USO DE TÉCNICAS ACÚSTICAS PARA VERIFICAÇÃO DE LOCUTOR EM SIMULAÇÃO EXPERIMENTAL

Aline de Paula MACHADO
Professor Responsável: Plínio Almeida Barbosa

RESUMO: Este projeto propõe, através de algumas técnicas de análise acústica, o reconhecimento de um indivíduo dentro de um grupo de cinco falantes do português paulista e assinalar quais parâmetros acústicos são relevantes para o reconhecimento naquele grupo.
As análises dos quatro primeiros formantes das vogais orais, da freqüência fundamental, da duração de unidades do tamanho da sílaba e da vogal e da intensidade de trechos escolhidos desses falantes servirão para identificar um indivíduo dentro daquele grupo. Todos os trechos escolhidos são de entrevistados em ambiente não tratado acusticamente. Além disso, trechos escolhidos em sala com tratamento acústico de um dos falantes (o 'criminoso') simularão o padrão questionado da situação forense.
Palavras-chave: fonética, identificação de locutor, simulação experimental

INTRODUÇÃO E JUSTIFICATIVA

A ciência forense encontra em uma área da Lingüística, a fonética, uma possibilidade de conciliar a tecnologia para o reconhecimento de voz de indivíduos, com o conhecimento fonético-acústico. Partindo de princípios da Fonética Acústica, será feita uma simulação de "crime" nos moldes das situações em Fonética Forense.

Há algum tempo, o *Grupo de Estudos de Prosódia da Fala*, onde a inicianda faz a Iniciação, dispõe de gravações de entrevistas com um grupo de cinco falantes paulistas (que agora denominaremos "suspeitos") feita o gravador digital Sony ICD-P630F (taxa de amostragem: 16 kHZ) ao ar livre. Entre esses falantes, sorteou-se, sem que a inicianda soubesse quem, um falante que é quem se deve identificar (que denominaremos "criminoso").

Além disso, este "criminoso" fez uma narrativa em ambiente tratado acusticamente, que será a base para sua verificação em comparação às demais gravações. Mais detalhes a esse respeito serão fornecidos na seção de Metodologia.

Na área de reconhecimento de fala, dois termos são usados para a pesquisa com a fala dos indivíduos, são eles: identificação de locutor (*Speaker Identification*) e verificação de locutor (*Speaker Verification*). Eles podem ser substituídos por "reconhecimento de locutor" (*Speaker Recognition*), que é uma maneira genérica de dizer que tanto a *SPID* (*Speaker* Identification) quanto a *SV* (*Speaker Verification*) têm o objetivo de reconhecer uma pessoa no discurso. Porém, diferem em relação aos meios desse reconhecimento.

A *SPID* identifica um indivíduo através da análise de sua fala em ambientes que podem ter distorções e ruídos. A acústica desse ambiente influencia na precisão da identificação daquele que deve 'descobrir'. O criminoso em potencial é totalmente desconhecido, podendo compreender uma situação em que se tenha milhares de suspeitos (quando não há pistas, os falantes de uma comunidade são "suspeitos" em potencial).

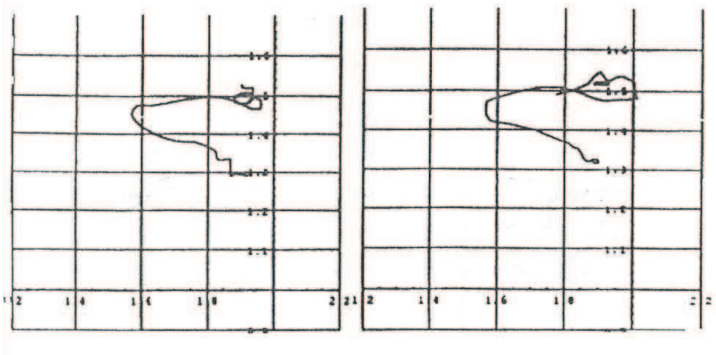
A SV busca a identificação do falante através de um ambiente acústico favorável, é utilizada basicamente para chaves de segurança em estações de energia nuclear, instalações militares, laboratórios de pesquisa e centros computacionais, segurança em geral.

