

TUTORIAL SOBRE FONÉTICA FORENSE¹

Anders Eriksson²

anders.eriksson@ling.su.se

RESUMO: Este tutorial apresenta a fonética forense como área de atuação prática e como campo de pesquisa para cientistas da fala. O texto está dividido em três partes. Na primeira, a fonética forense é apresentada em seu contexto histórico e em seu desenvolvimento moderno. Na segunda, são apresentados temas e questões fundamentais de pesquisa na área e como os conhecimentos gerados no campo podem ser aplicados em situações práticas. São discutidos assuntos como o reconhecimento e a discriminação da voz humana, bem como de que forma memória, familiaridade, língua, disfarce vocal, entre outros fatores, afetam essas habilidades. Na terceira, discute-se o uso de informações presentes na voz e em imagens do cérebro para a detecção de mentira e tenta-se traçar uma linha divisória entre métodos cuja aplicação na prática forense pode ser justificada com base em provas científicas.

Palavras-chave: Fonética forense; Ciência da fala forense.

INTRODUÇÃO

Este tutorial vai tratar de três áreas.

Primeiramente, vamos posicionar a ciência da fala forense em seu contexto histórico. Descreveremos brevemente seu desenvolvimento, o de suas origens modernas com as impressões vocais (*voiceprints*)³ nos Estados Unidos da América e com a fonoscopia (*phonoscopy*) na União Soviética de Stalin, até a presente situação, dando uma noção dos avanços e também das controvérsias.

¹ Publicado originalmente em: ERIKSSON, Anders. Tutorial on forensic speech science. Part I: Forensic phonetics. In: *Interspeech 2005 - Eurospeech 2005*. Proceedings of the 9th European conference on speech communication and technology. Lisbon, Portugal. September 4-8, 2005.

² Departamento de Linguística, Universidade de Estocolmo, Estocolmo, Suécia. O texto serviu de base para uma sessão tutorial na *9th European Conference on Speech Communication and Technology* (Interspeech, 2005), realizada em Lisboa, em setembro de 2005. Tradução de Pablo Arantes (Departamento de Letras, Universidade Federal de São Carlos) e Suska Gutzeit (Universidade Federal de São Carlos).

³ Nos casos em que não existe uma tradução bem estabelecida para palavras ou expressões técnicas presentes do texto original, sugerimos uma tradução, mas deixamos o termo em inglês entre parênteses (N.T.).

Em seguida, vamos apresentar algumas questões fundamentais na pesquisa fonética forense e ver como o nosso conhecimento atual pode ser aplicado em situações práticas. Nessa parte do tutorial, o foco será o reconhecimento e a discriminação da voz humana e como memória, familiaridade, língua, disfarce vocal, entre outros fatores, influenciam nessas duas habilidades.

Mentiras e fraudes são problemas históricos em investigações forenses. Assim, não é surpresa que muitas pessoas vêm tentando encontrar caminhos para detectar sinais de insinceridade na fala. Vamos descrever tentativas para usar as informações transmitidas pela voz humana na detecção de mentiras e recentes experimentos usando métodos de ressonância magnética funcional (fMRI) para lidar com o problema. Na descrição desses trabalhos, vamos fazer também uma tentativa de traçar uma linha divisória entre métodos cuja aplicação, na prática forense, pode ser justificada com base em provas científicas e aqueles que são, na melhor das hipóteses, especulativos e abertamente fraudulentos no pior dos casos.

Para cada tópico, vamos apresentar uma lista de leituras complementares.

1. LEITURAS DE BASE

Muitos livros sobre fonética forense foram publicados durante as últimas décadas, muitos dos quais podem ser encontrados em bibliotecas universitárias. Mencionaremos aqui dois livros recentes que proporcionam uma boa e abrangente introdução ao campo. *Forensic Voice Identification*, escrito por Hollien, é um livro introdutório que delineia os antecedentes históricos do campo e trata de tópicos como reconhecimento automático da fala, memória e procedimentos para a montagem de filas de reconhecimento de voz (*voice lineup*). O livro é razoavelmente acessível a não especialistas e não requer qualquer conhecimento fonético profundo. O livro de Rose, *Forensic Speaker Identification*, é consideravelmente mais técnico. Ele trata de assuntos como identificação automática de falantes e cobre com certa profundidade algumas das técnicas usadas no campo, como a análise cepstral, por exemplo. É uma excelente introdução ao campo dos métodos automáticos para o leitor que tem um domínio razoável de tecnologia da fala ou fonética acústica. O livro também trata de questões estatísticas e métodos usados para avaliar o resultado da comparação de falantes, como a estatística bayesiana. Ambos os livros são leituras altamente recomendadas.

2. ANTECEDENTES HISTÓRICOS

O uso da identificação de voz em casos criminais tem uma história mais longa do que se pode imaginar. Esse recurso provavelmente foi usado por milhares de anos para identificar suspeitos de terem cometido crimes que foram ouvidos, mas não vistos. Em tempos mais recentes, registram-se muitos casos do uso da identificação de falantes como prova em tribunais. Um caso antigo, muito conhecido, é o julgamento de William Hulet, em 1660. Hulet foi acusado de ter executado o Rei Charles I. Uma testemunha, Richard Gittens, afirmou que ouviu o executor, cuja face estava oculta, pedir o perdão do rei e que ele sabia que era Hulet “por causa da sua voz”. O júri considerou Hulet culpado de alta traição, e ele foi sentenciado à morte. Mas esse é também um dos primeiros casos conhecidos de identificação incorreta de um falante. Antes que a sentença de morte fosse executada, descobriu-se que o assassino verdadeiro foi o carrasco oficial, que mais tarde confessou, e Hulet, conseqüentemente, foi libertado. Os detalhes do caso Hulet são típicos de muitos casos mesmo hoje, quase 350 anos depois. Ouve-se o criminoso, mas ele não é visto. A testemunha pode se sentir segura de que a identificação esteja correta, mas – como este e muitos outros casos mostram – isso não é uma garantia. A voz escutada pode ser conhecida ou desconhecida. No caso Hulet, a testemunha pensou que a voz pertencia a uma pessoa conhecida por ela, mas isso foi obviamente um erro.

Mesmo sendo comum que parte das provas, em um caso, sejam depoimentos de uma testemunha auricular (que, ao mesmo tempo, pode ou não ser a vítima), na maioria dos casos, a voz do criminoso não pertence a alguém conhecido pela testemunha. Também é muito comum que, entre o crime e uma tentativa posterior de decidir se a voz do suspeito é a mesma que a do criminoso, tenha passado um período de semanas, ou mais tempo. Quando esse é o caso, a acuidade da memória da vítima em relação à voz se torna uma questão crucial. Uma questão importante, neste contexto, é como a memória de uma voz deteriora-se ao longo do tempo. A primeira tentativa de responder a esta questão (ver seção 4.2) foi inspirada por um testemunho no caso Lindbergh. O filho do famoso aviador Charles Lindbergh foi sequestrado em primeiro de março de 1932. Uma carta de resgate foi encontrada no quarto do menino na qual o sequestrador exigia \$ 50.000. As negociações prosseguiram através de cartas e de anúncios em um jornal local. A família Lindbergh concordou em pagar o resgate e na noite de 02 de abril de 1932, Lindbergh levou seu

negociador, Condon, até um cemitério onde o dinheiro do resgate foi entregue. Lindbergh estava esperando em seu carro e conseguiu ouvir (mas não ver) o sequestrador chamando por Condon, dizendo: “Aqui, doutor. Por aqui! Por aqui!”. Cinco semanas depois, o menino foi encontrado morto. A polícia eventualmente rastreou um suspeito, Bruno Hauptmann, e prendeu-o. Em setembro de 1934, 29 meses depois de ouvir a voz no cemitério, Lindbergh foi disfarçado ao escritório do promotor e ouviu Hauptmann repetir o chamado ouvido no cemitério. Lindbergh afirmou ter reconhecido a voz como sendo a que ele havia ouvido na ocasião do pagamento do resgate. No julgamento, em janeiro de 1935, ele testemunhou sob juramento que havia reconhecido a voz de Hauptmann.

