variam em função dos órgãos articuladores de cada falante, sendo importantes, por isso, para a sua identificação.

Os factores condicionantes acima referidos podem impedir que o nível de confiança da conclusão atinja o máximo da escala (o de *identificação* ou *eliminação*), como aconteceu no Caso C_2002 (nome fictício), que passamos a descrever.

Um utilizador da banca *online* (aqui denominado UB1) negou ser o autor de uma chamada telefónica em que era dada uma ordem de venda de acções, depois de se ter verificado que o valor destas subira consideravelmente após a operação. A entidade bancária detinha a prova da ordem dada e defendeu-se apresentando as gravações das respectivas conversas. As entidades judiciais solicitaram uma peritagem com o objectivo de avaliar a possibilidade de se demonstrar que a voz presente nas gravações correspondia à voz do suspeito. Assim, foram feitas gravações com UB1 (Gravações K), de modo a verificar se havia uma correspondência com o autor das chamadas telefónicas apresentadas (Gravações X).

Durante a peritagem, foram utilizados procedimentos de comparação *standard* (Nolan, 1983; Rose, 2002; Müller, 2007), que envolvem os parâmetros que se seguem:

- (i) carta de formantes;
- (ii) zonas formânticas das consoantes fricativas;
- (iii) duração de VOT das consoantes oclusivas.

Os resultados determinaram uma conclusão de *possível identificação*, de acordo com a escala apresentada na figura 1. Este nível da escala pressupõe que 80% das palavras que é possível comparar sejam muito semelhantes auditiva e acusticamente, obtendo-se não menos do que dez pares de palavras comparáveis com, pelo menos, dois formantes visíveis (cf. ABRE).

O nível baixo de identificação neste caso resultou da existência de:

- (i) baixa visibilidade dos formantes, em especial de F1 (mesmo descontado o *mobile effect*; Byrne & Foulkes, 2004); este factor, devido ao limite inferior da banda telefónica disponível, que se inicia em 300 Hz, condiciona sobretudo a análise das vogais altas (ou seja, as que têm um F1 baixo);
- (ii) visibilidade limitada das zonas formânticas das consoantes fricativas; esta limitação deve-se ao facto de a banda telefónica disponível não ultrapassar os 3400 Hz e condiciona, sobretudo, a análise das consoantes fricativas palatais e dentais, de duas formas: (a) reduzindo fortemente a discriminação dessas consoantes; e (b) impedindo a detecção de eventuais outras zonas formânticas que, não sendo definidoras dos segmentos, são definidoras do falante;
- (iii) um intervalo superior a seis anos entre as gravações X e K (que, por si só, impossibilita que se atinja o nível máximo da escala, tal como vimos acima).

Devido a estas limitações, no decurso do julgamento a defesa alegou fraqueza e falibilidade da prova, por ser muito próxima do nível *inconclusivo*. Sugeriu, em consequência, que os resultados da peritagem não deveriam ser tidos em conta como prova. Apesar disso, na fase final do julgamento, UB1 veio a reconhecer a sua responsabilidade no caso, confessando ser autor da gravação X.

Este desfecho, que corresponde à confirmação da autoria do falante UB1 na gravação X, mostra que os vários factores condicionantes impediram que a prova se fizesse com base nos parâmetros acústicos analisados; ou seja, as exigências do modelo de avaliação utilizado para se subir / descer níveis na escala de conclusões acabaram por limitar decisivamente uma conclusão (identificação do falante) que, para todos os efeitos, correspondia à realidade. Tal deve-se à permeabilidade que os parâmetros analisados têm aos factores referidos.

A procura de novos parâmetros de análise é, desse modo, uma necessidade sentida por todos aqueles que trabalham na área. Só com a sua introdução se poderá obter uma independência de factores externos que permita resistir ao abaixamento sucessivo na escala de conclusões. Idealmente, os novos parâmetros deveriam apresentar as seguintes características:

- (i) ser detectáveis nas frequências presentes na banda telefónica;
- (ii) ser pouco influenciados por ruído externo;
- (iii) estar presentes mesmo nas situações de disfarce (Zhang & Tan, 2008);
- (iv) estar ausentes na imitação;
- (v) não ser afectados de forma significativa por existirem intervalos de tempo prolongados entre as gravações X e K.

Procurando responder a esse objectivo, este trabalho reflecte a nossa preocupação de refinar a análise acústica com fins forenses. Exploramos a possibilidade de introduzir um novo parâmetro de análise que ajude a melhorar a performance dos resultados das perícias que, hoje em dia, são aplicadas globalmente. O parâmetro estudado é a duração da fracção de pré-vozeamento que antecede a realização de qualquer vogal na transição a partir de uma consoante não-vozeada (designaremos o parâmetro pela abreviatura *PreVoz*). A duração do *PreVoz* não deve ser confundida com a duração do VOT (Lin & Wang, 2011): o VOT é originado pela co-articulação existente entre a consoante e a vogal que se lhe segue, ou seja, está indexado ao filtro; o *PreVoz* está indexado à fonte: não distingue consoantes ou co-articulações, mas provavelmente os falantes. Por isso, decidimos pesquisar se a antecipação do vozeamento, que corresponde ao tempo de preparação das cordas vocais para iniciar a produção do segmento vozeado, pode ser importante para traçar o perfil individual da voz de cada falante.

