

## Explorando os dois lados do cérebro: a percepção holística da fala reversa (PHFR) sob a perspectiva forense

Gerson Albuquerque da Silva

Instituto de Criminalística do Estado de São Paulo -Superintendência da Polícia Técnico Científica de São Paulo

[gerson.gas@policiacientifica.sp.gov.br](mailto:gerson.gas@policiacientifica.sp.gov.br); tel +55 (11)987382379

---

### Resumo

Quando se reverte um sinal de fala, observam-se distorções em vários níveis linguísticos, bem como drásticas mudanças em suas características perceptivas. Informações fonéticas das estruturas segmentais e suprasegmentais são definitivamente perdidas. Contornos de energia e curvas de F0 são revertidas e, como resultado, o conteúdo da enunciação é obscurecido de tal forma que o que resta em nada lembra a língua original onde foram feitos os enunciados. Apesar das distorções, muitos parâmetros são mantidos, como, por exemplo, F0 médio, contorno de F0, valores médios dos formantes, espectro de longo termo e qualidade de voz. De acordo com os resultados da primeira pesquisa internacional sobre práticas de Comparação Forense de Locutor, a frequência fundamental e a qualidade de voz são os parâmetros suprasegmentais mais usados e, além disso, a qualidade de voz é considerada o parâmetro de maior poder discriminativo. A tarefa de reconhecimento da voz é um processo de interação entre a análise fonética componencial e uma tarefa de reconhecimento de padrão. Em virtude de se reter parâmetros importantes para tarefa de discriminação sem nenhum benefício de sugestões semânticas ou contextuais, a comparação de amostras reversas de fala é uma tarefa cujos processos psico – perceptivos subjacentes operam de forma holística. Nesse artigo, abordamos a percepção holística da fala reversa, de acordo com os dados científicos disponíveis, mostrando que a natureza ambígua desse estímulo, ao manter parâmetros altamente discriminativos, possibilita o seu uso na etapa perceptiva dos exames de Fonética Forense como procedimento alternativo em testes cegos por agrupamento.

*Palavras-Chave:* psicolinguística, neurolinguística, psicoacústica, Comparação Forense de Locutor; criminalística.

### Abstract

When a speech signal is reversed, there are distortions at various linguistic levels, as well as drastic changes in its perceptual characteristics. Phonetic information of the segmental and suprasegmental structures are definitely lost. Energy contours and F0 curves are reversed, and as a result, the content of the enunciation is obscured in such a way that what remains in no way resembles the original language in which the utterances were made. Despite the distortions, many parameters are maintained, such as F0 average, F0 contour, mean values of formants, long term spectrum and voice quality. According to the results of the first international survey on Forensic Speech Comparison practices, fundamental frequency and voice quality are the most widely used suprasegmental parameters and in addition voice quality is considered the parameter of highest discriminative power for almost the totality of the research participants. The task of voice recognition is a process of interaction between componencial phonetic analysis and a pattern recognition task. Because retaining important parameters for discrimination task with no benefit from semantic or contextual cues, comparison of speech samples is a pattern recognition task whose underlying psycho - perceptual processes operate in a holistic way (Gestalt processing). In this article, we approach the holistic perception of reverse speech, according to available scientific data, showing that the ambiguous nature of this stimulus, by maintaining highly discriminative parameters, can be used in Forensic Speaking Comparison caseworks, as an alternative procedure in blind grouping.

*Keywords:* psycholinguistics, neurolinguistics, psychoacoustics, forensic speaker comparison, criminalistics.

## 1. INTRODUÇÃO

A análise auditiva precede todas as etapas de um exame de Comparação Forense de Locutor. A abordagem perceptiva clássica envolve a oitiva crítica da gravação, procurando-se detectar quaisquer aspectos do sinal de fala que sejam distintivos ou notáveis [1]. Dada a variabilidade do sinal de fala, ao realizar essa tarefa perceptiva o perito sempre encontrará similaridades e diferenças entre os materiais confrontados [2]. Em relação aos traços convergentes, o perito levará em consideração o poder discriminativo de cada parâmetro observado. Já em relação às diferenças, o perito avaliará se é esperado que elas sejam observadas caso tenham sido produzidas por um mesmo falante (intravariabilidade) ou se são esperadas quando se leva em consideração a hipótese de terem sido oriundas de diferentes locutores (intervariabilidade). O perito examinará parâmetros segmentais e suprasegmentais. Dentre os primeiros, estão incluídas as características auditivas das vogais e das consoantes, por exemplo. Já dentre os últimos, estão incluídas a entoação e a qualidade de voz.

O método mais comumente aplicado durante a etapa perceptiva é o analítico [3]. Por meio dessa metodologia, a fala é dividida em seus constituintes mais elementares, os quais são investigados separadamente ou com relação à influência de um sobre o outro (processos coarticulatórios). Cada componente pode ser transcrito por meio do Alfabeto Fonético Internacional, no caso de segmentos acústicos, ou por meio de um protocolo de análise de qualidade de voz [4], [5], [6] e [7], dentre outros, para a documentação de propriedades notáveis relacionadas à qualidade de voz.

Além da metodologia analítica, o perito pode se valer de uma abordagem holística [3]. Nessa abordagem, o material de fala não é “segmentado”, mas considerado de forma global. Aqui, o que se procura é analisar os eventos de fala não de modo componencial, mas de forma global, concentrando-se na tarefa de se saber se as amostras são oriundas de um mesmo falante ou de

diferentes locutores.

Essas duas tarefas, a saber, a analítica e a holística (percepção Gestalt), são, como se verá, desempenhadas em hemisférios cerebrais distintos e de modo independente. Como será visto, tarefas relacionadas ao reconhecimento de padrões, por exemplo, são processadas no hemisfério direito do cérebro.

Os fundamentos da Gestalt foram estabelecidos por Wertheimer [8], [9], já no início do século passado. Trabalhando com exemplos visuais e auditivos, o autor mostra a coalescência dessas “formas” elementares em grupos e contornos, argumentando, ao cabo de seu artigo, que a experiência sensorial é organizada em padrões e não é aprendida como um registro descritivo dos elementos individuais.

No contexto forense, ao se deparar com amostras reversas o perito se vê diante de um estímulo descontextualizado do ponto de vista linguístico, que só pode ser explorado por meio de um processamento holístico. Nesse artigo, abordaremos a percepção Gestalt da fala reversa. Em seguida, baseando-se na literatura da área, mostraremos como essa percepção holística - doravante chamada de PHFR - pode ser incorporada em exames de Comparação Forense de Locutor, focando sua integração em testes cegos por agrupamento.

No item a seguir, será feita uma breve descrição das evidências que emergiram dos estudos neurológicos e comportamentais, as quais confirmam que há, de fato, um processamento paralelo da informação do conteúdo linguístico e de identidade do falante.

## 2. O HEMISFÉRIO DIREITO E O PROCESSAMENTO GESTALT

É de longa data a investigação dos centros cerebrais especializados na fala e na linguagem. Grande parte do conhecimento sobre a relação entre a organização do cérebro e a linguagem é oriunda de dados obtidos do estudo das lesões cerebrais. Dados que revelam a relação

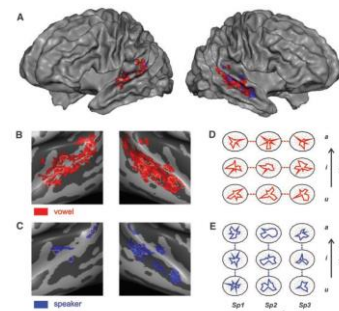
entre lesões no hemisfério esquerdo e déficits linguísticos podem ser encontrados já em 1836 nos trabalhos do médico escocês John Abercrombie [10]. No entanto, é atribuída a Paul Broca, neurologista e antropólogo francês, a consideração de que a perda de linguagem era muito mais frequente depois de uma lesão cerebral no lado esquerdo do que depois de uma lesão no lado direito do cérebro [11]. Hoje se sabe que os dois hemisférios cerebrais são em parte especializados para desempenho de diferentes tarefas. Assim, na maioria dos destros o hemisfério esquerdo é particularmente responsável pelas atividades analíticas, relacionadas ao raciocínio lógico e à linguagem, controlando, ainda, tarefas que progridem, passo a passo, em uma dimensão - como se sucede com nosso senso de progressão temporal, por exemplo<sup>1</sup>. Já o hemisfério direito é especializado, dentre outras tarefas, naquelas que envolvem a intuição, a criatividade, a localização espacial, bem como na realização de tarefas em três dimensões - por exemplo, a consciência de nossa posição em um espaço tridimensional -, e no processamento holístico [12], [13], [14].