A invenção do telefone e dos equipamentos de gravação abriu novas áreas para a fonética forense. Ferramentas de análise capazes de fazer análise acústica da fala também foram desenvolvidas. Um marco nesta última área foi a invenção do espectrógrafo. A maior parte do desenvolvimento desse aparelho foi feita nos laboratórios da companhia telefônica Bell (*Bell Telephone Laboratories*) do final dos anos 1930 em diante. A construção do espectrógrafo desenvolvido lá foi baseada em ideias sugeridas por Steinberg (JASA, vol. 8, 1934, pp. 16-24). Um espectrógrafo basicamente do mesmo tipo foi produzido mais tarde pela empresa Kay Elementrics e vendido comercialmente com o nome comercial *Sonagraph*.

A motivação original por trás do desenvolvimento do espectrógrafo foi o estudo fonético da fala – “um método para o estudo e mensuração da produção da fala” (Steinberg). Também foi produzido um espectrógrafo que operava em tempo real, chamado *Direct Translator*, no qual o espectrograma era exibido em uma tela fluorescente. Imaginava-se que o equipamento fosse ser usado no treinamento da pronúncia de surdos e estudantes de língua estrangeira.

Como todos sabemos hoje, o espectrógrafo foi, e é, uma ferramenta muito valiosa para a pesquisa fonética. Sua utilidade como um recurso de auxílio aos surdos ou no ensino de língua estrangeira nunca alcançou as expectativas iniciais, mas essas expectativas ao menos geraram pesquisas relevantes. Assim, pode parecer surpreendente que quase nada tenha sido publicado durante os primeiros anos do desenvolvimento dessa ferramenta (final dos anos 1930 até 1945). A explicação para isso é o fato de que o projeto foi considerado um segredo de guerra. Em uma das primeiras publicações sobre o assunto lançadas depois da guerra, Potter (1945), que era pesquisador nos laboratórios Bell, escreveu: “O trabalho aqui descrito começou

antes da guerra. Por causa de interesses relacionados à guerra, ele foi considerado oficialmente como projeto de guerra e avançou o suficiente durante o período da guerra para justificar que ele fosse trazido à atenção do público”. É muito improvável que o exército norte-americano considerasse o desenvolvimento de uma ferramenta de auxílio aos surdos como um projeto de guerra. Não sabemos muito sobre o que exatamente eles esperavam conseguir com o projeto, mas especula-se que o que os militares realmente queriam era um método confiável para identificação de falantes. Meuwly (2003) reforça: “A participação dos Estados Unidos na Segunda Guerra Mundial deu origem a um projeto de aplicação militar do espectrógrafo sonoro: a identificação de navios inimigos por meio da voz de seus operadores de rádio”.

Essa afirmação parece razoável, mas, como dissemos, quase nada foi publicado sobre essa parte dos trabalhos de pesquisa. Vale notar, entretanto, que, quando dois pesquisadores dos laboratórios Bell, Grey e Kopp, publicaram um relatório interno, no final da guerra, eles usaram o termo impressão vocal para se referirem a espectrogramas com uma referência obviamente metafórica às impressões digitais, e o mesmo termo é usado em seu primeiro trabalho publicado (GREY; KOPP, 1944). Impressão vocal também é um termo que veio a ser usado em conexão com a controversa história da identificação de falantes através de espectrogramas que começaria alguns anos depois, iniciada por um ex-engenheiro dos laboratórios Bell, chamado Lawrence Kersta (ver seção 4.1).

Também é interessante notar que, quando os pesquisadores dos laboratórios Bell começaram a publicar novamente depois da guerra, poucas menções à identificação de falantes podem ser encontradas; ao invés, reapareceram tópicos anteriormente mencionados: “os surdos serão muito beneficiados... e estudantes de línguas estrangeiras... Isso oferecerá um meio objetivo de verificar conceitos fonéticos existentes e de expandir nosso conhecimento sobre a língua falada” (KOPP; GREEN, 1946). Deve-se notar também que os autores parecem enfatizar mais as similaridades entre os falantes do que diferenças individuais: “Padrões visíveis observados nas mesmas palavras e sentenças ditas por indivíduos diferentes mostram que as semelhanças na produção da fala são muito maiores do que as diferenças”. Potter (1945), ao mesmo tempo que reconhece diferenças individuais, também enfatiza as similaridades nos padrões exibidos por falantes diferentes.

Se as pessoas nos laboratórios Bell financiadas secretamente pelos militares trabalharam com “impressões vocais” para finalidades de identificação de falantes,

como temos boas razões para acreditar, então os primeiros momentos da história das “impressões vocais” seguem caminhos muito paralelos na União Soviética, incluindo o fato de que sabemos muito pouco sobre o assunto. Possivelmente o único relato dos esforços soviéticos que temos é o romance *O Primeiro Círculo*, de Soljenítsin. A trama do romance se desenrola em um espaço de tempo de apenas três dias durante as festas natalinas em 1949. O cenário é a prisão Marfino, nos arredores de Moscou, onde o regime stalinista mantinha cientistas rebeldes presos, entre os quais estava Soljenítsin. A prisão, que era dividida em várias seções com diferentes especialidades, abrigava em torno de trezentos prisioneiros. Era no *Setor Sete*, considerado o mais importante e prestigiado, onde acontecia o trabalho em diversos projetos ligados à tecnologia de fala. Além de seu próprio *Laboratório de Acústica*, o setor tinha outro, chamado *Laboratório de Fala Cifrada*, onde se buscava desenvolver meios eficientes de codificar a fala de modo a tornar sua decodificação difícil ou impossível pelo “inimigo”. Um dia, o foco mudou, pelo menos temporariamente, da “codificação” para o “reconhecimento” de voz. Uma chamada telefônica “criminosa” de um falante desconhecido que se acreditava estar trabalhando no Ministério das Relações Exteriores dirigida a um professor de medicina, advertindo-o contra compartilhar seus resultados de pesquisa com colegas estrangeiros, havia sido interceptada e gravada. Cinco pessoas no ministério eram os principais suspeitos, embora a polícia não soubesse qual deles tinha feito a chamada. Então, os cientistas do *Laboratório de Acústica* receberam uma fita contendo a chamada e gravações de amostras de voz de cinco suspeitos e receberam a tarefa de identificar um dos suspeitos como o autor da chamada. Eles receberam apenas dois dias para completarem a tarefa, com a deportação para a Sibéria como destino provável caso falhassem. A descrição da tecnologia e das ferramentas de análise feita por Soljenítsin não é detalhada o suficiente para tirar conclusões definitivas, mas é óbvio, pelo vocabulário usado (por exemplo, *voiceprints* e *vocoder*), que os cientistas estavam muito cientes de trabalho correspondente feito fora da União Soviética. Eles não usaram um espectrógrafo no sentido moderno da palavra, mas é provável que tenha sido algo similar ao modelo descrito no trabalho de Steinberg (1934).

Dadas as alternativas, não é surpreendente que eles tenham alcançado sucesso em seus esforços de reconhecimento de voz, ainda que só tenham podido diminuir o número de suspeitos para dois entre cinco, fato que acharam perturbador. É óbvio, no entanto, que eles viam com bastante otimismo a possibilidade de reconhecimento

confiável de voz. Segue abaixo uma citação do romance que pode dar ao leitor uma ideia da leve euforia que acompanhou esse sucesso inicial. Rubin, o responsável pelo experimento, faz um primeiro relato a um agente que aparece para ser informado sobre os resultados.

RUBIN: Somente o começo! Somente as mais provisórias deduções, Adam Veniaminovich!

POLICIAL: E quais são elas?

RUBIN: Elas estão abertas ao debate, mas uma coisa é certa. A ciência da fonoscopia, nascida hoje, 26 de dezembro de 1949, tem uma base racional.

E o futuro não parecia menos promissor:

Eles conceberam o sistema, como o das impressões digitais, que deve um dia ser adotado: um banco de dados de áudio unificado com as impressões vocais de todos aqueles que alguma vez estiveram sob suspeita. Qualquer conversa criminosa seria gravada, comparada, e o criminoso seria pego imediatamente, como um assaltante que deixasse suas impressões digitais na porta do cofre.