Estabelecemos a seguinte hipótese de trabalho:

H-1: Os valores de duração de pré-vozeamento dependem do falante, podendo fazer parte do seu perfil vocal.

De acordo com H-1, supomos que haverá valores médios e níveis de variação possível implementados por cada falante, cujos valores estarão em zonas de conforto típicas, à semelhança do que acontece com outros parâmetros, como F0, os formantes, a duração de segmentos, etc..

Afigurou-se-nos necessário obter dados relativos a um número significativo de falantes que pudessem servir de referência para o estabelecimento de comparação com vozes gravadas em casos forenses, uma vez que tal base de dados ainda não existe para o Português Europeu. Nessa medida, recorreremos a um conjunto de entrevistas, previamente existentes, do CPE-Var (Rodrigues & Andrade, 1998; Rodrigues, 2003).

Além da obtenção desses valores de referência, aplicamos também a metodologia a um caso forense que será descrito mais adiante.

2. Metodologia

2.1. Descrição dos dados usados do CPE-Var

A utilização de gravações do CPE-Var justifica-se pelas seguintes razões: (i) as gravações têm um tipo de registo adequado, (ii) foram sociolinguisticamente controladas e classificadas, e (iii) encontram-se em fase adiantada de tratamento.

Do material disponível no CPE-Var, só foi usada a parte de diálogo informal em ambiente familiar ao falante (a parte das entrevistas sociolinguísticas que Rodrigues (2003) designou Discurso Informal). Essas conversas incidem em temas variados e apresentam grande espontaneidade discursiva, gerada pelo facto de ocorrerem após uma longa tarefa de leitura. Não sendo possível para este trabalho recriar uma situação mais comparável ao registo captado, de forma geral, nas gravações forenses, estas conversas apresentam um mínimo de características adequadas para o escrutínio das propriedades acústicas em registo informal. Estas conversas foram todas feitas pelo mesmo investigador, que procurou criar com todos os falantes uma situação discursiva semelhante.

As gravações foram obtidas com um gravador Marantz PMD 222 e um microfone Uher externo unidireccional, instalado num tripé de mesa em frente ao entrevistado. As gravações foram depois sujeitas a digitalização e a transcrição ortográfica em ficheiros EXMARaLDA (Partitur Editor 1.5.2). O software permite o alinhamento do ficheiro áudio com a respectiva transcrição, o que facilita a pesquisa dos dados pretendidos. Isso facilitou em certa medida a obtenção dos dados relevantes para a análise acústica, uma

vez que, por se tratar de fala espontânea, não é previsível a localização no sinal de segmentos ou sequências específicos.

O facto de o material estar classificado quanto ao perfil sociolinguístico dos falantes possibilitou o estabelecimento de comparações entre grupos de falantes, o que se revelou útil para a eventual identificação de padrões comportamentais e permitirá a sua utilização futura em qualquer comparação com novos dados.

Nesta fase, foram utilizadas gravações de catorze homens naturais de Braga, pertencentes a duas faixas etárias: 25-39 e 50-65 anos. Foram seleccionados homens destas faixas etárias, por haver maior número de delitos praticados por homens, segundo o *Relatório Anual de Segurança Interna* (RASI 2012, pp.85-86). Braga é também uma das cinco cidades portuguesas onde se regista maior número de delitos participados (RASI 2012, p. 48).

Os informantes serão designados pelo seu número no CPE-Var: informante 80, informante 142, etc..

2.2. Procedimentos

2.2.1. Medições do PreVoz

O registo dos valores é feito em segundos (s).

Utilizámos para identificar o PreVoz o programa Praat (Boersma & Weenink, 2013), em dois tipos de visualização: (i) espectrograma de banda larga (0.005s), alcance de 0-5000 Hz, *dynamic range* de 40 db e duração de 0.5s; (ii) representação da onda sonora, que, no programa Praat, surge acima da representação espectrográfica no ecrã principal. O segundo tipo de visualização foi utilizado de modo auxiliar, com o objectivo de confirmar a interpretação feita no espectrograma.

A obtenção dos valores foi realizada sempre na visualização espectrográfica, por meio de medição manual. Seleccionou-se o segmento temporal que corresponde ao PreVoz, utilizando o cursor; veja-se um exemplo na figura 2.

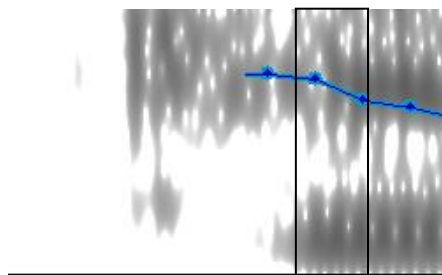


Figura 2: Exemplo de medição de PreVoz

O rectângulo indica o início e o fim do PreVoz. A linha azul representa o *pitch* no programa Praat, ferramenta utilizada auxiliarmente mas sem implicação na obtenção dos valores do PreVoz.