Hoje em dia, as evidentes diferenças entre o processamento de vozes familiares e não familiares são bastante conhecidas. Descobriu-se, por exemplo, que as funções de reconhecimento e discriminação são tarefas distintas [15], [16], [17], e que, portanto, o desempenho em tarefas de discriminação é afetado por danos em qualquer um dos hemisférios cerebrais, enquanto que o desempenho em tarefas de identificação de uma voz é afetado por danos no hemisfério esquerdo do cérebro. As mesmas regiões envolvidas nas tarefas de discriminação foram mais tarde associadas com o processamento de vozes não familiares e as regiões correlacionadas com o reconhecimento de vozes com o processamento de vozes familiares, sendo ainda que os dois tipos de processamento são duplamente dissociados [18].

Informações relevantes tanto de cunho fonético quanto necessárias para a identificação do falante são

<sup>1</sup> De acordo com Ramunssen & Milner (1977), mesmo em 70% dos sestros (canhotos) o hemisfério esquerdo é dominante.

eficientemente representadas no cérebro. A figura 01 foi extraída de [19] em artigo em que os autores demonstram ser possível decifrar tanto o conteúdo da mensagem quanto a identidade do falante por meio da observação da atividade neural do ouvinte, medida por meio de ressonância magnética funcional. Nos gráficos em coordenadas polares, é importante observar que os padrões discriminativos das vogais (em vermelho) se distribuem bilateralmente na região cortical superior do lóbulo temporal enquanto que os padrões de reconhecimento do falante (em azul) se encontram confinados mais no hemisfério direito do cérebro. Esse hemisfério é especializado em reconhecimento de padrão.



**Figura 01.** Representações do falante e das vogais no cérebro.

Fonte: Formisano et al., (2008)

Ao fecharmos este item, é importante ter em mente que, embora a linguagem seja predominantemente lateralizada no hemisfério esquerdo na vasta maioria das pessoas- mesmo nos canhotos -, o processamento de informação do falante é parte integrante da percepção da fala e, em que pese o processamento do conteúdo linguístico e de reconhecimento do falante ocorram em regiões distintas do cérebro, eles se sobrepõem e se influenciam mutuamente.

No item a seguir, faremos uma pequena revisão e relato de experimentos que avaliam como se dá o processo perceptivo e que mostram que tanto o reconhecimento quanto a discriminação de falas (“diretas” ou reversas) são tarefas desempenhadas com relativa facilidade pelos seres humanos.

### 3. A PERCEPÇÃO HOLÍSTICA DA FALA REVERSA

Num estudo conduzido por Ladefoged & Ladefoged [20], J. Ladefoged gravou três tipos de amostras (a palavra “hello”, uma sentença simples e 30 segundos de fala contínua) com 29 membros da família e conhecidos do autor, intercalados com 13 pessoas cujas vozes esse último ouvira poucas vezes e 11 cujas vozes certamente nunca ouvira. 31% das 29 vozes familiares foram corretamente identificadas a partir da locução “hello”; em relação à sentença simples, 66% foram corretamente identificadas e 83% das vozes foram reconhecidas a partir dos 30 segundos de fala contínua. Este exercício confirma, categoricamente, a noção de que vozes familiares podem ser reconhecidas facilmente. De fato, reconhecimento de vozes familiares é uma prodigiosa habilidade humana. Pesquisas mostram que os seres humanos são capazes de reconhecer as vozes de suas mães quando ainda em seus ventres [21].

Durante o curso de formação de peritos criminais oficiais, enviamos, para um pequeno grupo, amostras de fala reversa de celebridades conhecidas pelos respectivos ouvintes. Antes do envio, solicitamos aos sujeitos que registrassem em uma folha de papel quais as celebridades cujas vozes lhes soavam mais familiares. Considerando as respostas, enviamos uma lista com imagens de celebridades nacionais (pessoas famosas do mundo do esporte, da política, da música, do entretenimento e do jornalismo) e perguntamos aos sujeitos qual(is) voz(es) lhes era(m) mais fácil(eis) de ser(em) lembradas - ou, em outras palavras, que lhe(s) soava(m) mais familiar(es) [22]. Além das celebridades listadas, os sujeitos poderiam sugerir outras personalidades que não constavam na lista (figura 02).



Figura 02. Algumas das personalidades apresentadas aos participantes da pesquisa

Com base nas respostas obtidas, amostras de fala que incluíam as selecionadas previamente pelos sujeitos eram, então, amostradas a 22.5kHz quantizadas a 16bits, revertidas e, por fim, apresentadas para os participantes. A estes era dada a tarefa de agrupar os áudios segundo o grau de similaridade (segundo a proposição de terem sido provenientes de um mesmo falante) em uma fila de reconhecimento aurial [23], podendo os participantes ouvirem as amostras durante quanto tempo quisessem. Após terem levado a termo a tarefa que lhes fora proposta, os peritos analisavam, por meio de uma escala contínua de 0 a 100, o grau de similaridade entre as amostras atribuídas a um mesmo falante (intravariabilidade) e o grau de diferença entre os áudios que foram atribuídos a diferentes falantes (intervariabilidade).

Você conseguiu associar os áudios "que lhe foram enviados com as respectivas celebridades?"

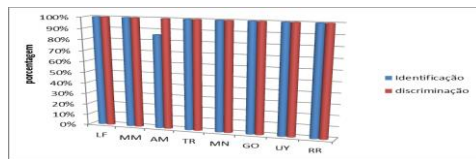
Item	Marcio Canuto	Silvio Santos	Lula	Palé	Nenhuma das celebridades das colunas anteriores
A1	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
A3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
A4	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
A5	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
A6	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
A7	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

[Click this button to continue](#)

Figura 03. Exemplo de agrupamento realizado por um dos participantes

**Figura 04.** Avaliação subjetiva da intervariabilidade

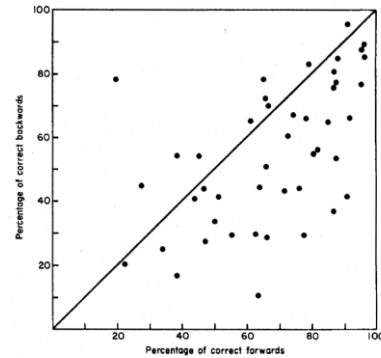
Inobstante o relato de que o estímulo era algo estranho e que pouco lembrava a fala humana – ou que lembrava uma língua que lhes eram desconhecidas –, obteve-se um grau de acerto de 100% em relação a todos os ouvintes na tarefa discriminativa (agrupamento de amostras similares) e somente um dos ouvintes não conseguiu identificar a voz de uma celebridade feminina, como se mostra na no gráfico da [figura 05](#). Ele nos informou, mais tarde, que quase nunca assistia ao programa televisivo em que a celebridade atuava.



**Figura 05.** Porcentagem de acerto nas tarefas de identificação e discriminação vozes reversas (Fonte: o autor)

O estudo informal acima descrito é uma variante da investigação levada a termo em [15], em que os autores investigaram como a habilidade dos ouvintes é afetada quando esses são submetidos a vozes reversas (estímulos de 4s apresentados “de trás para frente”). Nesse estudo, os autores descobriram que, não obstante a *performance* não ser homogênea, os ouvintes obtiveram, em média, um bom desempenho nas tarefas de reconhecimento. Tal como ocorre com a voz “natural”, no entanto, o reconhecimento de amostras revertidas de vozes familiares varia em função do falante e do ouvinte, sendo que algumas vozes são reconhecidas sem nenhum problema (tanto “natural” como quando revertidas), enquanto outras, não obstante possam ser facilmente

reconhecidas -quando não manipuladas - tornam-se irreconhecíveis quando revertidas. Embora bem menos frequente, havia ainda vozes que eram melhor reconhecidas quando revertidas. O gráfico da [figura 06](#) mostra as porcentagens de identificação correta de fala direta e revertida. A diagonal mostra igual índice de acerto.



**Figura 06.** Porcentagem de acerto nas tarefas de identificação e discriminação vozes reversas (Fonte: Van Lancker et al.)

Algumas observações podem ser feitas a partir dos estudos levados a efeito em [15]. Primeiramente, eles mostram que o reconhecimento de uma voz familiar se dá por meio de um processamento holístico (Gestalt). O fato de o reconhecimento de vozes ser essencialmente uma tarefa de reconhecimento de padrões é o que nos permite entender como as pessoas podem reconhecer os falantes mesmo quando suas vozes são revertidas.