O romance termina aqui, e o que aconteceu em seguida não sabemos. Até onde sabemos, nenhum outro relato de desenvolvimentos posteriores da identificação de falantes durante aquela era estão disponíveis. Vale notar que o termo *fonoscopia*, usado para se referir à fonética forense e cunhado em Marfino, ainda é usado na Rússia e em muitos ex-países do leste europeu.

3. UM OLHAR DETALHADO SOBRE ALGUMAS QUESTÕES FUNDAMENTAIS

Nas seções seguintes, apresentaremos uma seleção de questões importantes em fonética forense e sugeriremos leituras importantes para aqueles que queiram explorá-las um pouco mais a fundo. Os livros e artigos aos quais nos referimos estão na maioria dos casos disponíveis em bibliotecas universitárias razoavelmente bem equipadas.

3.1 IMPRESSÕES VOCAIS

Não havia nada de controverso sobre o uso da expressão “impressão vocal” quando ela inicialmente apareceu no artigo de Grey e Kopp (1944) e é compreensível que os autores tenham sido inspirados pelos padrões que viram nos espectrogramas a fazer comparações com as impressões digitais e a especular sobre a possibilidade de usar espectrogramas para identificar falantes, assim como impressões digitais são

usadas para identificar indivíduos. A controvérsia surgiu bem depois, quando um engenheiro dos laboratórios Bell chamado Lawrence Kersta começou a usar as assim chamadas impressões vocais em aplicações forenses. Não sabemos muito sobre o envolvimento científico de Kersta no trabalho inicial com o espectrógrafo, realizado nos laboratórios Bell. Ele foi um engenheiro e não um pesquisador, mas foi ou tornou-se o chefe do laboratório de pesquisas naquela instituição, de modo que deve ter se familiarizado com as questões científicas também.

Como vimos no parágrafo anterior, pesquisadores dos laboratórios Bell não promoveram a ideia das impressões digitais de voz quando começaram a publicar depois da guerra, concentrando-se em questões de pesquisas tradicionais em fonética e aplicações em educação e terapia. Se eles também trabalhavam em identificação de falantes nós não sabemos. Pode muito bem ser o caso que essa parte do trabalho ainda fosse confidencial. De qualquer modo, houve um silêncio total dos laboratórios Bell sobre o assunto da identificação de falantes de suas primeiras publicações em 1944 até 1962, quando Kersta publicou um trabalho na revista *Nature*, intitulado *Voiceprint identification*, e depois, no mesmo ano, fez um discurso no encontro anual da Sociedade Americana de Acústica, intitulado *Voiceprint-identification infallibility*. Suas afirmações a respeito da precisão da identificação de falantes por meio de impressões vocais eram extraordinárias. Baseando-se em comparações visuais de algumas palavras cruciais, era possível alcançar nada menos do que 99% de identificações corretas e, em alguns casos, mais do que isso. Em 1966, Kersta deixou os laboratórios Bell, fundou sua própria empresa, chamada *Voiceprint Laboratories Corporation*, e começou a oferecer seus serviços em investigações criminais e treinar pessoas na identificação de impressões vocais. Até aquele momento, suas afirmações permaneceram largamente incontestadas pela comunidade científica. Assim, ele aproveitou um sucesso inicial, e seus testemunhos eram aceitos como evidência por tribunais em alguns estados americanos, mas não em todos.

Porém, ele logo começou a encontrar a resistência. Sujeitos em um estudo de Young e Campbell (1967), usando a técnica de impressão vocal, obtiveram 78,4% de identificações corretas usando um material de treino constituído por duas palavras ditas em isolamento. Quando as mesmas palavras, extraídas de contextos diferentes, eram usadas, a identificação baixou para 38,3%. Stevens *et al.* (1968) testou sujeitos que realizaram uma tarefa de identificação de falantes auditivamente através de fones de ouvido e visualmente por meio de espectrogramas. A margem de erro para os

testes em que os estímulos eram apresentados auditivamente foi de 6% comparados aos 21% para a identificação visual. A margem de alarme falso foi alta também na tarefa de identificação visual. Em ambos os estudos houve uma variação considerável nas taxas de identificação de falantes individuais, alguns falantes sendo bem mais difíceis de reconhecer do que outros. Numerosos outros estudos chegaram basicamente aos mesmos resultados, altas taxas de erro na identificação de impressões vocais e, em alguns casos, muito altas.

Seria injusto, no entanto, não mencionar que o método de Kersta também tinha adeptos. A maioria deles não eram pesquisadores em campos relevantes, mas pelo menos um deles, Tosi, era um foneticista qualificado. Tosi montou um laboratório em seu departamento e começou a testar as ideias de Kersta por meio de numerosos experimentos. Depois de dois anos de trabalho, ele publicou um artigo (TOSI *et al.*, 1972) no qual os resultados de suas pesquisas foram resumidos. As taxas de erro relatadas foram tipicamente entre 5 e 15%.

A controvérsia a respeito do assunto continuou até o final dos anos 1980, com Koenig (1986) e outros defendendo o método de Kersta e Shipp *et al.* (1987) e outros criticando-o. Análises baseadas na ideia de impressões vocais continuam sendo feitas por detetives particulares e outros “especialistas” não acadêmicos, embora ninguém na comunidade científica da fala ainda acredite na sua utilidade para finalidades forenses.

Para aqueles que leem francês, uma revisão excelente da controvérsia das impressões vocais pode ser encontrada em Meuwly (2003a, b).

Para um relato divertido, feito do ponto de vista de um leigo interessado no assunto, recomendo o livro de Block se for possível encontrá-lo.

3.2 RECONHECIMENTO AUDITIVO DE VOZ E MEMÓRIA

Conforme mencionamos na seção 3, o caso Lindbergh levantou questões sobre a acuidade do reconhecimento de voz e a memória. Um pesquisador que questionou se é possível lembrar com precisão uma voz desconhecida após um período de mais que dois anos foi a psicóloga Francis McGehee. Ela realizou dois estudos (McGEHEE 1937, 1944) que testaram o reconhecimento de voz em função do intervalo de tempo entre a primeira escuta da voz e uma tentativa posterior de reconhecimento da mesma voz em uma fila. No primeiro experimento, os ouvintes escutaram (mas não viram) um falante lendo uma passagem de texto contendo 56 palavras. Os ouvintes

foram então divididos em grupos que escutaram uma fila de vozes com cinco distratores, mais a voz do falante cuja voz havia sido ouvida anteriormente, em intervalos de um, dois e três dias, uma, duas e três semanas e um, três, e cinco meses, respectivamente. A taxa de reconhecimento variou em função do tempo, começando com um pouco mais de 80% de identificações corretas depois de um lapso de um dia ou uma semana. Depois de duas semanas, a taxa de reconhecimento havia caído para 69%; depois de um mês, para 57%; depois de três meses, para 35%; e depois de cinco meses, para 13%, o que é menos do que se poderia esperar se os sujeitos se comportassem aleatoriamente. Os resultados desse experimento de reconhecimento de voz estão, no geral, de acordo com outros estudos sobre a piora na retenção de memórias com o passar do tempo. A segunda série de experimentos apresentadas em McGehee (1944) diferiu da primeira, principalmente pelo fato de que foram usadas vozes gravadas em vez de vozes ao vivo atrás de um biombo, como foi o caso na primeira série de experimentos. Mas os resultados dos dois experimentos são muito semelhantes. Ela também realizou sub-estudos nos quais outros fatores variavam, tais como a familiaridade do ouvinte com a língua, mas esses resultados não serão discutidos aqui. Estudos posteriores confirmaram, de forma geral, seus resultados, ainda que a taxa exata de piora na retenção da memória varie de estudo para estudo. Referências para alguns dos estudos posteriores são encontrados ao final do tutorial.