2.2.2. Tratamento estatístico dos dados

Os dados foram submetidos a tratamento estatístico usando o programa STATISTICA. Verificou-se que os dados seguiam uma curva de distribuição normal através da aplicação do teste de Shapiro-Wilk. Aplicámos a One-way ANOVA, através da avaliação a partir de intervalos de confiança de 95% e por meio de aplicação do teste de Tukey (definição de grupos homogéneos)..

2.3. Identificação dos contextos

2.3.1. Contexto consonântico

Foram seleccionadas as três consoantes oclusivas não vozeadas da língua, [p,t,k], com base em dois critérios:

- (i) consoantes com o modo de articulação oclusivo, porque pretendíamos maximizar a diferença entre C e V: [p,t,k] (Ladefoged & Johnson, 2011);

- (ii) consoantes com os três pontos de articulação disponíveis na língua, porque poderia existir dependência relativamente ao ponto de articulação, tal como acontece com o tempo de explosão + VOT, fenómeno articulatorio próximo do PreVoz.

2.3.2. Contexto vocálico

Com o objectivo de verificar se existe influência da qualidade das vogais nos valores de duração do PreVoz, foi efectuado um pré-teste. Para tal foram utilizados os dados do informante 142 do CPE-Var, de onde foram medidas doze ocorrências de várias vogais antecedidas das três consoantes [p], [t] [k].

Os resultados, patentes nas figuras 3 e 4, mostram que não existem diferenças significativas nos tempos de PreVoz das diferentes vogais. Mais concretamente, a figura 3 revela uma sobreposição efectiva (com o mesmo intervalo de confiança) do espaço de duração do PreVoz em todas as vogais. As barras verticais representam intervalos de confiança com 95%.

Os valores de PreVoz não são significativamente diferentes usando o teste de Tukey ($p < 0,05$), como se pode observar na Figura 4, dado que formam um grupo estatisticamente homogéneo.

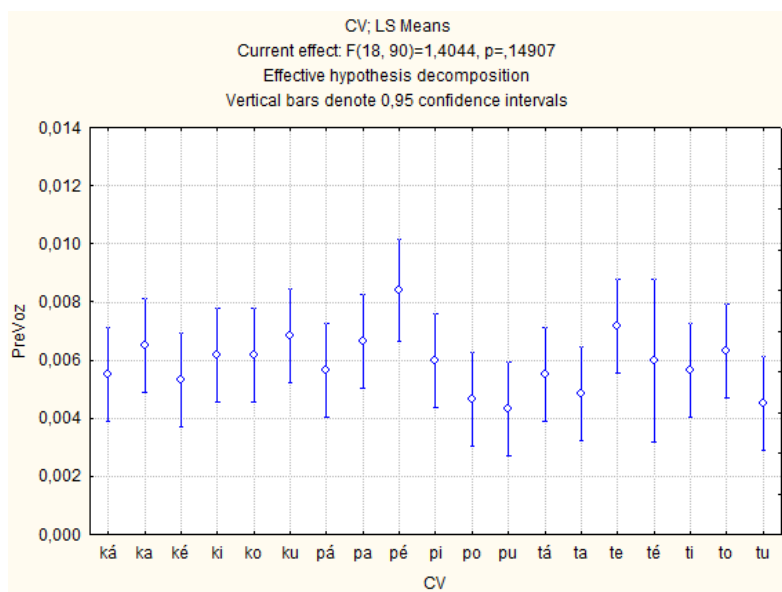


Figura 3: Valores das médias e DP³ de PreVoz em função das vogais do informante 142

³ DP = desvio-padrão.

Tukey HSD test; variable PreVoz (142_k_p_t)			
Homogenous Groups, alpha = .05000 (Non-Exhaustive Search)			
Error: Between MS = .00000, df = 90,000			
Cell No.	CV	PreVoz Mean	1
12	pu	0,004333	****
19	tu	0,004500	****
11	po	0,004667	****
14	ta	0,004833	****
3	ké	0,005333	****
13	tá	0,005500	****
1	ká	0,005500	****
7	pá	0,005667	****
17	ti	0,005667	****
16	té	0,006000	****
10	pi	0,006000	****
4	ki	0,006167	****
5	ko	0,006167	****
18	to	0,006333	****
2	ka	0,006500	****
8	pa	0,006667	****
6	ku	0,006833	****
15	te	0,007167	****
9	pé	0,008400	****

Figura 4: Teste de Tukey aplicado aos valores de PreVoz em função das vogais do informante 142

Ora, se a duração efectiva do PreVoz deste locutor não depende da qualidade da vogal, importa, para além de verificar se ela a condiciona noutros falantes, verificar se a duração do PreVoz se deve à natureza da consoante precedente, nomeadamente no que se refere ao seu ponto de articulação.

Em função dos resultados do pré-teste, deixámos de usar a qualidade das vogais como variável nos restantes falantes, passando, por isso, a considerar-se como variável só o ponto de articulação da consoante oclusiva. Ficou garantido, no entanto, o equilíbrio entre as diferentes vogais no número de ocorrências analisadas em cada consoante, por informante.

No que diz respeito à potencial influência do tipo de consoante na duração do PreVoz, uma vez que nos limitamos à classe das oclusivas orais, só poderemos apresentar resultados relativos aos três pontos de articulação utilizados distintivamente em português. Esses resultados serão apresentados na secção 3.