Embora a Fonética Forense seja associada à tarefa de identificação de locutor, ou seja, a identificação de uma única pessoa em uma população (reconhecimento indireto de um sujeito), na prática acaba sendo verificação, pois trabalha no fim, com um número finito de sujeitos.

Uma das principais diferenças entre a Fonética Forense e o reconhecimento de locutor (SR, speaker recognition) comercial, segundo Künzel (1994), é a questão da cooperação do sujeito. No primeiro caso o falante não quer ser identificado, portanto usa de técnicas como o disfarce em sua gravação. A segunda lida com um ambiente acústico propício para identificação, o sujeito se deixa identificar.

Na Fonética Forense não há uma fala prévia conhecida para o sujeito, um texto, por exemplo, que será enunciado com intuito de comparar cada palavra com uma gravação anterior. Um principal ponto que 'favorece' o SR comercial, por outro lado, é justamente o número limitado de amostras, falantes, palavras a serem examinadas. Há basicamente três abordagens de Reconhecimento de Falante: reconhecimento auditivo, verificação visual de espectrogramas de banda larga e sistemas de SR semi-automáticos e por computador. Essa interpretação visual de espectrograma foi uma técnica muito usada nos Estados Unidos, países da Europa, Israel e demais países durante os anos sessenta e setenta, estendida até hoje. (Künzel, 1994).

A utilização apenas de espectrogramas é inconclusiva e muito controversa para análises de verificação de voz, como foi publicado pelo *Comitê de avaliação de espectogramas sonoros* e ratificada por R.J. Bolt, F.S. Cooper, D.M. Green (1979); Doddington, G.R. (1988); Hollie, H. (1974); B.E. Koenig, D. S. Ritenour, B. A. Kohus; A. S. Kelly (1987) e T. Shipp, E.T. Doherty, H. Hollien (1987). A razão disso é que as variações intrafalantes nos espectrogramas de voz não podem ser mais drásticos do que entre falantes. Por exemplo, em Holien (2002, p. 143) examine-se a Figura 1:



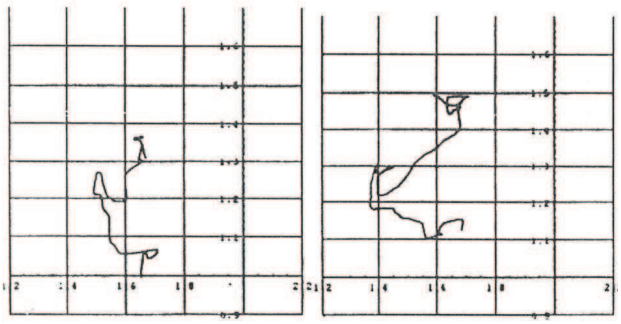


Fig. 1. Comparação de voz entre três falantes pronunciando a palavra /owie/
(estraído de Hollie, 2002, p. 143. Fig. 7.1)

Os blocos da esquerda são de um falante X produzindo a mesma palavra /owie/ duas vezes; os da direita dois falantes diferentes pronunciando a mesma palavra. Como podemos observar, as trajetórias superiores à esquerda e à direita são mais similares que as demais trajetórias, a do mesmo sujeito à esquerda e de outro indivíduo à direita inferior.

Esse tipo de variação nos leva a concluir que um conhecimento estatístico deve ser usado como ferramenta de validação/análise de dados, como usado nos experimentos de: Maturi (1990), Mella (1994), R.N. Monsen, A.M. Engebretson (1983), entre outros. A análise estatística é crucial para experimentos que apresentam variações de Freqüência Fundamental, por exemplo, de avaliação de deficiências de pronúncia e da fala.