3.3 AMOSTRAS DE FALA NÃO CONTEMPORÂNEAS

O termo se refere a amostras de fala que são obtidas em diferentes pontos no tempo e usadas posteriormente em um processo de identificação. Sabemos, naturalmente, que a voz humana muda ao longo do tempo. Mas a mudança acontece em geral bastante lentamente. A questão relevante, no contexto da fonética forense, é determinar, entre duas gravações, o intervalo de tempo decorrido a partir do qual mudanças na voz se tornam um problema para a comparação das amostras. É razoável imaginar que, no decorrer de um longo período de tempo, aconteçam mudanças marcantes. Por razões óbvias, há poucos estudos longitudinais desse tipo. Naquele realizado por Endres *et al.* (1971), gravações de sete falantes coletadas ao longo de um período de tempo de 29 anos foram comparadas. Os autores encontraram uma tendência de queda, em função do aumento da idade, para a frequência fundamental e as frequências formânticas. Em casos forenses, é incomum, para dizer o mínimo, que um espaço de tempo da ordem de décadas separe uma

gravação de um suspeito e uma tentativa posterior de identificar o falante usando uma gravação recente. Mas espaços de um ano ou mais não são incomuns. Assim, é importante saber se mudanças na voz que acontecem em um período de um ou alguns anos podem afetar a precisão do reconhecimento de falantes. Essa questão foi abordada em uma série de estudos por Hollien e Schwartz (2000; 2001). Em seus experimentos, os autores testaram latências entre gravações entre quatro semanas e vinte anos. Houve uma queda na identificação correta de 95% para amostras contemporâneas para 70-85% para latências entre 4 semanas e 6 anos (sem nenhuma tendência observável no intervalo). Para a latência de 20 anos, no entanto, uma redução forte para 35% pôde ser observada. Dois fatores, além da latência de tempo entre gravações, também foram testados: experiência do ouvinte e semelhanças entre vozes. Como seria de esperar, foneticistas experientes tiveram um desempenho melhor do que estudantes. Para os foneticistas, a identificação correta chegou a 76% para a latência de 20 anos. A semelhança entre as vozes, entretanto, provocou um efeito dramático de piora sobre o reconhecimento. A performance caiu de um pouco abaixo de 95% para amostras contemporâneas para um pouco mais de 40% para amostras gravadas 4 semanas depois. Para as latências típicas em investigações forenses, amostras não contemporâneas de fala pareceram, assim, afetar a identificação somente marginalmente, pelo menos se foneticistas treinados forem usados como ouvintes e as vozes não sejam similares demais.

3.4 OUTRAS QUESTÕES ENVOLVENDO A AMOSTRA DE FALA

Outros fatores que podem influenciar a acuidade da identificação são, primeiramente, a duração e a qualidade acústica da amostra. Se considerarmos a influência da duração da amostra, podemos observar que, em investigações reais, as amostras de fala podem ser bem curtas, na maioria das vezes, algumas palavras ou uma frase ou duas, o que significa que sua duração será da ordem de alguns poucos segundos. Em um estudo antigo feito por Pollack *et al.* (1954), os autores observaram que a acuidade na identificação a partir de palavras monossilábicas aumentou conforme o aumento do tamanho da amostra, mas somente até 1,2 segundos. Em amostras mais longas, os autores alegam que a variação fonética torna-se o fator mais importante. Eles concluem dizendo que “nós acreditamos que a duração da amostra de fala *per se*, não é particularmente importante, exceto na medida em que ela permite a obtenção de uma amostra estatística maior ou menor do repertório de fala

do falante”. Esta descoberta, até certo ponto surpreendente, foi, no entanto, confirmada em outros estudos. Em um estudo reportado em Compton (1963), quinze segmentos da vogal [i] gravados por nove diferentes falantes foram apresentados a ouvintes familiarizados com aqueles falantes. Os segmentos diferiam apenas em sua duração (25-2500 ms). Para segmentos maiores que 75 ms não houve aumento na taxa de reconhecimento. Bricker e Pruzansky (1966) apresentaram estímulos que variavam tanto em duração quanto e em composição fonêmica. Eles descobriram que a taxa de identificação aumentava com a duração somente se o estímulo mais longo contivesse também mais variação fonêmica e que “a acuidade na identificação aumentou diretamente com o número de fonemas na amostra mesmo quando a duração foi controlada”. Em um estudo publicado por Orchard e Yarmey (1995), a taxa de identificação correta foi substancialmente maior para estímulos de oito minutos comparada com estímulos de trinta segundos. Não foram feitas tentativas, no entanto, para estimar as respectivas contribuições de duração e variação fonológica, mas é provável que a variação fonológica deve ter sido maior nos estímulos mais longos.

É importante ressaltar, no entanto, que, enquanto um aumento nas identificações corretas é desejável, é igualmente desejável manter baixo o número de alarmes falsos. O reconhecimento de vozes gravadas por telefone foi investigado em um trabalho anterior de Yarmey (1991). Nesse estudo, o número de identificações corretas aumentou quando a duração das amostras aumentou de 3,2 para 4,3 minutos e finalmente para 7,8 minutos. Mas os alarmes falsos também aumentaram. O mesmo tipo de relação entre acertos e alarmes falsos foi observado num estudo similar feito por Yarmey e Matthys (1992): “O efeito facilitador da identificação de durações de amostras de voz mais compridas foi neutralizado pelas taxas altas de alarme falso tanto na condição suspeito-presente quanto suspeito-ausente”. Minimizar alarmes falsos é, certamente, muito importante em situações forenses reais embora seja uma questão muitas vezes negligenciada.

A qualidade acústica é um tópico de grande abrangência e pode ser pensado de diferentes maneiras. Os problemas mais comuns de qualidade que afetam a prática da fonética forense são o ruído de fundo e a largura de banda das gravações ou transmissões. Entendemos ruído de fundo como tudo aquilo que não é sinal da fala, o que inclui coisas tão diversas como um barulho ambiental qualquer, a fala de outra pessoa ou um rádio tocando ao fundo. Há um grande número de estudos na área do

reconhecimento automático da fala e de sistemas de diálogo em que essas questões têm sido investigadas, mas não há muitos trabalhos que tratam da influência do ruído de fundo sobre o reconhecimento de falantes feito por humanos. Embora esse seja um tópico importante, ele não será explorado neste trabalho.

Uma grande proporção de ameaças é feita pelo telefone, e criminosos muitas vezes usam o telefone quando planejam ou coordenam a execução de crimes. A qualidade da fala no telefone tem, portanto, recebido atenção nos estudos sobre fonética forense. Linhas telefônicas têm largura de banda limitada. A maioria das frequências relevantes para a percepção da fala está presente, mas não todas. Frequências menores que 300 Hz, por exemplo, são deixadas de fora. Com telefones celulares, problemas relacionados à codificação da fala são introduzidos. Esses efeitos são particularmente notáveis para vozes femininas.

Questões importantes no contexto forense são se a qualidade de som mais pobre de conversas telefônicas afeta a identificação de voz e, se sim, em que medida e como isso ocorre. Além disso, de um ponto de vista metodológico, é importante saber se se devem usar somente vozes gravadas por telefone em filas de reconhecimento nas quais a chamada incriminadora foi gravada por telefone. Surpreendentemente, há poucos estudos que abordam essa questão, mas há alguns resultados que indicam que o problema pode não ser tão sério quanto se poderia imaginar. Rathborn, Bull e Clifford (1981, citado em YARMEY, 1991), por exemplo, “não acharam quaisquer diferenças significativas na identificação de voz de uma voz-alvo ouvida originalmente pelo telefone e testada usando uma fila de reconhecimento apresentada pelo telefone, em contraste com identificação de voz ouvida originalmente pelo telefone e testada diretamente a partir de uma fila apresentada a partir de uma gravação em fita”.

Uma questão que tem recebido alguma atenção ultimamente é a influência da filtragem passa-faixa que ocorre nas transmissões telefônicas sobre a análise acústica de amostras de voz. Em um estudo recente, Künzel (2001) constatou que o valor relativamente alto do limite inferior da faixa de rejeição (300 Hz) tem o efeito de elevar o F1 em vogais do alemão quando comparadas com exemplares correspondentes em uma gravação simultânea não filtrada. O tamanho médio do deslocamento foi de 6,6% para os homens e de 6,1% para as mulheres, e todas as diferenças foram significativas para um nível de 5% ou melhor. Outros artefatos de pequena magnitude foram observados. Como consequência, Künzel adverte contra o

uso de dados de formantes para propósitos de identificação de falantes se as gravações foram feitas a partir de telefones. Seus resultados não foram questionados, mas sua rejeição total do uso de dados de formantes na identificação de falantes baseado em gravações telefônicas foi posta em dúvida por Nolan (2002).