3. Apresentação e discussão dos resultados

3.1. Comparação entre faixas etárias

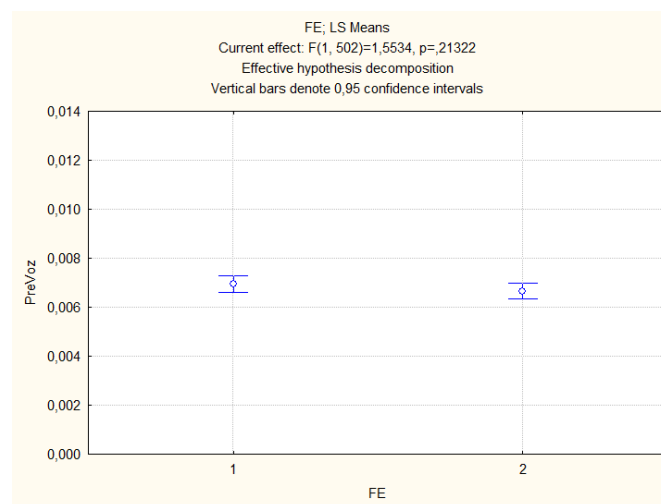


Figura 5: Valores das médias e DP de PreVoz em função da faixa etária (FE)
FE1 = 25-39 anos e FE2 = 50-65 anos.

Os valores de PreVoz não variam em função da idade, ou seja, não existem diferenças estatisticamente significativas entre faixas etárias, como se pode ver na figura 5.

Deste modo, cada falante da amostra foi analisado separadamente com vista ao estabelecimento de um perfil vocal próprio.

3.2. Poder de discriminação do parâmetro em pares de falantes

Em Fonética Forense, uma das metodologias aplicadas frequentemente consiste na comparação binária de gravações de fala com o objectivo de determinar se pertencem ao mesmo falante ou, melhor, se as gravações não apresentam um perfil vocal diferente. Assim, antes de extrairmos os resultados globais dos informantes, veremos três exemplos de aplicação dessa comparação que apresentam informação diferenciada. Em primeiro lugar, comparamos dois falantes com variação reduzida dos valores de PreVoz; em segundo lugar, dois falantes com médias de duração do PreVoz muito distintas e, em terceiro lugar, dois falantes que apresentam tanto distinções de valores de duração média do PreVoz, como do espaço ocupado de duração média.

a) Falante 80 vs. falante 142 — Estes dois falantes têm ambos médias de duração do PreVoz com variabilidade reduzida (142: DP = 0.41ms; 80: DP = 0.30ms), embora ocupem espaços diferentes. O 142 tem um valor médio = 5.20ms e o 80 tem um valor médio = 10.0ms. Os informantes têm, assim, perfis distintos, como as figuras 6 e 7 demonstram. O teste de Tukey ($p < 0,005$) sistematizado na figura 7 coloca os falantes 80 e 142 em dois grupos homogêneos diferentes.

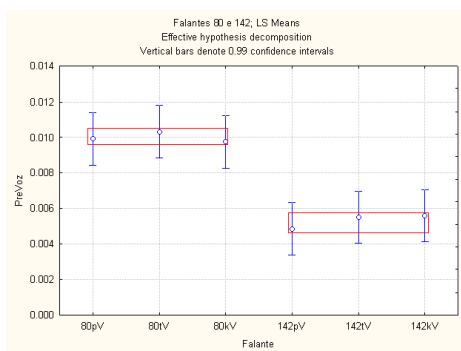


Figura 6: Valores médios de PreVoz em função de [p] [t] e [k] nos falantes 80 e 142

Tukey HSD test; variable PreVoz (80_142)
Homogenous Groups, alpha = .01000
Error: Between MS = .00000, df = 66.000

Cell No.	Falante	PreVoz Mean	1	2
4	142pV	0.004833	****	
5	142tV	0.005500	****	
6	142kV	0.005583	****	
3	80kV	0.009750		****
1	80pV	0.009917		****
2	80tV	0.010333		****

Figura 7: Teste de Tukey aplicado aos valores de PreVoz em função das vogais dos informantes 80 e 142.

b) Falante 107 vs. falante 126 — Estes falantes apresentam médias de duração do PreVoz com variabilidades distintas (126: DP = 1.30ms; 107: DP = 0.73ms), apesar de ocuparem espaços próximos: 126: valor médio = 6.3ms; 107: valor médio = 6.7ms. Deste modo, os informantes não apresentam resultados significativamente diferentes, como fica patente das figuras 8 e 9. Logo, os dois falantes pertencem a um mesmo grupo homogêneo ($p < 0,05$).

Todavia, se observarmos os valores do desvio padrão do PreVoz associado a cada uma das consoantes, podemos ver que não se trata do mesmo padrão, porque a duração apresenta valores mais altos com consoantes diferentes nos dois falantes. Com a comparação deste par de falantes ficamos a

perceber que, mesmo nos casos de sobreposição parcial das áreas relativas à duração do PreVoz, é possível encontrar informação distintiva relevante nos valores específicos associados a cada um dos pontos de articulação das consoantes aqui estudadas.