Em segundo lugar, as autoras mostram que a familiaridade é condição para a identificação de falante. Em outras palavras, não é possível identificar um falante se não tivermos familiaridade com sua voz. É importante dizer que, ainda que “familiaridade” seja algo bastante subjetivo, ela parece estar relacionada com o tempo que o ouvinte foi exposto à voz em questão; ou, em outras palavras, “familiaridade” depende da quantidade de material de fala a que o ouvinte foi exposto.

Na tarefa de Comparação Forense de Locutor, por outro lado, o perito parte de uma amostra de voz que

não lhe é familiar e a familiarização é obtida por meio de uma ouvira crítica e reiterada, numa complexa interação entre um processo de reconhecimento de padrão e uma análise fonética componencial. Na medida em que o perito vai adquirindo mais familiaridade com a voz, essa interação migra num *continuum* de um inicial processamento *bottom up* para uma tarefa de reconhecimento de padrão [24].

No item 4, será feita uma apreciação acústica do sinal reverso de fala. Mostraremos, assim, que medidas acústicas que se relacionam com a qualidade de voz se mantêm inalteradas quando a amostra de fala é revertida. Os resultados ali apresentados são profundamente significativos em termos práticos, pois, como se sabe, a qualidade de voz é considerada o parâmetro de maior relevância fonético forense pelos próprios peritos foneticistas [25].

#### 4. MEDIDAS ACÚSTICAS EM UM SINAL DE FALA REVERTIDO

A reversão do sinal é uma das operações mais básicas que se pode fazer com um sinal. Matematicamente, tal operação pode ser representada pela seguinte relação [26]:

$$x_r[n] = x[-n],$$

em que [n] são as amostras do sinal digitalizado.

A figura 07 mostra um sinal simples e arbitrário amostrado e seu reverso:

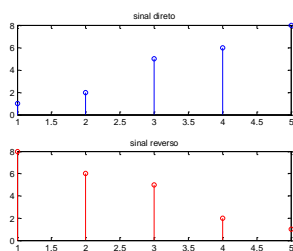


Figura 07. Sinais revertidos no tempo (o sinal em vermelho foi obtido a partir da reversão do sinal azul)

Em um sinal de fala, a reversão irá causar uma mudança em sua estrutura, mantendo-se, porém, importantes parâmetros acústicos. É de se esperar, portanto, poucas variações em relação aos parâmetros de fonte e filtro. Na figura 08, mostramos o fone [ɛ:] -à esquerda-, e a sua reversa -à direita-, produzidos, de forma sustentada, por um mesmo falante.

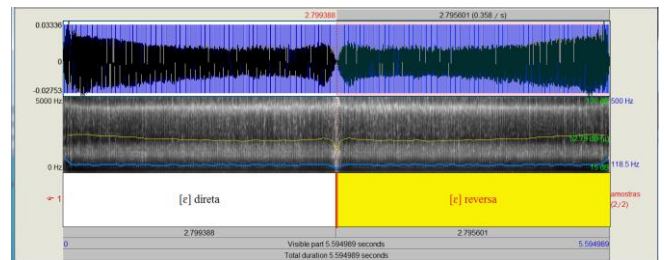


Figura 08. Fone [ɛ] e sua contrapartida reversa.

[ɛ:]	F0 (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)	HNR (dB)	Jitter (%)	Shimmer (dB)
Direta	106	391	1928	2419	22	0.544	0.244
Reversa	106	391	1928	2419	22	0.544	0.244

Tabela II. Algumas medidas acústicas relativas ao fone [ɛ:]

Os valores médios dos formantes e de F0, mostrados na tabela II, são iguais. É possível também verificar que outras medidas comumente associadas à qualidade de voz são coincidentes. A tabela mostra valores coincidentes tanto no que diz respeito à relação harmônico-ruído (HNR, que mede a relação entre a componente periódica - oriunda da vibração das pregas vocais - e aperiódica do sinal - oriunda do ruído glotal), quanto em relação ao jitter (relacionado à perturbação da frequência fundamental ciclo a ciclo) e ao shimmer (perturbações da amplitude da frequência fundamental ciclo a ciclo).

Dada a compatibilidade que se verifica quando são comparadas medidas de filtro, espera-se que encontremos convergências quando submetemos os áudios a sistemas automáticos de comparação de locutor,

uma vez que os coeficientes mel-cepstrais utilizados em sistemas automáticos de comparação de locutor estão, em certa medida, relacionados com as ressonâncias nas cavidades supralaríngeas [27], que são preservadas numa operação de reversão. Assim, dependendo da qualidade das amostras, as diferenças que se apreciam em vozes reversas são mínimas quando submetidas a sistemas automáticos. Quando comparamos amostra(s) padrão, obtida(s) em canal microfônico, com amostra(s) questionada(s) oriunda(s) de um telefone celular, há, em geral, um valor menor de LR atribuído à(s) amostra(s) questionada(s) revertida(s) em relação à(s) amostra(s) “direta(s)”, e os valores são tanto mais discrepantes para valores críticos de relação sinal-ruído (menor do que 15dB) e de duração (menor do que 60s). Mas, na maior parte dos casos, mesmo em condições acústicas pouco favoráveis, a comparação de um material padrão com amostras questionadas reversas e diretas fornece resultados compatíveis entre si.

A [tabela IV](#) mostra valores bastante próximos de LR, como resultado de comparações entre uma amostra de referência, obtida de um depoimento dado em juízo em um ambiente não isolado acusticamente - e, conforme mostrado na [Tabela III](#), apresentando duração e razão sinal/ruído muito distantes do ideal - e uma mesma amostra questionada - que também foi revertida.

Amostra	Tempo de Fala Líquida (s)	Relação Sinal / Ruído (dB)
Padrão	30	24
Questionada Direta	71	20
Questionada Reversa	71	20

**Tabela III.** Tempo de fala e relação sinal-ruído de amostras padrão (direta) vis-à-vis questionada (direta e reversa)

Valor da Razão de Verossimilhança		
Amostra	Questionada Direta	Questionada Reversa
Padrão (Direto)	33.82	33.76

**Tabela IV.** Valores de razão de verossimilhança resultante do confronto entre uma amostra padrão (direta) vis-à-vis uma questionada (direta e reversa) oriunda de um mesmo falante

Percebe-se, portanto, que muitos parâmetros são mantidos quando se reverte um sinal de fala, e, dentre eles, estão os que se relacionam com a qualidade de voz.

A essa altura, já estamos prontos para apreciar as duas principais linhas de pensamento acerca da natureza da qualidade de voz, bem como entender de que forma a percepção da voz reversa não pode ser apreendida senão por um mecanismo de reconhecimento de padrão que opera por Gestalt. É o que faremos no item a seguir

## 5. PHFR NA QUESTÃO DA QUALIDADE DE VOZ EM CONTEXTO FORENSE

Qualidade fonética se refere àqueles aspectos do som da fala que veiculam informação linguística – mais especificamente informação fonológica [2]. São as realizações fonéticas de unidades linguísticas como consoantes e vogais. Por exemplo, a qualidade fonética do fonema /o/ na palavra <vocal> falada por um paulista pode ser descrita como uma vogal meio fechada (média-alta) posterior arredondada [o]. Há outro alofone possível no Português Brasileiro. Para um falante de Recife, por exemplo, a qualidade fonética da mesma vogal pretônica será descrita como uma vogal meio aberta (média-baixa) posterior arredondada [ɔ].

A qualidade fonética não se limita a segmentos acústicos, como consoantes e vogais, mas também abrange categorias linguísticas suprasegmentais, como entonação, tonicidade e ritmo. Assim, a diferença de tonicidade entre <análise> e <análise> também constitui um aspecto da qualidade fonética, pois sinaliza a diferença linguística entre um verbo e um substantivo.

Qualidade de voz, por seu turno, é o que se pode ouvir quando a qualidade fonética é removida, como, por exemplo, quando alguém pode ser ouvido falando em um cômodo adjacente, mas o que ele/ela está realmente dizendo não é de todo inteligível. A filtragem passa - baixo exercida pelas paredes elimina toda

informação em nível segmental [27], restando apenas as propriedades prosódicas, isto é, relacionadas com a evolução temporal da frequência fundamental, da duração e da intensidade do sinal de fala.