3.5 FAMILIARIDADE COM O FALANTE

Todos nós já tivemos a experiência de reconhecer uma voz familiar, e este reconhecimento é, na maioria das vezes, rápido e exato. Mesmo estímulos não linguísticos curtos como uma tosse são, muitas vezes, suficientes para reconhecer uma pessoa familiar. Mas essas observações informais não são suficientes. Na fonética forense, precisamos ser mais precisos sobre a influência da familiaridade na acuidade do reconhecimento de voz. Há alguns estudos que abordaram essa questão. Hollien *et al.* (1982) estudaram como a familiaridade de ouvintes em relação a vozes afeta a tarefa de reconhecimento de falantes em três condições de fala diferentes: normal, fala com disfarce vocal e sob efeito de estresse. Os resultados revelaram que ouvintes familiarizados com os falantes cujas vozes foram usadas no experimento tiveram um desempenho significativamente melhor em todas as condições. Esses resultados foram, de forma geral, confirmados em outros estudos (e.g. SCHMIDT-NIELSEN; STERN, 1985).

É importante destacar, no entanto, que, embora as taxas de reconhecimento sejam geralmente altas para vozes com as quais os ouvintes são familiarizados, o reconhecimento está longe de ser perfeito. Para falantes e ouvintes individuais, a taxa de erro pode chegar a 30-40% se os enunciados forem curtos e pertencerem a um conjunto aberto relativamente grande (LADEFOGED; LADEFOGED, 1980). A influência do tamanho do enunciado sobre o reconhecimento de vozes familiares também foi verificada em outros estudos. Em uma série de experimentos relatados por Rose e Duncan (1995), o reconhecimento de vozes familiares variou de aleatório a quase perfeito em função da duração dos enunciados.

Assume-se geralmente que, no reconhecimento de vozes, a discriminação constitui o passo inicial, com reconhecimento ocorrendo numa fase posterior em um processo unificado. Van Lacker *et al.* (1985) mostraram, no entanto, que esse não parece ser o caso e sim que “discriminação e reconhecimento não são estágios de um processo coerente, mas habilidades dissociadas, não ordenadas temporalmente”. Assim, é perfeitamente possível que um ouvinte que seja bom em reconhecer vozes

familiares tenha um desempenho ruim se a tarefa é a de discriminar entre falantes com vozes não familiares.

3.6 DISFARCE VOCAL

Disfarces vocais, na medida em que são usados, podem ser um problema sério para a identificação de falantes. Em um dos extremos do espectro de possibilidades, encontramos a manipulação eletrônica ou até mesmo a comunicação via síntese de fala, que tornaria a identificação de falantes impossível na prática. No mundo do trabalho forense real, no entanto, o emprego dos disfarces vocais tende a ser muito pouco sofisticado. Künzel (2000) nota, baseado na experiência do BKA (*Bundeskriminalamt*, a polícia federal alemã), que “falseto, voz crepitante, sussurros, imitação de sotaque estrangeiro, e segurar as narinas” são os tipos mais comuns. Basicamente, as mesmas observações foram feitas em estudos experimentais. Em um estudo de Masthoff (1996), em que alunos de graduação serviram como sujeitos, a maioria dos disfarces escolhidos (35%) foram de nível fonológico (sussurro, aumento ou diminuição da frequência fundamental). Disfarces de nível articulatorio (imitação de padrões dialetais, sotaque estrangeiro, etc.) também foram usados (20%). Os disfarces restantes eram combinações de dois tipos. Mensagens eletronicamente manipuladas ainda são raras, embora Künzel note ter havido um aumento nos últimos anos, principalmente na forma de edição de vozes gravadas.

Mesmo que os tipos de disfarce usados, na maioria das vezes, não sejam sofisticados, o disfarce pode, ainda assim, ter um efeito negativo considerável na identificação de falantes. Em um estudo de Reich e Duke (1979), em que vários tipos de disfarce foram usados, todos os tipos produziram identificações significativamente menos corretas. A hipernasalidade produziu o maior efeito negativo, embora, na maioria dos casos, não tenha havido diferenças significativas entre diferentes tipos. O emprego do sussurro, um dos tipos mais comuns, resultou em acentuadamente menos identificações corretas em um estudo de Orchard e Yarmey (1995) quando amostras com sussurros eram comparadas a amostras vozeadas. Nos casos em que tanto a referência quanto as amostras de teste eram sussurradas, a diferença era menos pronunciada.

O uso de disfarce vocal não é tão comum quanto se pensa. Künzel (2000) relata que: “Ao longo das últimas duas décadas, entre 15 e 25% dos casos anuais com

os quais a sessão de identificação de falantes do BKA lida apresentaram pelo menos um tipo de disfarce”.

3.7 SOTAQUES E LÍNGUAS ESTRANGEIRAS

A influência de sotaques estrangeiros ou línguas estrangeiras na identificação de falantes foi investigada em uma série de estudos. Geralmente se pensa que sotaques estrangeiros tornam a identificação mais difícil, mas a diferença é geralmente pequena ou inexistente. O trabalho de McGehee (1937) mencionado acima incluiu um estudo da influência de sotaques estrangeiros. O reconhecimento de um falante de inglês com um sotaque alemão foi testado. Nenhuma diferença na taxa de reconhecimento foi encontrada. “Uma voz estrangeira não familiar (falando alemão) foi reconhecida aproximadamente pela mesma percentagem de ouvintes como uma voz não familiar norte-americana quando cada uma ocorreu em condições semelhantes”. Em outros estudos, no entanto, diferenças foram encontradas. Em um estudo de Doty (1998), falantes nativos de inglês dos EUA e da Inglaterra e falantes de inglês como língua estrangeira da França e de Belize foram gravados lendo sentenças em inglês. Com falantes nativos de inglês como ouvintes, a taxa de reconhecimento foi dramaticamente maior para outros falantes nativos do que para falantes com sotaques estrangeiros. Os resultados de um estudo publicado por Goldstein *et al.* (1981) sugerem uma situação intermediária: “Em amostras de fala relativamente longas, o reconhecimento de vozes com sotaque não foi mais difícil do que o de vozes sem sotaque; a redução da duração das amostras de fala diminuiu o reconhecimento de vozes com e sem sotaque, mas a redução foi maior para vozes com sotaques”. Vê-se, portanto, que os resultados são de certo modo ambíguos, embora possamos concluir que vozes com sotaque tendem a ser menos reconhecidas, ainda que a diferença seja em geral pequena. É muito possível também que profissionais experientes, como linguistas, sejam melhores do que ouvintes leigos para reconhecer vozes com sotaques.

A influência de uma língua estrangeira também foi assunto de muitos estudos. Em Thompson (1987), seis estudantes masculinos bilíngues gravaram mensagens em inglês, espanhol e inglês com um sotaque espanhol forte. A amostra apresentada na fila de reconhecimento era falada na mesma língua e sotaque que a mensagem inicial. As vozes foram identificadas melhor (por ouvintes falantes apenas de inglês) quando falando inglês e pior quando falando espanhol. A acuidade de identificação foi

intermediária para a condição de sotaque. Schiller e Köster (1996) testaram norte-americanos sem nenhum conhecimento de alemão, norte-americanos com alguns conhecimentos de alemão e falantes nativos de alemão usando gravações de seis falantes nativos de alemão. Participantes com nenhum conhecimento de alemão fizeram significativamente mais erros que outros sujeitos. Participantes com um pouco de conhecimento de alemão tiveram desempenho similar a falantes nativos de alemão.