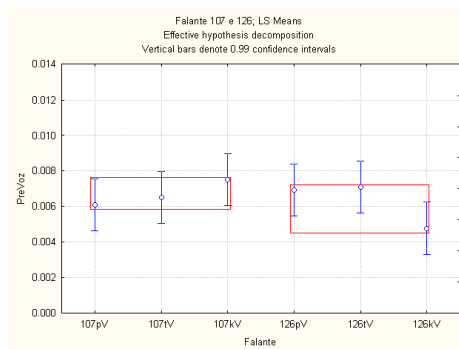


Figura 8: Valores médios de PreVoz em função de [p] [t] e [k] nos informantes 107 e 126

Tukey HSD test; variable PreVoz (107_126)			
Homogenous Groups, alpha = .01000			
Error: Between MS = .00000, df = 66.000			
Cell No.	Falante	PreVoz Mean	1
6	126kV	0.004750	****
1	107pV	0.006083	****
2	107tV	0.006500	****
4	126pV	0.006917	****
5	126tV	0.007083	****
3	107kV	0.007500	****

Figura 9: Teste de Tukey aplicado aos valores de PreVoz em função das vogais dos informantes 107 e 126

c) Falante 80 vs. falante 84 — A comparação destes dois informantes mostra uma diferença significativa tanto na variabilidade das médias (84: DP = 1.88ms; 80: DP = 0.30ms) como nos espaços da duração média do PreVoz: 84: valor médio = 6.2ms; 80: valor médio = 10.0ms. O teste de Tukey ($p < 0.05$) revela que os dois informantes têm valores de PreVoz significativamente distintos. O informante 80 apresenta um perfil homogêneo, ao contrário do 84, ou seja, os falantes têm dois perfis diferentes; as figuras 10 e 11 sistematizam os dados relevantes.

Em consequência dos resultados vistos para os vários falantes, podemos perceber que, tanto pelos valores específicos associados a cada uma das consoantes como pela diferença de valores objectivos apresentados por cada falante, a sua distinção é inequívoca. Estimamos que cada falante apresente, por isso, valores próprios de duração do PreVoz, um ajustamento ao vozeamento vocálico, que variam em função do que lhe é confortável. As áreas de conforto da fonação, como as articulatórias, devem fazer parte do perfil vocal de cada falante.

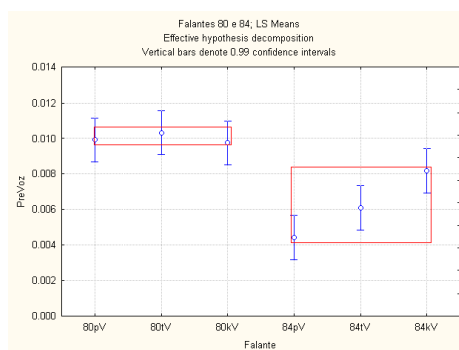


Figura 10: Valores médios de PreVoz em função de [p] [t] e [k] nos falantes 80 e 84

Tukey HSD test; variable PreVoz (80_84) Homogenous Groups, alpha = .01000 Error: Between MS = .00000, df = 66.000					
Cell No.	Falante	PreVoz Mean	1	2	3
4	84pV	0.004417		****	
5	84tV	0.006083	****	****	
6	84kV	0.008167	****		****
3	80kV	0.009750	****		
1	80pV	0.009917	****		
2	80tV	0.010333	****		

Figura 11: Teste de Tukey aplicado aos valores de PreVoz em função das vogais dos informantes 80 e 84

3.3. Resultados globais do CPE-Var

As figuras 12, 13 e 14 apresentam os resultados médios totais dos dados do CPE-Var. Delas podemos derivar a figura 15, em que, a partir da comparação global entre os catorze falantes, se apresenta um valor de medida que contribui para a caracterização do perfil do falante: cada falante é identificado por um código específico, determinado pelas sequências de letras (em substituição dos algarismos utilizados nas imagens das subsecções anteriores, por exemplo no código 1, 2, 3 da figura 11). Não existem nos dados falantes que apresentem o mesmo código.

Podemos concluir da observação destes valores médios que a duração do PreVoz oscila entre 0 e 0.012s, ainda que essa duração oscile entre 0 e 0.017s em valores absolutos. Isso é importante na medida em que se trata de uma característica de duração tão reduzida que deve ser resistente ao disfarce e, por não poder ser manipulável conscientemente, deve ser igualmente resistente à imitação. Dentro do PreVoz de cada falante não se verificam diferenças significativas.