Qualidade da voz se refere, portanto, às propriedades suprasegmentais resultantes da configuração do trato vocal. Tal termo -embora se refira, amíúde, a propriedades supraglóticas, como nasalidade, por exemplo - é frequentemente utilizado também para descrever tipos de fonação (voz soprosa, falseto, creaky voice, entre outros) [28]. Esse termo se refere àquelas características presentes mais ou menos durante todo o tempo em que uma pessoa está falando [29], algo como a soma da configuração do trato vocal, da anatomia da laringe e de uma componente “aprendida” - de cunho social, portanto [30].

O método perceptivo analítico se baseia no fato de que os sons da fala são entidades estruturadas de forma componencial e, como tais, podem ser tanto analisados como descritos em termos de seus componentes [31]. Assim, podermos, por exemplo, descrever uma consoante como fricativa labiodental vozeada ou uma vogal da forma como descrevemos mais acima, qual seja, como meio fechada posterior não arredondada.

O protocolo de Laver fornece uma terminologia bastante útil acerca da qualidade de voz em relação a ajustes como levantamento/abaixamento da laringe, arredondamento dos lábios, abertura da mandíbula, posição da ponta, da lâmina e do dorso da língua, constrição ou expansão da faringe, posição do véu e o estado da glote. O sistema de descrição é componencial e, desta forma, assim como uma consoante pode ser descrita como “alveolar plosiva vozeada”, a qualidade de voz pode ser descrita, no arcabouço de Laver, como “palatalizada, denasal creaky voice”, por exemplo [32].

No arcabouço teórico proposto em [7], a qualidade de voz é entendida como tendo dois componentes: um componente orgânico e um componente

de ajustes (configuração ou *settings*). O componente orgânico refere-se a aspectos do som que são determinados pela anatomia e fisiologia do trato vocal do falante, como o comprimento do trato vocal ou o volume de sua cavidade nasal, e sobre a qual o falante não têm controle. A configuração anatômica de um falante tipicamente impõe limites à gama de características vocais - um bom exemplo de uma característica anatomicamente determinada seriam os limites superior e inferior da faixa de frequência fundamental de um locutor.

O segundo componente da qualidade de voz - frequentemente chamado de ajustes ou configuração articulatória -, refere-se aos ajustes musculares habituais que um indivíduo adota quando fala. Um falante pode habitualmente falar com lábios ligeiramente arredondados, com muita nasalização e uma faixa preferencial de uso de frequência fundamental baixa. Como esses recursos de ajustes são deliberadamente adotados, eles diferem do primeiro componente pelo fato de que, aqui, eles estão sob o controle do locutor.

Os componentes dos ajustes articulatórios de um indivíduo (por exemplo, arredondamento de lábio) são concebidos como desvios de uma configuração neutra idealizada do trato vocal. Um falante pode falar com o corpo da língua ligeiramente para trás e para cima a partir de uma posição neutra, resultando no que é descrito como voz uvularizada (o desvio está na direção da úvula). Isso significaria que todos os segmentos sonoros suscetíveis de serem influenciados pelo ajuste articulatório são articulados mais para trás e levemente mais altos que o normal.

Como já mencionamos, o fato de a qualidade de voz ser um dos parâmetros que se mantém durante o processo de reversão da fala é altamente significativo, pois, de acordo com a primeira pesquisa internacional sobre práticas em Comparação Forense de Locutor, qualidade de voz é o parâmetro fonético forense com maior poder discriminativo [25]. Ali vemos que, não obstante o fato de 94% dos participantes afirmarem que



o exame da qualidade de voz fizesse parte de seus procedimentos globais de análise, apenas 61% o faz utilizando-se de algum protocolo de análise reconhecido (como o proposto por Laver ou alguma versão modificada de seu *Vocal Profile Analysis Protocol*, a exemplo do elaborado pelos linguistas da Universidade de York, o qual possibilita a análise de 32 ajustes dentre os supralaríngeos, fonatórios, velofaríngeos e de tensão muscular).

Para muitos pesquisadores, entretanto, o uso de tais protocolos é de alcance bastante limitado em casos forenses reais. Embora considere o protocolo de Laver a mais abrangente ferramenta para descrição da qualidade da voz, consultando 30 casos forenses ocorridos nas Ilhas Britânicas de 1988 a 2002, Nolan notou que as descrições feitas pelos foneticistas forenses em seus laudos periciais acerca da qualidade de voz se limitavam a menções de ocorrências de nasalização ou levantamento da laringe, por exemplo, ou, simplesmente, eram observações do tipo: “há similaridades na qualidade de voz” [32]. Ao procurar as causas do uso limitado da análise componencial da qualidade de voz, Nolan concluiu que, para além da falta de treinamento, do desconhecimento, por parte dos peritos, acerca da disponibilidade de tal ferramenta e do fato de ser uma tarefa que consome bastante tempo para ser levada a efeito, o principal empecilho para a análise componencial da qualidade de voz se assenta na própria qualidade das amostras, bastante alteradas em virtude das distorções introduzidas pela transmissão telefônica, que é o cenário típico de um caso real de fonética forense.

Nolan exemplifica mostrando como o canal telefônico causa distorções no primeiro harmônico. Sabe-se que, enquanto os harmônicos mais altos estão relacionados com a descontinuidade que ocorre com o impacto das pregas vocais, a amplitude do primeiro harmônico está associada à excursão lateral da laringe e sua diferença em relação ao segundo harmônico está inversamente correlacionada com o coeficiente de contato, sendo um bom indicativo de sopro [33],

[34], [35], [36]. Além disso, também o ruído aditivo introduzido pelo canal telefônico pode interferir com a percepção de sopro, por exemplo. Nolan argumenta ainda que a qualidade de voz pode ser determinada, em parte, por diferenças dialetais ou variações linguísticas, e não ditadas unicamente pela intervareabilidade, a qual é o foco do exame de Comparação de Locutor.

Em nossa opinião, os argumentos de Nolan não invalidam o uso de protocolos de qualidade de voz uma vez que nem toda casuística em Fonética Forense se dá por meio de interceptações telefônicas - mesmo nesse tipo de canal, muitos traços associados à qualidade de voz podem ser inspecionados pela observação instrumental - *creaky voice*, por exemplo, pode ser constatada pela observação dos pulsos glóticos. Acreditamos, ainda, que há muita propriedade na asserção de que as distorções introduzidas pelo canal telefônico degradam a qualidade de voz, mas esta nem sempre se perde em absoluto.

Há, por outro lado, um grupo de pesquisadores para os quais a qualidade de voz não pode ser descrita por meio de seus componentes. Tais pesquisadores são respaldados por um considerável corpo de dados comportamentais e neurológicos [37], [38], [39], [40], [41], [42], [43] e [44], por exemplo, que mostram que os ouvintes percebem a qualidade de voz como um padrão que só pode ser apreendido de forma integral e não como uma lista de descritores. Destarte, o motivo pelo qual a descrição da qualidade de voz por seus componentes básicos é problemática repousa em sua própria complexidade [24]. Sob tal perspectiva, resta claro que a qualidade de voz não pode ser descrita adequadamente de forma analítica, pois ela é mais do que simplesmente um sinal acústico: assim como os parâmetros psicoacústicos *pitch* e *loudness*, a qualidade de voz resulta de uma interação entre o ouvinte e o sinal de fala, e, sendo assim, é um processo não passivo, porquanto altamente dependente do ouvinte.

A limitação da análise de perfil vocal é que ela modela a percepção em termos do processo de produção

da fala, mas sem referência estabelecida ou documentada quanto ao ouvinte. Ou seja, ao descrever a qualidade de voz em termos detalhados da suposta configuração fisiológica subjacente, a análise de perfil vocal indica onde a informação perceptual acerca da qualidade de voz *pode* ser encontrada. Tal análise não indica, entretanto, *quais* dos vários aspectos especificados são mais significativos ou, de fato, perceptíveis aos ouvintes - *como* os ouvintes usam diferentes informações a respeito dos ajustes para avaliar a qualidade, *se* alguns aspectos são mais importantes que outros ou *de que forma* as diversas dimensões em um *setting* interagem perceptualmente [24]. Mesmo a suposição de que os ouvintes têm habilidades para separar os diferentes componentes de um *setting* de modo perceptivo é algo, segundo os autores, altamente questionável. Mais do que isso, os ouvintes parecem largamente incapazes de isolar dimensões elementares de uma amostra de fala [40].