Köster e Schiller (1997) replicaram o experimento feito por Schiller e Köster usando ouvintes espanhóis e chineses. “Descobriu-se que os ouvintes espanhóis e chineses que eram familiares com o alemão mostraram taxas de reconhecimento melhores que ouvintes espanhóis e chineses sem conhecimento de alemão, enquanto os ouvintes espanhóis e chineses com o conhecimento de alemão tiveram um desempenho pior que os ouvintes alemães e norte-americanos com conhecimentos de alemão”.

Podemos resumir os resultados dizendo que os ouvintes sem nenhum conhecimento da língua têm desempenho pior no reconhecimento da fala que ouvintes com algum conhecimento ou falantes nativos, enquanto ouvintes com um pouco de conhecimento da língua tendem a apresentar um desempenho no mesmo nível que falantes nativos.

3.8 TESTEMUNHAS AURICULARES

Fatores que são relevantes para o reconhecimento de falantes em geral, como memória, familiaridade e disfarce, descritos acima, também são relevantes quando falamos sobre testemunhas auriculares, embora haja fatores adicionais sobre os quais ainda não sabemos tanto quanto gostaríamos. Como Bull e Clifford (1984) apontam, “a maior parte dos (relativamente poucos) estudos sobre testemunho auricular tem poucas semelhanças com as circunstâncias reais vividas por testemunhas. A maioria usou situações não estressantes com sujeitos preparados que participaram em situações criadas em laboratório”.

Em primeiro lugar, não apenas o estresse que uma testemunha pode sofrer em uma situação real nunca poderá ser totalmente recriado em um experimento em laboratório, como também estudiosos e testemunhas não dispõem de muitas experiências a que poderiam recorrer para ajudar a determinar exatamente como e em que grau pode ser afetada a habilidade de uma vítima traumatizada para

reconhecer e discriminar vozes. Segundo, “experiências pessoais de reconhecimento de voz são sempre de vozes familiares – as vozes que em geral *não* são aquelas a serem identificadas em situações criminais” (Bull e Clifford). E como sabemos a partir do trabalho de Van Lancker e Kreiman (ver seção 3.5), reconhecer uma voz familiar e diferenciar entre aquelas não familiares são habilidades independentes. Em terceiro lugar, enquanto os participantes de um experimento de laboratório estão, em maior ou menor grau, preparados para a situação, testemunhas reais, na maioria dos casos, são pegas de surpresa. Estudos mostraram (e.g., CLIFFORD; DENOT, 1982, citados em BULL; CLIFFORD, 1984) que a acuidade da identificação de voz é bastante mais baixa em condições em que os ouvintes não estão preparados para a tarefa. A confiança das testemunhas em relação ao seu julgamento também não é de grande ajuda. Bull e Clifford (1984) relatam correlações significativas entre a acuidade e a confiança, mas outros estudos (e.g. YARMEY, 2001) não encontraram correlação semelhante.

3.9 FILAS DE RECONHECIMENTO COM TESTEMUNHAS AURICULARES

Uma fila reconhecimento de voz (*lineup* ou *voice parade*) apresentada a uma testemunha auricular pode ser considerada o equivalente auditivo de uma fila de reconhecimento visual apresentada a uma testemunha ocular (*eyewitness lineup*). É usada quando uma pessoa ouviu, mas não viu o criminoso. Como foi mencionado na seção 3.8, normalmente a testemunha não tem familiaridade com a voz. Nas filas de reconhecimento, gravações da voz de um suspeito e vozes de outros falantes, usadas como distratoras, são apresentadas para a testemunha, cuja tarefa é comparar as vozes gravadas com a memória da voz escutada e determinar se alguma das vozes gravadas é compatível com a memória da voz do criminoso.

Duas questões importantes em conexão com as filas de reconhecimento aplicadas a testemunhas auriculares são: quantas vozes devem ser apresentadas e o quão similar à voz do suspeito devem ser as vozes distratoras?

Descobriu-se que, em filas de reconhecimento com poucas vozes, observam-se fortes efeitos da posição da voz na fila. Descobriu-se também que o número de identificações corretas cai quando o tamanho da fila aumenta. Então, a questão é se há um tamanho ideal no qual o efeito de posição é minimizado e a queda nas identificações corretas aumenta. Há um número de estudos que tem tratado dessas questões, mas aqui iremos citar apenas um. Bull e Clifford (1984) testaram, em dois

experimentos, a influência do tamanho das filas de reconhecimento no desempenho das testemunhas. No primeiro experimento, cinco ou onze vozes distratoras foram usadas e, no segundo experimento, quatro, seis ou oito. Houve diferenças significativas entre a condição com quatro vozes distratoras quando comparada às condições com seis ou oito distratoras, com diferenças mínimas nos outros casos. Os resultados sugerem como regra geral, assim, o uso de cinco ou seis distratoras. Eles também descobriram um efeito de posição somente quando o alvo acontecia na posição inicial.

Quão similar ao alvo devem ser as vozes distratoras? Essa é, com certeza, uma questão difícil com muitas complicações, mas pelo menos dois extremos devem ser evitados. A voz-alvo não deve destacar-se do resto. Os falantes precisam ser razoavelmente semelhantes com respeito a características gerais como idade, dialeto, etc. Por outro lado, não devem ser excessivamente parecidas. Quando Rothman (1977, citado em HOLLIEN, 2002) usou falantes com vozes parecidas (irmãos, pais e filhos), a identificação caiu de 94% (distratores não parecidos) para 58% (distratores parecidos). Resultados semelhantes foram obtidos por Hollien e Schwartz (2000, ver 4.3). Assim, as vozes distratoras devem contemplar um grau razoável de variação, mas evitar os extremos.

3.10 DETECÇÃO DE MENTIRAS USANDO FMRI E OUTRAS TÉCNICAS

Recentemente foram feitas tentativas para usar métodos de imageamento do cérebro para estudar a possibilidade de detectar diferenças consistentes em padrões de atividades do cérebro que podem ser usados para separar mentiras ou insinceridades de declarações verdadeiras. Ainda que a pesquisa nesse campo esteja apenas em seu início, resultados muito interessantes foram obtidos. Apesar disso, aqui, abordaremos esses estudos apenas marginalmente, uma vez que este tutorial é sobre fonética forense, e poucos resultados derivados dessas pesquisas são diretamente relevantes para a fonética forense como tal.

Langleben *et al.* (2002) usaram ressonância magnética funcional (fMRI) para ver se era possível detectar diferenças na atividade cerebral quando seus sujeitos contavam uma mentira e quando falavam a verdade. Seus resultados indicam que houve, de fato, uma diferença: “Este achado indica que há uma diferença neurofisiológica entre mentira e verdade no nível de ativação cerebral que pode ser

detectada com fMRI". Resultados aproximados foram obtidos em outros estudos (e.g., LEE *et al.*, 2002).

Trabalhos interessantes usando a técnica fMRI para buscar correlatos da percepção de voz (BELIN *et al.*, 2004) e da familiaridade de pessoa (voz e rosto) também estão em curso (SHAH *et al.*, 2001).

Imagens térmicas de alta resolução que podem detectar mudanças espaciais de pequenas dimensões no fluxo do sangue, por exemplo, também foram usadas na tentativa de desenvolver métodos para detectar mentiras (PAVLIDIS; LEVINE, 2002).

Como alguns dos próprios autores das pesquisas citadas apontam, devemos estar atentos para o fato de que esses são resultados bastante preliminares. É preciso ter sempre em mente que resultados como os relatados aqui são fruto de experimentos em laboratório, na maioria das vezes, altamente sofisticados, demorados e caros! Quando e se esses métodos poderão vir a ser usados no campo forense, não saberemos por mais alguns anos. Precisamos estar também cientes de que poderá haver um longo caminho entre resultados de pesquisa e aplicações confiáveis que podem ser usadas em campo. Infelizmente, esse nem sempre é o caso. “Tecnologias cuja validade não está estabelecida estão atraindo cada vez mais a atenção das forças policiais norte-americanas e agências de segurança... Ferramentas de laboratório – de sensores infravermelhos a rastreamento ocular – estão sendo transformadas em detectores de mentira” (KNIGHT, 2004).