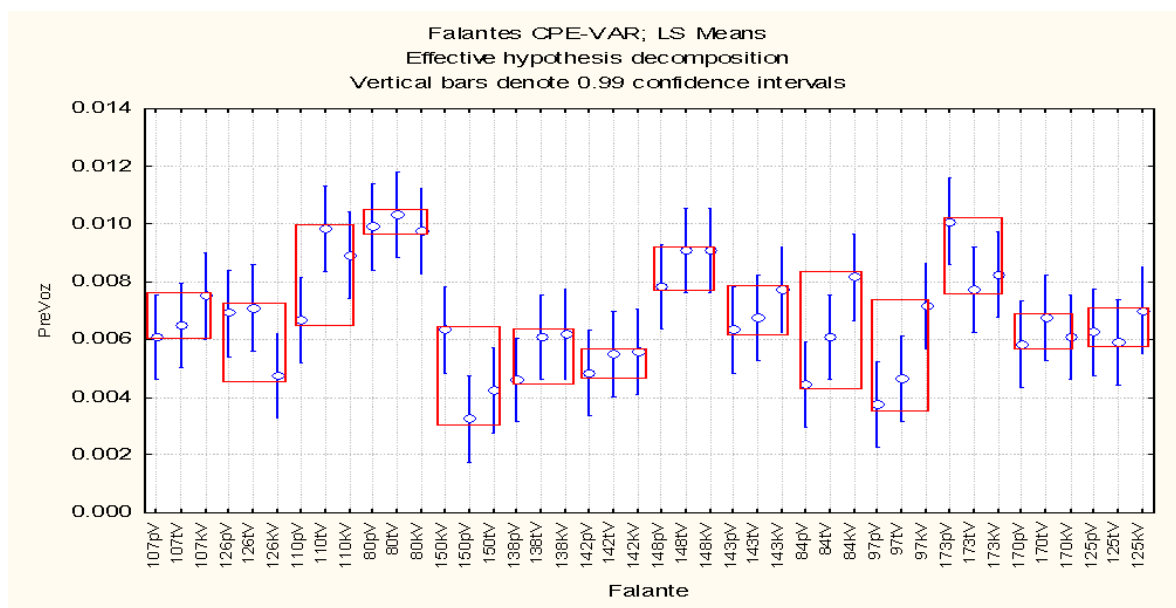


Figura 12: Valores médios e DP de PreVoz de 14 falantes do CPE-Var

107kV			C	D	E	F	G	H	I	J	K	L	M
107pV	A	B	C	D	E	F	G	H					
107tV	A	B	C	D	E	F	G	H	I	J	K		
110kV							G	H	I	J	K	L	M
110pV	A	B	C	D	E	F	G	H	I	J	K	L	
110tV									J	K	L	M	
125kV		B	C	D	E	F	G	H	I	J	K	L	M
125pV	A	B	C	D	E	F	G	H	I				
125tV	A	B	C	D	E	F	G	H					
126kV	A	B	C	D	E	F							
126pV		B	C	D	E	F	G	H	I	J	K	L	M
126tV		B	C	D	E	F	G	H	I	J	K	L	M
138kV	A	B	C	D	E	F	G	H	I				
138pV	A	B	C	D									
138tV	A	B	C	D	E	F	G	H					
142kV	A	B	C	D	E	F	G	H					
142pV	A	B	C	D	E	F							
142tV	A	B	C	D	E	F	G						
143kV			C	D	E	F	G	H	I	J	K	L	M
143pV	A	B	C	D	E	F	G	H	I	J			
143tV	A	B	C	D	E	F	G	H	I	J	K	L	
148kV								H	I	J	K	L	M
148pV				D	E	F	G	H	I	J	K	L	M
148tV								H	I	J	K	L	M
150kV	A	B	C	D	E	F	G	H	I	J			
150pV	A												
150tV	A	B	C										
170kV	A	B	C	D	E	F	G	H					
170pV	A	B	C	D	E	F	G	H					
170tV	A	B	C	D	E	F	G	H	I	J	K	L	
173kV						F	G	H	I	J	K	L	M
173pV												L	M
173tV			C	D	E	F	G	H	I	J	K	L	M
80kV									I	J	K	L	M
80pV											K	L	M
80tV													M
84kV					E	F	G	H	I	J	K	L	M
84pV	A	B	C	D									
84tV	A	B	C	D	E	F	G	H					
97kV		B	C	D	E	F	G	H	I	J	K	L	M
97pV	A	B											
97tV	A	B	C	D	E								

Figura 15: Códigos definidores do perfil dos falantes

A observação dos resultados de catorze falantes revela o que seria expectável: a existência de falantes que se assemelham e de outros que são totalmente diferentes no que respeita à duração do PreVoz, tal como os indivíduos. O que nos parece mais interessante nos valores concretos obtidos na análise deste parâmetro é que os valores de cada falante não se sobrepõem nunca integralmente aos de outro, por razões que se prendem com a estrutura do seu aparelho vocal ou com os seus hábitos de fonação. Tal como as pessoas podem dar, por exemplo, passos com a mesma exacta medida, também podemos prever que existam pessoas diferentes a produzir valores idênticos do PreVoz. Apesar disso, como nas perícias forenses habitualmente não se compara um falante com todos os falantes existentes, mas só com um leque restrito de suspeitos, a utilização destes valores como referência pode ser útil no traçado do perfil vocal do falante não identificado.

Estes resultados são tanto mais expressivos quanto, mesmo em situações de sobreposição de dois falantes numa mesma área preferencial para a realização do PreVoz, eles se revelam distintivos pelo desvio padrão apresentado para cada uma das consoantes e por cada um dos falantes. O estudo deste parâmetro é portanto promissor, necessitando de ser ainda prosseguido em mais falantes, tendo como base outras sequências CV e mesmo sequências consonânticas com a 2ª consoante vozeada.