Assim, tais esquemas de descrição de qualidade vocal supõem um espaço perceptivo para qualidade de voz bem definido, aplicável a todas as vozes e verdadeiro para todos os ouvintes, os quais o explorariam de um mesmo modo [24]. Todas as evidências e considerações teóricas, ao contrário, indicam que os ouvintes exploram o sinal da fala de modo único. Por fim, pode-se notar que, sob esse enfoque, o processo de reconhecimento de voz partilha da mesma propriedade do reconhecimento facial, no sentido de que, assim como a face é um complexo padrão visual, a voz é um complexo padrão auditivo [24].

Sob a perspectiva de um exame fonético forense interessa-nos o fato de que, quando o sinal de fala é revertido, há drásticas mudanças do ponto de vista perceptivo. A informação é distorcida em vários níveis de linguagem (fonético, fonológico, semântico), de tal forma que o estímulo outrora estruturado da emissão normal se converte, quando revertemos o sinal de fala, em um estímulo desestruturado e ambíguo - para muitos, algo bastante difícil de se processar [15] e que não pode ser examinado senão por Gestalt. Ou seja, na ausência de referentes, a PHFR começa a se desenvolver sem a

participação de certos elementos que caracterizam as tarefas perceptivas clássicas, de tal forma que os princípios que regem a separação dos distintos objetos auditivos da análise de cena auditiva não chegam a se realizar [45], e o material de fala só pode ser acessado de forma holística. A percepção da fala reversa é, portanto, inevitavelmente holística, tarefa essa que é, como mostramos, desempenhada unicamente pelo lado direito do cérebro.

Neste ponto, faz-se importante nos determos na observação de que, para além dessa natureza holística, a fala reversa é descontextualizada do ponto de vista linguístico. Amostras revertidas de fala, ao promover a neutralização das pistas relacionadas aos padrões articulatorios, ela elimina os efeitos oriundos do processamento *top down*, os quais são tanto mais frequentes quanto mais ruidosa for a gravação analisada. Assim, o uso de amostras revertidas ajuda a diminuir o viés cognitivo uma vez que coloca o ouvinte diante de um estímulo descontextualizado do ponto de vista linguístico, explorando tão somente o prodígio humano de reconhecer vozes mesmo que elas lhe sejam apresentadas de forma revertida.

Partindo das observações acima, mostraremos adiante de que forma a PHFR, ao se constituir em resposta a um estímulo linguisticamente descontextualizado, pode ser utilizada como um método para redução do viés cognitivo nos exames de Comparação Forense de Locutor. Como a questão do viés cognitivo perpassa todas as disciplinas da Criminalística, antes de focarmos na utilização da PHFR nos exames de Comparação Forense de Locutor, faremos uma breve revisão da literatura que trata do viés cognitivo em Criminalística

## 6. O VIÉS COGNITIVO NAS DISCIPLINAS CRIMINALÍSTICAS

Nas ciências sociais, o termo “viés” é, por assim dizer, “neutro”. Descreve meramente uma situação em que quaisquer erros constatados tendem a ser realizados em uma ou outra direção, em oposição a um erro não aleatório, onde há direcionalidade do erro. Chamamos de viés cognitivo a um padrão de desvio no julgamento que pode levar a desvios sistemáticos de lógica e a decisões irracionais.

É de longa data o reconhecimento de que vieses cognitivos são tendências inerentes à nossa espécie, noção esta implícita na frase *“fere libenter homines id quod volunt credunt”* (“os homens acreditam de bom grado naquilo que desejam acreditar”), atribuída a César [46]. Para a literatura científica dos últimos séculos, o viés cognitivo é descrito como oriundo de nossa necessidade de processar eficientemente o descomunal fluxo de informação que recebemos do mundo externo. Não tivéssemos algum sistema de categorização dessas informações, o mundo se tornaria insuportavelmente confuso [47]. É provável, portanto, que muitos dos vieses cognitivos a que estamos susceptíveis sejam resultado de condições adaptativas que nos propiciaram evoluir como espécie. Em alguns contextos, vieses podem ser até desejáveis quando vão na direção em que o erro em questão é menos custoso do que suas alternativas. Esse necessário sistema de categorização, interpretação e seletividade pode, entretanto, envolver extremo desgaste em certas áreas da atividade humana; particularmente, naquelas disciplinas em que há muito grau de subjetividade e de interpretação humana.

A expressão “viés de confirmação forense” abrange uma gama de processos psicológicos que podem influenciar a condução, decisão, julgamento e interpretação do perito. Esses processos incluem a

exposição a informações contextuais ou incriminadoras do suspeito, fatores emocionais como o desejo de “achar o culpado”, de “encerrar o caso”, o medo de errar, a satisfação de se sentir recompensado cognitivamente [48], ou de concordar com o cliente [49], por exemplo.

Dentre os vieses mais recorrentes no mundo forense estão o:

- viés de confirmação: a tendência de se lembrar, interpretar ou pesquisar por informações de maneira a confirmar crenças ou hipóteses iniciais;
- viés de contexto (ou contextual): a tendência de uma interpretação ser influenciada por informação estranha ao caso;
- viés de expectativa: também conhecido como viés do experimentador, onde o perito tende a analisar dados que concordam com suas expectativas para o caso e descartar ou desvalorizar evidências conflitantes com suas expectativas.

Nunca se deve deixar de observar que, em várias circunstâncias, alguns desses vieses tendem a intensificar os vieses já existentes. É o que ocorre, muitas vezes, com o viés de expectativa. Quando as pessoas são levadas, pelas circunstâncias, a esperar por algum fato ou condição, elas tendem a perceberem apenas tal fato ou condição quando se defrontam com situações ambíguas do ponto de vista informacional e isto pode levar a erros enviesados em direção da expectativa. Quando aquilo que a pessoa espera ver é resultado de sua própria geração de hipótese, teorias ou cenários sobre o que tal ou qual caso deva ser, o investimento de caráter pessoal nessa hipótese vai necessariamente reforçar a tendência de perceber ou valorizar as informações que confirmem suas expectativas, esquecendo ou desvalorizando irracionalmente as informações que não confirmem tal hipótese. Além disso, quando tais hipóteses enviesadas

são formuladas por pessoas de status superior em uma equipe de trabalho, ou quando estão envolvidos fatores ligados à autoestima ou motivacionais (busca de alguma recompensa cognitiva ou medo de alguma sanção), o viés de confirmação tende a ser amplificado. Como resultado, mesmo a mais óbvia e inequívoca informação, caso não confirme a expectativa, tende a ser desacreditada, desvalorizada ou mesmo esquecida.

Não nos surpreende, portanto, a constatação segundo a qual, de maneira geral, o mundo forense é altamente suscetível ao enviesamento cognitivo [50]. O fato de muitas vezes estar informado acerca da natureza do crime em análise - e de se saber de detalhes acerca do mesmo -, de se sentir pressionado institucionalmente, de ser chamado ao tribunal como testemunha de uma das partes em um sistema legal baseado no contraditório - ou mesmo se sentir como membro de uma das partes -, o efeito *CSI*, além do excesso de confiança que os próprios peritos têm em relação às suas habilidades - todos esses fatores são considerados potenciais fatores de enviesamento. E isto é particularmente relevante naquelas abordagens que envolvem interpretação subjetiva do exame (isto é, aquelas que não são baseadas em base de dados representativos de uma população relevante e em modelos estatísticos). É a situação com a qual nos deparamos, por exemplo, em um exame de cunho puramente auditivo- perceptivo. Conforme mostraremos a seguir, a PHFR representa uma boa ferramenta para se lidar com o viés cognitivo, o qual é, em certa extensão, ainda negligenciado nos exames de Comparação Forense de Locutor.

## 7. PHFR NA REDUÇÃO DO VIÉS COGNITIVO

Comparados com outros campos da fonética forense, como a identificação por testemunhos auditivos [51] ou análise de locuções disputadas e transcrições [52], [53], [54], [55], são recentes - e ainda escassos - os estudos voltados para a redução de enviesamento cognitivo nos exames de Comparação Forense de Locutor

[56], [57], [58]. Há, por outro lado, várias razões para se suspeitar que o exame de Comparação de Locutor seja bastante suscetível a vieses cognitivos.

Diferentemente do que ocorre em outras disciplinas, os foneticistas forenses se deparam, logo de início, com um mecanismo perceptual que tende a estabelecer uma hipótese inicial [58]. Isso pode levar à “visão de túnel“ (um tipo de viés de confirmação no qual os atores do sistema criminal se concentram em um suspeito, selecionando e filtrando as evidências que confirmem sua hipótese em relação a ele e, ao mesmo tempo, ignorando ou mesmo suprimindo evidências que não confirmem tal expectativa [59]).