3.11 GENERALIZAÇÃO EXCESSIVA, CHARLATANISMO E FRAUDE

A possibilidade de detectar mentiras ou insinceridades com algum tipo de detector de mentiras automático é, naturalmente, algo desejado pela polícia e outros investigadores. No mundo dos filmes e das histórias em quadrinhos, detectores de mentira perfeitamente confiáveis existem há muito tempo. Todavia, na realidade, não chegamos tão longe. O mais conhecido “detector de mentira” é o chamado polígrafo (*polygraph*). Sua primeira aparição, em uma forma bastante preliminar, pode ser datada de 1917. Uma versão mais refinada produzida no começo dos anos 1920 foi usada em um tribunal, em 1923, e ele tem sido usado desde então com alguns aperfeiçoamentos. A ideia básica por trás do uso do polígrafo como “detector de mentiras” é que mentiras aumentam o nível de estresse na pessoa que está mentindo e, se for possível registrar precisamente as reações involuntárias que sabemos ser

correlacionadas com estresse – tais como respiração, pulsação, pressão sanguínea, resposta galvânica da pele – pode-se também detectar mentiras. O problema de usar o polígrafo como um detector de mentiras reside na interpretação. Correlações entre níveis de estresse e pulsação, por exemplo, são encontrados quando se analisam resultados de grupos de pessoas. Generalizar de resultados coletivos para indivíduos particulares não é, certamente, um passo válido. Também não é um passo válido concluir que uma pessoa que sente estresse deve necessariamente estar mentindo. Pessoas que defendem o uso do polígrafo evitam chamá-lo de detector de mentiras, muito embora usem-no como se fosse um, o que é quase a mesma coisa.

O polígrafo como tal é relevante apenas marginalmente no contexto da fonética forense, mas o princípio geral de tentar encontrar correlatos de mentira e insinceridade que possam ser medidos de forma confiável é fundamental também em tentativas de construir detectores de mentiras baseados na análise de voz. A ideia é que alguma propriedade da voz pode ser usada como uma pista para a detecção de mentira. Sugeriu-se que a presença de microtremores na voz pudesse ser usada para detectar mentiras, e alguns analisadores baseados nessa ideia são comercializados. Não há, no entanto, nenhuma base científica para as afirmações de que esses analisadores possam detectar mentiras. Hollien (1987) revisou a literatura sobre o tema e concluiu que: “a habilidade de analisadores de voz para detectar estresse na fala – ou para identificar mentiras que são faladas – tem sido negativa ou inconclusiva”. Ele também realizou testes usando analisadores de voz comerciais que revelaram ter um desempenho aleatório: “identificações de estresse/não estresse aconteceram somente de forma aleatória; os resultados das identificações mentira/não mentira eram bastante semelhantes”. Shipp e Izdebski (1981) testaram a ideia usando eletrodos inseridos nos músculos da laringe, mas nenhum padrão de microtremor foi encontrado. Produtos baseados nesse princípio, no entanto, ainda continuam sendo usados por detetives particulares e, em alguns casos, até pela polícia. Dada a fraca ou inexistente base científica subjacente a esses recursos, pode-se justificadamente chamar seu uso de charlatanismo no melhor dos casos.

Mas há também pura e simplesmente fraudes. Uma companhia com base em Israel comercializa ferramentas que supostamente fazem coisas extraordinárias, incluindo tanto detectores de mentira quanto detectores de amor. A técnica por trás do detector de mentira é chamada de *Layered Voice Analysis* (LVA), e o pressuposto é que

todo “evento” que passa pelo cérebro vai deixar suas “impressões digitais” no fluxo da fala. A tecnologia LVA ignora o que o sujeito está dizendo, e foca somente em sua atividade cerebral. Em outras palavras, o “como” é dito é fundamental e não o “o quê”.

Eles são cuidadosos em não chamar o recurso explicitamente de detector de “mentiras”, mas não há absolutamente nenhuma questão de que é nisso que querem que acreditemos: “A LVA é capaz de detectar a intenção por trás da mentira e, fazendo isso, pode levar você a identificar e revelar a mentira propriamente dita”.

Como até mesmo pessoas com o menor conhecimento em análise de voz sabem, não há nem vestígio de evidência de uma relação entre voz e atividade cerebral do tipo proposto. Além disso, a análise cuidadosa da descrição do método nos documentos que solicitam a concessão de patente para o produto nos EUA confirma a suspeita de que o método é um completo disparate, talvez melhor descrito como estatísticas baseadas na digitalização de artefatos.

Você poderia pensar que uma companhia que comercializa máquinas que produzem impressões digitais do cérebro e detectores de amor levantariam suspeitas ou pelo menos cautela em seus clientes prospectivos, mas isso parece não ter sido o caso em geral. A companhia é um negócio de milhões de dólares tendo como cliente, entre outros, algumas companhias de seguro do Reino Unido ou dos EUA. Também há relatos que seus produtos são usados por departamentos de polícia nos EUA e talvez em outros lugares.

Podemos aprender alguma coisa a partir de experiências anteriores, especificamente que há certo perigo em ignorar completamente os charlatões. Leigos podem erroneamente interpretar o silêncio como aceitação, não importando o quão exorbitante, até mesmo ridículas, as afirmações possam parecer para um especialista no campo. Por outro lado, pode ser bastante demorado expor charlatanices, tempo que deixará de ser usado para coisas mais importantes cientificamente. Aqui reside um dilema que temos de enfrentar.

REFERÊNCIAS

1. Leituras de base

1. HOLLIEN, Harry. (2002). *Forensic Voice Identification*. San Diego, CA: Academic Press.
2. ROSE, Phil. (2003). *Forensic Speaker Identification*. New York: Taylor & Francis.

2. Antecedentes históricos

3. GREY, G.; KOPP, G. A. (1944). "Voiceprint identification." *Bell Telephone Laboratories Report*: 1–14.
4. HOLLIEN, Harry. (2002). *Forensic Voice Identification*. San Diego, CA: Academic Press.
5. KOPP, G. A.; GREEN, H. C. (1946). "Basic phonetic principles of visible speech." *Journal of the Acoustical Society of America* 18: 74–89.
6. MEUWLY, D. (2003). "Le mythe de « L’empreinte vocale » (I)." *Revue internationale de criminologie et de police technique et scientifique* 56(2): 219–236.
7. POTTER, R. (1945). "Visible patterns of speech." *Science* November: 463–470.
- SOLZHENITSYN, A. I. (1968). *The First Circle*. Evanston, Ill: Northwestern University Press.
8. STEINBERG, J. C. and N. R. French (1946). "The portrayal of visible speech." *Journal of the Acoustical Society of America*, 18: 4–18.

3.1 Impressões vocais

9. BLOCK, E. B. (1975). *Voiceprinting: How the Law Can Read the Voice of Crime*. New York: David McKay Company, Inc.
10. KERSTA, L. G. (1962). "Voiceprint identification." *Nature*, 196:1253–1257.
11. KERSTA, L. G. (1962). "Voiceprint-identification infallibility." *Journal of the Acoustical Society of America*, 34:1978.
12. KOENIG, B. E. (1986). "Spectrographic voice identification: A forensic survey." *Journal of the Acoustical Society of America* 79: 2088-2090.
13. MEUWLY, D. (2003a). "Le mythe de « L’empreinte vocale » (I)." *Revue internationale de criminologie et de police technique et scientifique*, 56(2): 219–236.
14. MEUWLY, D. (2003b). "Le mythe de « L’empreinte vocale » (II)." *Revue internationale de criminologie et de police technique et scientifique* 61(3): 361–374.
15. SHIPP, T. E. T. Doherty and H. Hollien. (1987). "Some fundamental considerations regarding voice identification." *Journal of the Acoustical Society of America* 82: 687-688
16. STEVENS, K. N. et al. (1968). "Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material." *Journal of the Acoustical Society of America* 44: 1596–1607.
17. TOSI, O. et al. (1972). "Experiment on voice identification." *Journal of the Acoustical Society of America* 51: 2030– 2043.
18. YOUNG, M. A.; R. A. CAMPBELL (1967). "Effects of context on talker identification." *Journal of the Acoustical Society of America* 42: 1250–1254.