Este trabalho se baseia na verificação do locutor. Esta área difere da identificação pela forma de como a busca pelo "criminoso" é realizada. Basicamente, "o falante quer ser reconhecido" (HOLLIEN, 2002). Ou seja, é uma forma de reconhecer alguém através de ambientes acústicos totalmente favoráveis para esta identificação (como áreas de segurança ativadas por comando de voz, etc.). Os suspeitos de algum crime são levados em laboratório para gravação de material padrão, uma amostra de sua fala. Essas amostras, portanto são de boa qualidade e, através de equipamentos e experimentos eficazes, pode-se chegar a um resultado relativamente positivo, dependendo da qualidade do material questionado (a amostra em situação real).

Para uma boa análise, a gravação questionada deve ter qualidade mínima para análise, pois, dependendo do filtro, pelo qual a gravação passa, a obtenção de formantes das vogais e frequência fundamental pode estar comprometida (isso ocorre muito em gravações telefônicas).

Em seu experimento, Maturi (1990), relata a importância de uma boa gravação.

A banda de passagem do filtro (telefônico) do material questionado prejudicou informações linguisticamente pertinentes como o primeiro harmônico, primeiro formante de vogais fechadas e o ruído de fricativas.

As características acústicas que serão extraídas para a verificação de locutor são: quatro primeiros formantes (de vogais orais), frequência fundamental, duração de unidades do tamanho da sílaba e da vogal e ênfase espectral (medida de identidade relativa de bandas).

Um formante é um modo de vibração ressoante em um tubo acústico que assinala maximização local de energia sonora. Para a Fonética Lingüística o mais importante a ser estudado são os três primeiros formantes para a percepção de uma vogal, incluindo-se o valor e a largura da banda (Carlson & Granström, 1979 apud MONSEN; ENGEBRETSON, 1983; Carlson, Granström, & Klatt, 1979 apud MONSEN; ENGEBRETSON, 1983) e a intensidade

dos mesmos. Optou-se pela verificação dos quatro primeiros formantes nessa iniciação para uma pesquisa mais eficaz, podendo alguns traços pertinentes das vogais, sobretudo os relativos às características próprias de um indivíduo, não serem achados nos três primeiros formantes.

Em seu artigo, Mella (1994) estuda a eficiência dos três primeiros formantes de diferentes vogais do francês para caracterização de locutor e qual fator é favorável para identificar um falante. Para o formante F1, as vogais do francês /a/, /œ/ e /ɔ/ são relevantes para a identificação de falante. O formante F2, das vogais /i/ e /e/ apresentam grande relevância por sua vez, afilicado à cavidade anterior. O F3 está relacionado à labialização, aumenta a sua freqüência em contexto uvular e durante as vogais arredondadas. Porém, as questões levantadas sobre este formante funcionam apenas para o francês.

A frequência fundamental é a mais baixa freqüência de uma série harmônica de um som. Ela é modificada também pela entoação, emoção, disfarce do indivíduo, entre outros. Nesse trabalho, descritores estatísticos da mesma serão usados em trechos selecionados para auxílio à verificação de locutor.

A duração é uma medida do tempo transcorrido entre dois eventos singulares, como os segmentos acústicos dos sons da fala, como sílabas, entre outros. No português brasileiro, um fator interessante a ser observado sobre a duração das vogais é que quanto mais alta (ou fechada) é a vogal, menor é a duração. A duração de unidades do tamanho da sílaba é o parâmetro principal que define o ritmo da fala, o acento, as ênfases (essa última juntamente com a freqüência fundamental).

A intensidade de um som é proporcional ao quadrado de sua amplitude e é tanto maior quanto maior o esforço ou a força ilocutória. Trechos enfáticos têm maior intensidade. Para evitar o problema de sua variação com a posição do microfone, pode ser medida através de diferenças entre bandas do escpetro. (Eriksson, 2005).