Na secção seguinte aplicamo-lo às gravações do caso forense acima descrito com o intuito de mostrar a sua capacidade para aclarar o valor obtido na perícia realizada só com base nos parâmetros standardizados.

4. Aplicação ao caso real C_2002

Fazemos aqui a comparação entre os falantes das gravações K e X, utilizando a mesma metodologia utilizada na análise dos dados do CPE-Var. Salientamos que foi possível colher na gravação telefónica (X) e na gravação de controlo (K) os valores do PreVoz – o que deve ser tido em conta, dada a restrição da banda telefónica. O valor deste parâmetro pode ser recolhido mesmo acima dos 300Hz e não foi particularmente afectado pela existência de ruído.

Os resultados apontam para a confirmação da perícia inicial, ou seja, para a conclusão de que não há razões para dizer que as gravações K e X correspondam a falantes distintos. Tanto K como X apresentam variabilidade próxima (K: DP = 0.45ms; X: DP = 0.29ms) e espaços de duração do PreVoz semelhantes: K: valor médio = 6.2ms; X: 6.3ms. As figuras 16 e 17 sistematizam os resultados com intervalos de confiança de 95%.

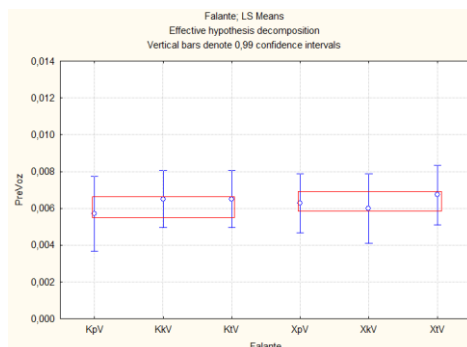


Figura 16: Valores médios de PreVoz nas Gravações K e X

Tukey HSD test; variable PreVoz (Caso forense)				
Homogenous Groups, alpha = .01000				
Error: Between MS = .00001, df = 63.000				
Cell No.	Falante	PreVoz Mean	1	
5	XkV	0.004000	****	
6	XtV	0.005286	****	
1	KpV	0.005714	****	
4	XpV	0.005750	****	
3	KtV	0.006500	****	
2	KkV	0.006500	****	

Figura 17: Teste de Tukey aplicado aos valores de PreVoz nas gravações X e K

Os resultados da figura 16 mostram a existência de valores médios um pouco diferentes nas duas gravações nas três consoantes. Uma vez que neste caso o suspeito confessou ser autor das gravações K e X, sabemos que essa diferença mínima só indica que cada pessoa pode oscilar ligeiramente dentro de uma gama preferencial de valores para este parâmetro (como para todos os outros). Vemos ainda que, apesar de o limite de tempo entre gravações ter sido superior a seis anos impedindo a conclusão de *identificação*, o falante mantém valores semelhantes do PreVoz, o que nos leva a equacionar a hipótese de testar gravações de outros falantes com intervalos superiores, para verificar se apresentam valores diferentes do PreVoz. Aqui verificou-se que esse prazo previsto no modelo ABRE-IAFPA não tem consequências claras.

Os valores de PreVoz nas gravações K e X permitem constituir um grupo homogêneo (figura 17), mostrando que a inclusão deste parâmetro na análise teria beneficiado as conclusões da perícia. Dessa forma, a força da prova pericial em tribunal teria sido maior e ter-se-ia podido, pelo menos, atingir o nível de *provável identificação*, esse sim próximo da *identificação* (unicamente inatingível neste caso devido ao intervalo temporal existente entre as gravações).

Saliente-se que propor a inclusão do PreVoz nas perícias forenses implica que estas continuem a utilizar os mesmos procedimentos habituais, porque os valores de um parâmetro isolado não têm força de prova suficiente. Os seus resultados devem ser aliados sempre aos dos restantes parâmetros na definição do perfil do falante, para que a perícia seja credível.

5. Conclusões

Neste artigo, foram descritas as características da escala e da metodologia habitualmente usadas nas perícias forenses. Perante as dificuldades em se atingir resultados periciais totalmente fiáveis nas

instituições judiciais, explorámos a eficácia discriminativa do novo parâmetro PreVoz, de base quantitativa e ainda inexistente na literatura.

Uma vez que a cadeia falada engloba, além das informações linguísticas, as características específicas de cada falante, cremos que a análise exaustiva de mais parâmetros poderá reforçar o peso relativo dos resultados das perícias em Fonética Forense face aos restantes meios de prova.

Testámos o PreVoz em dois tipos de material: gravações de fala espontânea em entrevista sociolinguística do CPE-Var e gravações do Caso C-2002, anteriormente periciado com base nas técnicas standard com um desfecho não satisfatório em ambiente judicial. Foram analisadas sequências CV, em que a consoante é uma oclusiva não vozeada.

A análise dos dados dos catorze falantes do CPE-Var permitiu-nos concluir o seguinte:

- (i) a duração do PreVoz oscila entre 0 e 0.017s, o que o torna não manipulável conscientemente e resistente ao disfarce e à imitação;
- (ii) o PreVoz não se associa à qualidade da Vogal, pelo que pode ser usado na perícia mesmo com escassez de ocorrências analisáveis;
- (iii) o PreVoz apresenta valores potencialmente distintos para cada consoante em cada falante, embora não se possa dizer que isso se deva ao ponto de articulação da consoante;
- (iv) o PreVoz não varia, nos dados analisados, em função da faixa etária do falante adulto, nomeadamente as faixas etárias dos 25-39 anos e a dos 50-65 anos;
- (v) o PreVoz tem um elevado poder discriminativo tanto em comparações envolvendo dois falantes (o processo típico da análise forense) como em comparações com um número maior de falantes.