3.2 Reconhecimento auditivo de voz e memória

19. CLIFFORD, B. R.; RATHBORN, H.; R. BULL. (1981). "The effects of delay on voice recognition accuracy". *Law and Human Behavior*, 5, 201–208.

20. PAPANICOLAOU, G.; KREIMAN, J.; DAVIS, A. (1989). "Long-term memory for unfamiliar voices". *Journal of the Acoustical Society of America*, 85, 913–925.
21. MCGEHEE, F. (1937). "The reliability of the identification of the human voice". *Journal of General Psychology*, 17, 249–271.
22. MCGEHEE, F. (1944). "An experimental study of voice recognition". *Journal of General Psychology*, 31, 53–65.
23. SASLOVE, H.; YARMEY, A. D. (1980). "Long-term auditory memory: speaker identification". *Journal of Applied Psychology*, 65, 111–116.

3.3 Amostras de fala não contemporâneas

24. ENDRES, W.; BAMBACH, W.; FLÖSSER, G. (1971). "Voice spectrograms as a function of age, voice disguise, and voice imitation". *Journal of the Acoustical Society of America*, 49, 1842–1848.
25. HOLLIEN, H.; SCHWARTZ, R. (2000). "Aural-perceptual speaker identification: Problems with noncontemporary samples". *Forensic Linguistics*, 7, 199–211.
26. HOLLIEN, H.; SCHWARTZ, R. (2001). "Speaker identification utilizing noncontemporary speech". *Journal of Forensic Sciences* 46: 63–67.

3.4 Outras questões envolvendo a amostra de fala

27. Bricker, P. D. and S. Pruzansky (1966). "Effects of stimulus content and duration on talker identification". *Journal of the Acoustical Society of America* 40: 1441-1450.
28. Compton, A. J. (1963). "Effects of filtering and vocal duration upon the identification of speakers aurally". *Journal of the Acoustical Society of America* 35: 1748–1752.
29. Künzel, H. J. (2001). "Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies." *Forensic Linguistics* 8: 80–99.
30. Nolan, F. (2002). "The 'telephone effect' on formants: A response." *Forensic Linguistics* 9: 74–82.
31. Orchard, T. L. and A. D. Yarmey (1995). "The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification." *Applied Cognitive Psychology* 9(3): 249–260.
32. Pollack, I., J. M. Pickett, et al. (1954). "On the identification of speakers by voice." *Journal of the Acoustical Society of America* 26: 403–412.
33. Yarmey, A. D. (1991). "Voice identification over the telephone." *Journal of Applied Social Psychology* 21: 1868–1876.
34. Yarmey, A. D. and E. Matthys (1992). "Voice identification of an abductor." *Applied Cognitive Psychology* 6: 367–377.

3.5 Familiaridade com o falante

35. Hollien, H., W. Majewski and E. T. Doherty. (1982). "Perceptual identification of voices under normal, stress, and disguise speaking conditions." *Journal of Phonetics* 10: 139–148.

36. Ladefoged, P. and J. Ladefoged (1980). "The ability of listeners to identify voices." *UCLA Working Papers in Phonetics* 49: 43–51.
37. Rose, P. and S. Duncan (1995). "Naive auditory identification and discrimination of similar voices by familiar listeners." *Forensic Linguistics* 2: 1–17.
38. Schmidt-Nielsen, A. and K. R. Stern (1985). "Identification of known voices as a function of familiarity and narrow-band coding." *Journal of the Acoustical Society of America* 77: 658–663.
39. Van Lancker, D. and J. Kreiman (1985). "Unfamiliar voice discrimination and familiar voice recognition are independent and unordered abilities." *UCLA Working Papers in Phonetics* 62: 50–60.

3.6 Disfarce vocal

40. Künzel, H. (2000). "Effects of voice disguise on speaking fundamental frequency." *Forensic Linguistics* 7: 149–179.
41. Masthoff, H. (1996). "A report on a voice disguise experiment." *Forensic Linguistics* 3: 160–167.
42. Orchard, T. L. and A. D. Yarmey (1995). "The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification." *Applied Cognitive Psychology* 9(3): 249–260.
43. Reich, A. R. and J. E. Duke (1979). "Effects of selected vocal disguises upon speaker identification by listening." *Journal of the Acoustical Society of America* 66:1023–1028.

3.7 Sotaques e línguas estrangeiras

44. Doty, N. D. (1998). "The influence of nationality on the accuracy of face and voice recognition." *American Journal of Psychology* 111: 191–214.
45. Goldstein, A. G., P. Knight, et al. (1981). "Recognition memory for accented and unaccented voices." *Bulletin of the Psychonomic Society* 17: 217–220.
46. Köster, O. and N. O. Schiller (1997). "Different influences of the native language of listener on speaker recognition." *Forensic Linguistics* 4: 18–28.
47. Schiller, N. O. and O. Köster (1996). "Evaluation of a foreign language speaker in forensic phonetics: A report." *Forensic Linguistics* 3: 176–185.
48. Thompson, C. (1987). "A language effect in voice identification." *Applied Cognitive Psychology* 1: 121–131.

3.8 Testemunhas auriculares

49. Broeders, A. P. A. and A. C. M. Rietveld (1995). Speaker identification by earwitnesses. In A. Braun and J.-P. Köster (Eds.), *Studies in Forensic Phonetics*. Trier: Wissenschaftlicher Verlag, 24–40.
50. Bull, R. and B. R. Clifford (1984). Earwitness voice recognition accuracy. *Eyewitness Testimony: Psychological Perspectives*. G. L. Wells and E. F. Loftus. Cambridge, Cambridge University Press: 92–123.

51. Yarmey, A. D. (2001). “Earwitness descriptions and speaker identification.” *Forensic Linguistics* 8: 113–122.

3.9 Filas de reconhecimento com testemunhas auriculares

52. Broeders, A. P. A. and A. Amelvoort (1999). “Lineup construction for forensic earwitness identification”. *Proceedings of the 7th International Congress of Phonetic Sciences*. San Francisco: 1373–1376.
53. Hollien, H. et al. (1995). “Criteria for earwitness lineups.” *Forensic Linguistics* 2: 143–153.
54. Hollien, H. (1996). “Consideration of guidelines for earwitness lineups.” *Forensic Linguistics* 3: 14–23.

3.10 Detecção de mentiras usando fMRI e outras técnicas

55. Belin, P., S. Fecteau and C. Bédard. (2004). “Thinking the voice: neural correlates of voice perception”. *Trends in Cognitive Sciences*, 8, 129–135.
56. Knight, J. (2004). “The truth about lying”. *Nature*, 428(15 April 2004), 692–694.
57. Langleben et al. (2002). “Brain activity during simulated deception: an event-related functional magnetic resonance study.” *NeuroImage* 15(3): 727–732.
58. Lee, T. M. C., et al.. (2002). “Lie detection by functional magnetic resonance imaging”. *Human Brain Mapping*, 15, 157–164.
59. Pavlidis, I. and J. Levine. (2002). “Thermal image analysis for polygraph testing”. *IEEE Engineering in Medicine and Biology Magazine*, 21(6), 56–64.
60. Shah et al. (2001). “The neural correlates of person familiarity.” *Brain* 124(4): 804–815.

3.11 Generalização excessiva, charlatanismo e fraude

61. Hollien, H. (1987). “Voice stress evaluators and lie detection.” *Journal of Forensic Sciences* 32: 405–418.
62. Shipp, T. and K. Izdebski (1981). “Current evidence for the existence of laryngeal macrotremor and microtremor.” *Journal of Forensic Sciences* 26: 501–505.

ABSTRACT: This tutorial presents the field of forensic phonetics research. The text is divided in three parts. In the first part, forensic speech science is placed in its historical context and its modern development is briefly described. In the second part, the focus is on some of the fundamental issues in forensic phonetics research and how present knowledge can be applied in forensic fieldwork. This part reviews subjects such as human voice recognition and discrimination and how factors like memory, familiarity, language and vocal disguise influence these abilities. Lastly, attention is given to recent advancements in the field of lie detection using voice features and brain imaging and their scientific soundness is discussed.

Keywords: Forensic phonetics; Forensic speech science.