A análise LPC (*Linear Predictive Coding*) é uma técnica de codificação para reduzir o número de amostras que representam o sinal e que permite obter, para vogais orais, valores mais precisos que os espectrográficos para frequência dos formantes. Essa também será usada para esse fim. Todos os parâmetros elencados serão extraídos no Praat (cf. seção Material e Métodos).

Künzel (Ibid.) prevê que o número de sistemas de reconhecimento de locutor para uso forense crescerá constantemente e que haverá um refinamento de parâmetros e técnicas de uso. A qualificação de um perito em forense é necessária na ciência da fala, como foneticista, entre outros profissionais que usam a fala como instrumento de pesquisa.

O mesmo autor afirma também que “no domínio forense nós estamos ainda muito longe de construir um vetor característico poderoso o bastante para ser usado como modelo para tarefas de identificação”, sem contar algoritmos para avaliações estatísticas.

Este projeto proporciona aplicações nas áreas de verificação de locutor (chaves vocais, dispositivos de segurança vocais), assim como aplicações na área forense.

Também há repercussões na área de fonoestilística (o que difere de uma pessoa em relação a outra). Não se pretende aprender técnicas novas, mas realizar a primeira incursão de um graduando na área de reconhecimento de locutor, proporcionando um aprendizado de procedimentos de análise acústica e estatística específicos à área.

OBJETIVOS

Este trabalhona área de verificação de locutor é feita no intuito de usar procedimentos experimentais procedimentos experimentais que possam ser automatizados e servir como início à área da Fonética Forense. Também há o interesse no aprendizado aprofundado de técnicas de análise acústica, entre elas análise de formantes, frequência fundamental, duração e intensidade, que vão servir também para iniciação na área de fonoestilística visto que nos concentramos na variação inter-individual.

Além disso, pretendemos apontar quais características das vogais estudadas e dos parâmetros de F0, duração e ênfase espectral, podem ser relevantes para o reconhecimento de um falante.

METODOLOGIA

As gravações foram coletadas por uma colaboradora, a aluna Maíra Bueno Braga, que entrevistou um grupo de cinco falantes ("suspeitos") entre os quais se encontra um "criminoso". Este "criminoso" fez uma pequena narrativa, em estúdio, que será a base para sua verificação em comparação às demais gravações.

O programa PRAAT (Boersma e Weenink, 2010) será utilizado para a análise dos formantes das vogais, duração de vogais e unidades do tamanho da sílaba, frequência fundamental e intensidade.

ANÁLISE DE PARÂMETROS ACÚSTICOS E RESULTADOS

Na primeira parte do trabalho, tínhamos etiquetado as quatro gravações do *corpus* colhidas em ambiente aberto (apenas as vogais orais foram segmentadas, por resistirem mais ao ruído do que as consoantes, e serem menos variáveis do que as vogais nasalizadas) seguindo o modelo a seguir:

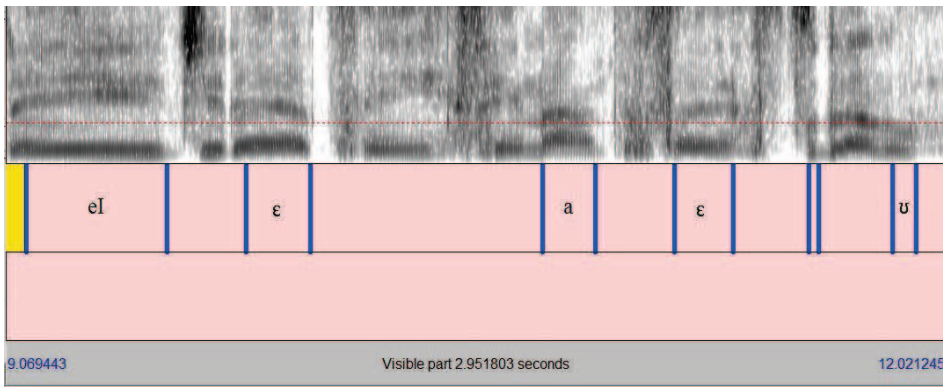


Fig. 1. Exemplo de segmentação no Praat, mostrando a separação das vogais orais da gravação em símbolos do IPA. O Sujeito 1 fala "*Entrei direto sem nada, só estudando.*".