Nem sempre o que ouvimos está contido no sinal da fala que atinge nossas orelhas. Antes, o que ouvimos pode ser suprido por nossos cérebros a partir do nosso conhecimento acerca da estrutura da mensagem ou de seu contexto não linguístico [2]. Esse fenômeno, em que a mensagem é suprida pelo cérebro - chamado de processamento *top down* -, é responsável tanto pela Restauração Fonêmica [60], como pelo Efeito de Expectativa [61], os quais serão, a seguir, brevemente detalhados.

Em [60], o autor expõe o ouvinte a amostras de fala com trechos digitalmente suprimidos. Warren o faz repetidamente e de modo progressivo, com um crescente aumento do nível do ruído de fundo. Em determinado nível, o ouvinte vai se tornando mais consciente dos trechos que já não eram audíveis e seu cérebro, a fim de outorgar um sentido semântico, trata de reconstruir os sons suprimidos com fonemas que podem ser ou não coincidentes com os elididos.

Em [61], o autor nos mostra como os ouvintes podem ser afetados pelo viés de expectativa. Selecionadas dentre um conjunto de 25 pessoas de seu laboratório pessoal, um subconjunto de onze vozes foi gravado a partir de um telefone fixo, sendo cada uma constituída de 30 segundos de fala espontânea. Uma décima segunda voz gravada - e que fora incluída no experimento -, pertencia a uma pessoa que não fazia parte

do conjunto das 25 pessoas que trabalhavam no laboratório. Nove dentre os dez ouvintes identificaram corretamente todos os onze membros do laboratório. O décimo ouvinte confundiu duas pessoas com as quais guardava menor grau de familiaridade. Metade dos ouvintes atribuiu erroneamente a voz da pessoa que não pertencia ao laboratório a um integrante do grupo que, entretanto, não fazia parte do subconjunto de vozes gravadas. Esse experimento mostra claramente que a expectativa afeta o desempenho da tarefa de reconhecimento.

Testes cegos são os procedimentos mais básicos para a eliminação de vieses cognitivos nas diversas disciplinas da criminalística, pois não há motivos para o perito receber informações estranhas à sua análise. A seguir, mostramos como amostras revertidas de fala podem ser integradas a testes cegos por agrupamento.

Na [figura 8](#), apresentamos uma fila de reconhecimento aural com seis amostras revertidas de 15s, obtidas a partir de um caso real em que utilizamos o arcabouço criado pelos peritos do Instituto Nacional Forense da Holanda [62]. Primeiramente, tais amostras “revertidas” são apresentadas ao perito que realizará o teste cego (essa fase é conhecida como fase de familiarização ou retenção), o qual pode ouvi-las tantas vezes quanto queira. De maneira geral, o material se constitui do áudio questionado, do material padrão e de distratores (que podem ser formados pelos locutores cujas conversas foram interceptadas, mas que não são os suspeitos ou acusados de terem cometido o crime). O perito que realizará a análise não possui nenhuma familiaridade com o caso, não possui nenhuma informação acerca dos materiais recebidos e nem sequer sabe, de antemão, o número de falantes que se encontra no material enviado – observe-se que, por meio desse procedimento, tanto o viés de contexto como o viés de confirmação ou de expectativa são dirimidos ou mesmo eliminados.

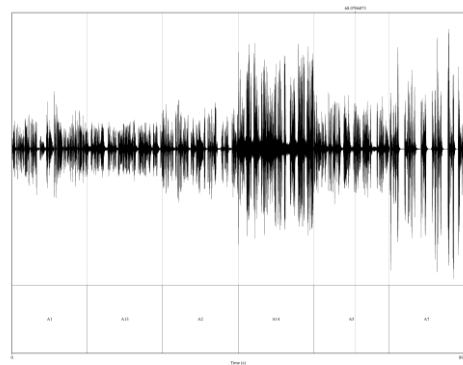


Figura 08. Fila de reconhecimento aural

O perito é, então, desafiado a agrupar as vozes que ele atribui a um mesmo falante. No exemplo da [figura 12](#), o analista considerou as amostras de fala A1 e A2 e as amostras A13 e A14, respectivamente, similares. Em relação às amostras restantes - A5 e A7-, o perito não as achou similares entre si e, por isso, colocou-as em camadas distintas.

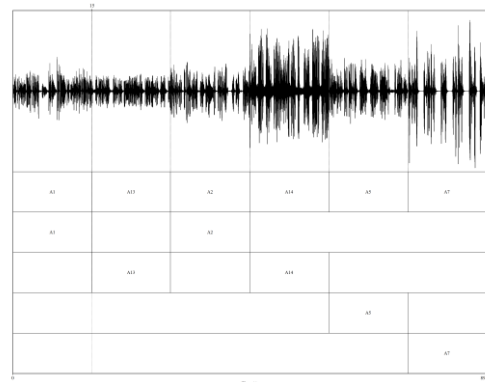


Figura 09. Classificação em fila de reconhecimento aural

Terminado o agrupamento, o analista é ainda desafiado a dizer o grau de similaridade intrafalante (no modelo adotado pela NFI - Netherlands Forensic Institute - é assinalado segundo uma escala subjetiva “três pontos”: “fraca”, “normal” ou “forte”). Ele também é encorajado a atribuir uma valoração subjetiva de nível de certeza em relação à variação interfalante. Por fim, o analista pode avaliar, por exemplo, se seu nível de certeza acerca de A13 e A14 - pertencentes a diferentes falantes - é alto,

baixo, normal ou, ainda, se não se é possível fazer quaisquer julgamentos. Esses procedimentos visam à avaliação do nível de consistência intrafalante e interfalante, respectivamente.

Se o propósito de um teste cego é fazer com que o analista não tenha acesso a informações contextuais, o agrupamento cego utilizando-se da PHFR representa um passo além na medida em que a descontextualização da informação se dá, de forma drástica, em termos da desestruturação do sinal de fala em vários planos linguísticos. A análise cega, conduzida nesses moldes, obriga, por assim dizer, que o perito desempenhe a tarefa comparativa de modo holístico tão somente, registrando sua real capacidade de (aprender a) agrupar as amostras de interesse do caso forense em questão quando apresentadas entre outras vozes de uma população relevante. Além disso, o arcabouço permite ao perito avaliar o nível de certeza que o mesmo pode auferir quando desempenha tal tarefa sem informação contextual e desestruturada em vários níveis linguísticos, utilizando-se tão somente da habilidade do lado direito de seu cérebro em tarefas de reconhecimento de padrão.

Note-se que o resultado de um teste cego por agrupamento, utilizando-se de PHFR é, como qualquer teste cego, um grupo de fragmentos formado a partir de um processamento de reconhecimento de padrão. Desta forma, além de eliminar vieses cognitivos, tais testes são realizados unicamente de maneira holística. Considerando os estudos mencionados ao longo deste artigo - os quais mostram que a qualidade de voz não pode ser caracterizada segundo uma lista de descritores devido a sua natureza complexa [24] -, a PHFR integrada em testes se revela uma abordagem coerente para se lidar com a qualidade de voz.

Embora peritos foneticistas treinados possam fazer testes cegos sem aparente dificuldades somente se valendo de amostras revertidas de fala, é altamente recomendável que, em sendo possível, proceda-se a testes duplos [50]. Um modo interessante de se fazê-los é

promovendo -os em duas etapas. Na primeira, pode-se utilizar de 10 a 20 fragmentos com um máximo de 20s de duração de fala natural. Na segunda etapa, utiliza-se material de fala reversa - no máximo, a metade do número de fragmentos utilizados na primeira etapa. Esses fragmentos podem ser oriundos da primeira etapa ou alguns sobre os quais se deseja alguma confirmação a respeito (áudios questionados e de referência do suspeito, por exemplo). Por meio deste procedimento duplo, o analista cego pode explorar sua capacidade de extrair informações do falante tanto por meio de uma análise linguística quanto também por meio de sua habilidade natural de reconhecimento de padrões. É sempre importante pontuar que quando se fala da utilização da PHFR em testes cegos, estamos nos referindo a um procedimento suplementar à análise e que, sendo assim, deve ser levado a termo sempre obedecendo à independência e discricionariedade do perito. Vale a pena dizer que, no entanto, enquanto procedimento padrão, embora os testes cegos costumem ser procedimentos complexos que consomem muito tempo para ser realizados, eles integram as melhores práticas em fonética forense em vários institutos, como o Centro Forense Nacional da Suécia e o grupo de áudio e fala do Instituto Nacional Forense da Holanda [62], [63], atendendo aos pré-requisitos de acreditação laboratorial demandada na norma ISO17025.