Aplicámos a metodologia ao Caso C_2002 e pudemos observar que nas gravações X e K os valores de PreVoz são consistentes com o mesmo perfil, o que está de acordo com a confissão de autoria do suspeito, efectuada após a perícia standard. Pudemos ver também que o PreVoz pode ser medido mesmo com as limitações da banda telefónica.

Os dados dos catorze falantes acima apresentados, em conjugação com os do Caso C-2002, permitem confirmar *H1 = Os valores de duração de pré-vozeamento dependem do falante, podendo fazer parte do seu perfil vocal*. De facto, a duração do PreVoz associa-se às características individuais de cada falante, o que sucederá não apenas em termos fisiológicos, mas também em termos de hábitos de fonação e confortabilidade da produção.

Em face destes resultados, propomos a inclusão do parâmetro testado nas análises acústicas forenses porque:

- (i) ele apresenta um nível de significância muito elevado;
- (ii) pode ser usado com qualquer número de amostras;
- (iii) pode ser obtido em gravações de duração relativamente curta, visto que basta que existam consoantes não vozeadas seguidas de vogal (o que é muito frequente na língua);
- (iv) não é possível ter consciência da sua existência e, por isso, não é imitável nem passível de disfarce;
- (v) está disponível nas amostras analisadas mesmo na banda limitada do telefone;
- (vi) não se mostrou consideravelmente afectado pelo ruído externo.

6. Desenvolvimento da pesquisa

No estudo subsequente deste parâmetro, assumirá prioridade a investigação de possíveis variações dos seus valores em gravações de um mesmo falante efectuadas com intervalos temporais superiores a seis anos. Essa questão coloca-se para satisfazer um dos critérios que a escala de conclusões utilizada nas perícias forenses impõe para que seja atingido o nível de Identificação.

Outros aspectos a explorar são:

- (i) a verificação de possíveis diferenças nos valores de PreVoz se a sequência contiver uma fricativa não vozeada, em vez da oclusiva;
- (ii) o mesmo tipo de abordagem a sequências consonânticas com C não vozeada *versus* C vozeada;

- (iii) os índices de variação da intensidade ao longo do PreVoz;
- (iv) a variação entre outras faixas etárias na realização do PreVoz;
- (v) aplicação do parâmetro a produções de falantes do género feminino;
- (vi) aplicação do parâmetro a um maior número de casos forenses.

Referências

- American Board of Recorded Evidence: <http://www.abreboard.us/> (consultado a 07/04/2014).
- Boersma, Paul & David Weenink (2013). *Praat: doing phonetics by computer*. Versão 5.3.52, obtida a 23/06/2013 em <http://www.praat.org/>.
- Byrne, Catherine & Paul Foulkes (2004) The mobile phone effect on vowel formants. *International Journal of Speech, Language and the Law* XI (1), pp. 83-102.
- EXMARaLDA - Partitur Editor 1.5.2 (Thomas Schmidt - University of Hamburg): <http://www.exmaralda.org>.
- IAFPA - International Association for Forensic Phonetics and Acoustics: <http://www.iafpa.net/> (consultado a 07/04/2014).
- Ladefoged, Peter & Keith Johnson (2011) *A Course in phonetics*, 6.^a edição. Boston: Cengage Learning.
- Lin, Chi-Yueh & Hsiao-Chuan Wang (2011) Automatic estimation of voice-onset time for word-initial stops by applying random forest to onset detection. *The Journal of the Acoustical Society of America* 130, pp. 514-525.
- Müller, Christian (2007) *Speaker Classification II*. Berkeley: Springer.
- Nolan, Francis (1983) *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.
- Rodrigues, Celeste & Ernesto d'Andrade (1998) CPE-Var - Corpus de Português Europeu-Variação. In *Actas do 14.º Encontro Nacional da Associação Portuguesa de Linguística* (Aveiro), II. Braga: Associação Portuguesa de Linguística, pp. 627-629.
- Rodrigues, Celeste (2003) *Lisboa e Braga: fonologia e variação*. Lisboa: Fundação para a Ciência e a Tecnologia / Fundação Calouste Gulbenkian.
- Rose, Philip (2002) *Forensic speaker identification*. Londres: Taylor & Francis.
- Relatório Anual de Segurança Interna (RASI) 2012* (2013). Sistema de Segurança Interna: <http://www.portugal.gov.pt/pt/os-ministerios/ministerio-da-administracao-interna/documentos-oficiais/20130327-rasi-2012.aspx> (consulta: 30/03/2014).
- Tosi, Oscar (1979) *Voice identification: theory and legal applications*. Baltimore: University Park Press.
- Zhang, Cuiling & Tiejun Tan (2008) Voice disguise and automatic speaker recognition. *Forensic Science International* 175(2), pp.118-122.