Para etiquetar as vogais no Praat seguimos o procedimento clássico na área, de se guiar pelo aparecimento dos dois primeiros formantes para o início da vogal, priorizando F2 em caso de conflito. Para o final da vogal, simetricamente, pelo esmaecimento dos dois primeiros formantes. A seguir, uma tabela contendo as informações dos dados analisados (duração da gravação, número de etiquetas e total).

	Duração da gravação (min)	Número de etiquetas
Sujeito 1	2:22	216
Sujeito 2	1:34	106
Sujeito 3	1:20	65
Sujeito 4	1:42	161
Criminoso	3:20	574
Total		1122

Com todos os dados já segmentados, incluindo os da gravação do "criminoso" em laboratório (de ambiente acusticamente favorável), os dados estatísticos começaram a ser contabilizados.

Os modelos que empregam técnicas estatísticas são baseados na distribuição de várias características pertencentes à fala do suspeito e em comparação a distribuição das mesmas características na população de referência com respeito à gravação questionada.

A variação que ocorre inter-falantes geralmente é tão mínima quanto as intrafalantes (Hollien, 2002), por isso um conhecimento estatístico deve ser usado como ferramenta de validação/análise de dados, como usado nos experimentos de: Maturi (1990), Mella (1994), R.N. Monsen, A.M. Engebretson (1983), entre outros. A análise estatística é crucial para experimentos que apresentam variações de frequência fundamental, por exemplo, de avaliação de variações na pronúncia e na fala.

Os dados desta pesquisa foram analisados através do programa R, tendo empregado um teste t de variáveis independentes com nível de significância igual a 5% para sempre comparar os dados de cada suspeito com o "criminoso". Este teste é importante para calcular a probabilidade de erro na rejeição da hipótese nula caso essa seja verdadeira (o suspeito x é o criminoso).

O cálculo da discrepância entre os parâmetros acústios dos sujeitos e criminoso sera analisado pelo programa R. O resultado de cada análise permite rejeitarmos ou aceitarmos a hipótese nula de que os parâmetros de suspeito e criminoso pertencem ao mesmo sujeito, para o nível de significância de 5%.

Após uma minuciosa análise de cada parâmetro e da realização do teste t, obtiveram-se os seguitnes resultados. Nas tabelas abaixo "formantes" e "duração de vogal" indica comparações para as mesmas vogais no criminoso e no suspeito (sujeitos 1 a 4). Baseline e ênfase espectral foram analisados independentemente da vogal, porque são medidas prosódicas.

Ao escutar cada gravação, inclinei-me para o sujeito 3, possui ritmo de fala e percepção da melodia parecidos. Os valores de ênfase espectral e baseline são menos confiáveis pois o estilo de elocução podem diferir bastante em um mesmo sujeito.

O “criminoso” não era nenhum dos suspeitos gravados. Por isso, com uma discrepância mínima entre os sujeitos 1 e 3, seria necessário utilizar um outro parâmetro para definir com uma maior precisão as diferenças dos suspeitos para com o “criminoso”.