## 7. CONSIDERAÇÕES FINAIS

Este artigo procurou mostrar que o uso da PHFR em testes cegos por agrupamento é um procedimento duplamente eficiente. Primeiramente, porque contorna a complexidade inerente à análise componencial da qualidade de voz – apontada, conforme expusemos, por diversos pesquisadores [24] – desobrigando o analista cego de abordar a qualidade de voz por meio de uma lista de descritores [62]. Em segundo lugar, a PHFR elimina o viés de confirmação e de expectativa, processos

cognitivos em relações aos quais a tarefa perceptiva clássica é bastante susceptível.

É importante pontuar que a PHFR pode ser útil em vários contextos fonético-forenses e sua utilização em testes cegos por agrupamento é apenas uma dessas possibilidades [64]. Recentemente, precisamos anonimizar nossa base de dados de material de fala para um experimento envolvendo sistemas computadorizados de Comparação de Locutor, e a melhor solução foi reverter as amostras de fala. Por meio desse procedimento, mantivemos inalterados dados acústicos importantes, ao mesmo tempo em que preservamos a identidade dos sujeitos utilizados no experimento.

De forma geral, sempre que o conteúdo linguístico puder enviesar o analista, o uso de amostras reversas se mostra uma boa opção. Quando a amostra de referência (material padrão) e o material questionado contém amostras em duas línguas diferentes<sup>2</sup>, por exemplo, testes cegos por agrupamento utilizando-se da PHFR revelam-se procedimentos suplementares bastante úteis.

Deve-se sempre considerar, entretanto, que, embora sejam bastante úteis para tarefas de discriminação, baseadas em julgamentos de similaridade, a tipicidade, entretanto, não é apreendida em uma abordagem holística [32]. De maneira geral, abordagens holísticas não respeitam os princípios bayesianos que devem orientar as disciplinas da criminalística que lidam com a questão de se determinar se duas amostras são ou não oriundas de uma mesma fonte [3]. Assim, por exemplo, a menos que o perito seja falante do dialeto em questão, ele pode ser enganado por uma similaridade mais característica desse dialeto do que devido a uma idiossincrasia do falante. Embora a PHFR faça com que parte das informações fonético-articulatórias ligadas ao dialeto seja perdida, ela não consegue indicar se as

propriedades suprasegmentais que se preservam são relacionadas a variações diatópicas ou se são específicas do locutor.

Não se deve esquecer, entretanto, que a análise perceptivo-auditiva é apenas parte de uma abordagem mais ampla que inclui também a análise acústico-instrumental [65], [66], [67]. Quando complementada por meio do uso de sistemas automatizados baseados em modelos estatísticos, também de natureza holística [3], o exame se torna não só mais acurado, transparente e cientificamente objetivo, como também mais robusto e resistente a vieses cognitivos. Por tudo isso, o método híbrido para Comparação Forense de Locutor (análise combinada mais o uso de sistemas automáticos de reconhecimento de locutor) é, hoje em dia, reconhecido universalmente como padrão de excelência nos exames de Comparação Forense de Locutor, sendo adotado nas melhores instituições policiais do mundo, confirmando a tendência já apontada em [25].

A PHFR, como complemento ou enfoque suplementar, deve, portanto, compor perspectivas mais amplas de análise fonético forense, onde a percepção Gestalt deve ser utilizada sempre de forma integrada à metodologias fonético analíticas tanto em relação aos parâmetros linguísticos (as características auditivas dos segmentos acústicos, por exemplo) quanto não linguísticos (como a qualidade de voz, por exemplo - a qual, como vimos, é, por meio da PHFR, apreendida de forma holística e por meio do uso de protocolos perceptivos, analisada em seus elementos constituintes).

Ao se utilizar de todo o ferramental que lhe é disponível na tarefa perceptivo comparativa, o perito está também explorando ao máximo todo o potencial de seu cérebro - literal e metafóricamente falando -, unindo sua destreza adquirida por treinamento e experiência à sua habilidade natural de reconhecimento de vozes.

---

<sup>2</sup> Comunicação pessoal de Dr. G.M. Cambier-Langeveld, *expert* em fonética forense e linguística forense do grupo de áudio e fala do NFI (Netherlands Forensic Institute) em 25 de agosto de 2017, recebida por correio eletrônico.

## 8.REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Morrison, G.S. *Forensic voice comparison*. In Expert Evidence 99, 1ed. London : Thompson Reuters, 2010. Chap.99, p.1051
- [2] Rose, P. J. (2002). *Forensic Speaker Identification*, Taylor & Francis , London & New York..
- [3] Jessen, M. (2008). “Forensic phonetics,” *Lang. and Ling. Compass* 2, 671–711Hollien, H. Considerations of guidelines for earwitness lineups. *Forensic Linguistics*, 1986 ,3, 14-23;
- [4] E.San Segundo, P.Foulks, J.P French, P. Harrison, &V. Hughes, “Voice quality analysis in forensic comparison:developing the vocal profile analysis scheme” in IAFPA 2016- Annual Conference of the International Association of Forensic Phonetic and Acoustics, July 24-27, York, UK .2016
- [5] Hirano,M.(1981).“GRBAS” scale for evaluating the hoarse voice & frequency range of phonation. In M. Hirano (Ed.),*Clinical examination of voice* (Vol.5,pp.83 – 84, 88 – 89). New York, NY: Springer-Verlag/ Wien
- [6] Hollien H, Didla G., Harnsberger J.D., Hollien K.A. (2016). The case for aural perceptual speaker identification. *Forensic Science International*, **269**, pp. 5-20. <http://dx.doi.org/10.1016/j.forsciint.2016.08.007>
- [7] Laver, John. *The phonetic description of voice quality*. Cambridge Studies in Linguistics London, v. **31**, p. 1-186, 1980.
- [8] Wertheimer, M. *Untersuchungen zur Lehre von der Gestalt*. *Psychologische Forschung*, 1923, **4**, 301–350.
- [9] Wertheimer, M. (1938). *Laws of organization in perceptual forms*. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71-88). London, England: Kegan Paul, Trench, Trubner & Company.
- [10] Forstl, H, “ The dilemma of localizing language: John Abercombie ‘s unexploited evidence”, *Brain and language* , **40**:145-150, 1991
- [11] Broca, P. Sur la faculté du langage articule. *Bulletin of Social Anthropology*. **6**: 493-94.1865
- [12] Bever, T.G. Cerebral asymetries in humans are due to the differentiation of two incompatible processes: Holistic and analytic. *Annals of the New York Academy of Science*, **263**, 251-262, 1975. DOI: 10.1111/j.1749-6632.1975.tb41589. x
- [13] Van Lancker, D. (1991). Personal relevance and the human right hemisphere.*Brain & Cognition*,**17**, 64–92.[CrossRefGoogle Scholar](#)
- [14] Van Lancker, D. R. & Canter, G. J. *Impairment of voice and face recognition in patients with hemispheric damage*. *Brain Cogn.* **1**, 185–195, 1982.
- [15] Van Lancker, D., Kreiman, J., &Emmorey, K. Familiar voice recognition: Patterns and parameters: Part I. Recognition of backward voices.*Journal of Phonetics*,**13**, 19–38, 1985.[Google Scholar](#)
- [16] Van Lancker, P., Kreiman, J., Wickens, T. Familiar voice recognition: Patterns and parameters. Part II. Recognition of ratealtered voices.*Journal of Phonetics*,**13**, 39–52, 1985.[Google Scholar](#)
- [17] Van Lancker, D., Kreiman, J. Voice discrimination and recognition are separate abilities. *Neuropsychologia*, **25**, 829–854, 1987. [CrossRefPubMedGoogle Scholar](#)
- [18] Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, **24**(2), 195-209, 1988.
- [19] Formisano E, De Martino F, Bonte M, Goebel R “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science* **322**: 970–973, 2008.
- [20] Ladefoged, P. & Ladefoged, J (1980). The ability of listeners to identify voices. *UCLA Working Papers in Phonetics*, **49**; 43-51
- [21] Kisilevsky BS, Hains SM, Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang Z 2003 *Effects of experience on fetal voice recognition*. *Psychol Sci* **14**: 220–224.
- [22] Papcun, G., Kreiman, J. & Davis, A. *Long-term memory for unfamiliar voices*. *J. Acoust. Soc. Am.* **85**, 913–925 (1989).