CONSIDERAÇÕES FINAIS

Com base nos artigos de Gay (1968) e Goldstein (1975), podemos entender a grande importância de informações dos formantes no reconhecimento de locutor não só nas frequências dos mesmos mas também no resto de sua estrutura não explorada. Com isso, o prolongamento da pesquisa (que já está sendo providenciado o início de bolsa PIBIC no mês de Setembro e também espera do pedido de aceitação do prolongamento da bolsa FAPESP) seria essencial para o estudo da estrutura de formantes das vogais, dando uma maior credibilidade ao resultado desta pesquisa, mesmo porque, fui informada após a redação desse relatório por meu orientador, o criminoso não é nenhum dos suspeitos. Portanto, com esses resultados já obtidos sera necessário acrescenter mais alguns parâmetros acústicos como movimentos de formantes, seguindo a teoria de Goldstein (1975) de que é necessário analisar a estrutura e não só as frequências de umformante para um resultado preciso. Também acrescentarei de taxa de elocução entre os parâmetros.

REFERÊNCIAS

HOLLIEN, H. **Forensic Voice Identification**. London: Academic Press. 2002.

PRAAT. Boersma e Weenihk, 2010. http://www.fon.hum.uva.nl/praat/R Project. R Development Core Team, 2011. http://www.r-project.org/.

BIBLIOGRAFIA

BOLT, COOPER, GREEN, et al.: **On the Theory and Practice of Voice Identification.** Washington DC: National Academy of Sciences, 1979.

BOTTI, ALEXANDER, DRYGAJLO**. An Interpretation Framework for the** Evaluation of Evidence in Forensic: Automatic Speaker Recognition with Limited **Suspect Data**. 2004.

DODDINGTON, G. R. **Speaker Recognition – Identifying People by Their Voices, IEEE-ASSP-Transactions.** 73:1651, 1985.

ERIKSSON, A. (2005). **Tutorial on Forensic Speech Science.**

FARRÚS, WAGNER, ANGUITA, HERNANDO. **How vulnerable are prosodic features to professional imitators?** 2008.

GAY, T. **Effect of Speaking Rate on Diphtong Formant Movements.** 1968.

GOLDSTEIN, U. **Speaker-identifying features based on formant tracks.** 2011.

HOLLIEN, H. **Forensic Voice Identification.** London: Academic Press, 2002.
_____.Peculiar case of "voiceprints". **JASA** 56: 210, 1974.

KINOSHITA, ISHIHARA, ROSE. **Beyond the Long-term Mean: Exploring the** Potential of F0: Distribution Parameters in Traditional Forensic Speaker **Recognition.** 2008.

KOENIG, RITENOUR, KOHUS, KELLY. **Reply to "Some fundamental considerations regarding voice identification".** *JASA* 82:688, 1987.

KOENIG, B. **Spectrographic voice identification: A forensic survey.** *JASA* 79:2088, 1986.

KÜNZEL, H. **Current Approaches to Forensic Speaker Recognition.** 1994.

MATURI, P. **Speaker Identification in Forensics: A simulation experiment.** 1994.

MELLA, O. **Extraction of formants of oral vowels and critical analysis for speaker characterization.** 1994.

MEUWLY, D. **Forensic speaker recognition: An evidence odyssey – Summary**. 2004.

MONSEN, ENGEBRETSON. The Accuracy of Formant Frequency Measurements: A Comparison of Spectrographic Analysis and Linear Prediction. In: **Journal of Speech and Hearing Research**. 1983.

NOLAN, F. Speaker Recognition and Forensic Phonetics. In: W. Hardcastle and J. Laver (eds), A Handbook of Phonetic Science. Oxford: Blackwell. 1997.

PELECANOS, CHAUDHARI & ROMASWANY. **Compensation of utterance length for speaker verification.** 2004.

ROSE, P. **Technical Forensic Speaker Identification from a Bayesian Linguist's Perspective.** 2004.

SHIPP, DOHERTY, HOLLIEN. **Some fundamental considerations regarding voice identification**. *JASA* 82:687, 1987.

SOLEWICZ, Y. **Noise Robustness in Forensic Speaker Verification.** 2001.

WATT, D. The Identification of the Individual Through Speech. In: WATT, D.; LLAMAS, C. (Eds.) **Language and Identities**. Edinburgo: Edinburgh University Press, 2010, p. 76-85.