- [23] Hollien, H. Considerations of guidelines for earwitness lineups. *Forensic Linguistics*, **3**, 14-23, 1986;
- [24] Kreiman, J., and Van Lancker Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception* (John Wiley and Sons, Chichester).
- [25] Gold, Erica; French, Peter. International practices in forensic speaker comparison. *International Journal of Speech, Language & the Law*, v. **18**, n. 2, 2011.
- [26] Lathi, Bhagwandas Pannalal. *Linear Systems and Signals*, Second Edition. Oxford University Press, 2005
- [27] French, P. and Stevens, L. (2013). Forensic speech science. In M. Jones and R. Knight, Eds., *The Bloomsbury Companion to Phonetics*. London: Bloomsbury
- [28] FOULKES P. and French P. (2012). Forensic phonetic speaker comparison. In Solan, I. and Tiersma, P. (eds.) *Oxford Handbook of Language and Law*
- [29] Abercrombie, D. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press, 1967
- [30] TITZE, I.R and Martin, D.W. Principles of voice production *Acoustical Society of America Journal*, **104**, 3 (1998), 1148.
- [31] Lass, R. (1984). *Phonology: an introduction to basic concepts*. Cambridge: Cambridge University Press.
- [32] Nolan, F. Forensic speaker identification and the phonetic description of voice quality. In W.J. Hardcastle & J. Mackenzie Beck (Eds.). *A figure of speech. A festschrift for John Laver* (pp.385-411). Mahwah. New Jersey: Lawrence Erlbaum Associates
- [33] Passetti, R. R. O efeito do telefone celular no sinal da fala : uma análise fonético - acústica com implicações para a verificação de locutor em português brasileiro. (Mestrado em Linguística)
- [34] Rose, P.J (2003). Effect of telephone transmission . In: Selby H. & Freckelton I. (eds.). *Expert Evidence*. Sydney: Thomson Lawbook Company, 2003.
- [35] Knösche, T. R., Lattner, S., Maess, B., Schauer, M., & Friederici, A. D.. Early parallel processing of auditory word and voice information. *NeuroImage*, **17**, 493-1503, 2002
- [36] Hammarberg B, Fritzel B, Gauffin J, Sundberg J, Wedin L Perceptual and acoustics correlates of abnormal voice qualities. *Acta Otolaryngol*
- [37] Hart DA, Hans S, Vassière J, Brassnu DA. Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia. *Eur Arch Otorinolaringol*, 260(4), 175-82, 2003
- [38] Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research*, 1114,161–172
- [39] Andics A, McQueen JM, Petersson KM, Gál V, Rudas G, Vidnyánszky Z. Neural mechanisms for voice recognition. *NeuroImage*. **52**, 1528–1540, 2010. <http://dx.doi.org/10.1016/j.neuroimage.2010.05.048>.
- [40] Kreiman J, Gerratt BR, Ito M. When and why listeners disagree in voice quality assessment tasks. *Journal of the Acoustic Society of America*. **122**, 2354–2364, 2007. <http://dx.doi.org/10.1121/1.2770547>. [PubMed]
- [41] Latinus M, McAleer P, Bestelmeyer PEG, Belin P. Norm-based coding of voice identity in human auditory cortex. *Current Biology*. **23**, 1075–1080, 2013. <http://dx.doi.org/10.1016/j.cub.2013.04.055>. [PMC free article] [PubMed]
- [42] Lavner Y, Rosenhouse J, Gath I. The prototype model in speaker identification by human listeners. *International Journal of Speech Technology*. **4**, 63–74, 2001. <http://dx.doi.org/10.1023/A:1009656816383>.
- [43] Li, X., & Pastore, R. E. Perceptual constancy of a global spectral property: Spectral slope discrimination. *Journal of the Acoustical Society of America*, **98**, 1956–1968, 1985.
- [44] Melara RD, Marks LE. Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception and Psychophysics*. **48**, 169–178, 1990. <http://dx.doi.org/10.3758/BF03207084>. [PubMed]
- [45] Bregman, A.S., 1990. *Auditory Scene Analysis*. MIT Press, Cambridge, MA

- [46] Caesar, Caesar's Commentaries on the Gallic War 155 (51 B.C.E) . Frederick Holland Dewey ed. Translation Publishing Co. 1918
- [47] William J. The Principles of Psychology. H. Holt and company; 1890.
- [48] Charlton, D; Fraser-Mackenzie, PAF; Dror, IE; Emotional Experiences and Motivating Factors Associated with Fingerprint Analysis. *JOURNAL OF FORENSIC SCIENCES* , **55** (2) pp. 385-393, 2010. [10.1111/j.1556-4029.2009.01295.x](https://doi.org/10.1111/j.1556-4029.2009.01295.x).
- [49] Whitman, G., & Koppl, R. Rational Bias in forensic science. *Law, probability and risk*, **9**(1), 2010
- [50] S. M. Kassin, I. E. Dror, J. Kukucka, "The forensic confirmation bias: Problems perspectives and proposed solutions", *J. Appl. Res. Memory Cognit.*, vol. **2**, no. 1, pp. 42-52, Mar. 2013.
- [51] Laubstein, A.S. Problems of voice line-ups. *Forensic Linguistics*, **4**, 262-279, 1997.
- [52] Fraser, H. (2003). Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases. *International Journal of Speech, Language and Law- Forensic Linguistics*
- [53] Fraser, H., Stevenson, B., & Marks, T. (2011). Interpretation of a crisis call: persistence of a primed perception of a disputed utterance.
- [54] Fraser, H. , & Australia , F.P (2014). Transcription of indistinct forensic recordings: Problems and solutions from the perspective of phonetic science . *Language and Law/ Linguagem e Direito*. Disponível em <http://ler.letras.up.pt/uploads/ficheiros/13353.pdf>.
- [55] Morrison, G. S., Lindh, J. & Curran, J. M. Likelihood ratio calculation for a disputed - utterance analysis with limited available data. *Speech Communication* , **58**, 81- 90, 2014.
- [56] Delgado Romero, Carlos. *La identificación de locutores en el ámbito forense*. Universidad Complutense de Madrid, Servicio de Publicaciones, 2004.
- [57] Cambier-Langeveld T. Current methods in forensic speaker identification: results of a collaborative exercise. *Int J Speech Lang Law* **14**(2), 223–243, 2007.
- [58] Rhodes, R (2014). Cognitive bias in forensic speech science. *Proceedings of IAFPA 2014*, Zürich, Switzerland. [Pholab.uzh.ch](http://Pholab.uzh.ch)
- [59] Findley, K. A, & Scott, M. S (2006). Multiple Dimensions of Tunnel Visions in Criminal Cases. *The Wis. L. Rev*, 291
- [60] Warren, R. (1999). *Auditory Perception: A New Analysis and Synthesis* (Cambridge University Press, Cambridge).
- [61] Ladefoged P. 'Expectation affects identification by listening' *Language and Speech*. **21**(4), 373–4, 1978.
- [62] Cambier-Langeveld, M; Rossum, M.V; Vermeulen, J *Whose Voice is that?* In: Caspers, Johaneke, Yiya, Willemijn Heeren, Jos Pacilly, Niels O. Schiller and Ellen van Zanten(eds), *Above and Beyond the Segments: Experimental Linguistics and phonetics*, 2014.xii, 363 pp. (pp.14-27)
- [63] M Sundqvist, T Leinonen, J Lindh, J Åkesson *Blind Test Procedure to Avoid Bias in Perceptual Analysis For Forensic Speaker Comparison Casework in IAFPA 2017- Annual Conference of the International Association of Forensic Phonetic and Acoustics*, July 9-12, Spli, Croatia .2017
- [64] Lindh, J., & Morrison, G. S *Humans Versus Machine: Forensic Voice Comparison on a Small Database of Swedish Voice Recordings* . ICPHS XVII, 2011, Hong Kong: 17-21.
- [65] Kunzel, H.J (1987) *Sprechererkennung: Grundzuge forensicher Sprachverarbeitung*. Heidelberg: Kriminalistik Verlag.
- [66] Kunzel, H.J (1995) 'Field procedures in forensic speaker recognition' . in Lewis (ed.) (1995): 68-84.
- [67] French, J.P (1994) 'An overview of forensic phonetics with particular reference to speaker identification'. *FL* ½ 169-84