

## UMA BREVE INTRODUÇÃO AOS ASPECTOS TEÓRICO-METODOLÓGICOS RELACIONADOS À COMPARAÇÃO FORENSE DE LOCUTOR

*Julio Cesar Cavalcanti<sup>1</sup>*

### 1. INTRODUÇÃO

Neste capítulo são abordados aspectos gerais referentes à tarefa de comparação forense de locutor, a partir de uma perspectiva fonético-acústica. Preceitos teórico-metodológicos relevantes para a prática e pesquisa na área são discutidos, de modo a fornecer ao leitor um breve panorama acerca dos diferentes fatores, internos e externos aos locutores, que merecem atenção quando à caracterização da identidade fonético-acústica de um falante se põe em foco.

Diferentes variáveis, de diferentes ordens, a exemplo da influência do canal de transmissão, das degradações de áudio, do estilo de fala e do próprio conceito de espontaneidade na fala são comentadas e critérios teóricos para a seleção de parâmetros fonético-acústicos são apresentados a partir da literatura especializada.

### 2. A FONÉTICA FORENSE

É possível pensar a fonética forense enquanto uma subárea da fonética aplicada inserida na intersecção entre as ciências da fala e as ciências de natureza forense, o que atribui a esse campo uma característica ou um traço híbrido.

---

<sup>1</sup> Universidade Estadual de Campinas (UNICAMP).

Em se tratando de definições, algumas tentativas de descrever o que é esse campo de estudo foram feitas por alguns autores, dentre eles, Jessen (2008), o qual descreve a fonética forense como sendo a aplicação do conhecimento, teorias e métodos da fonética geral para tarefas práticas que surgem em um contexto específico de investigação, bem como no desenvolvimento de saberes e métodos próprios do campo de investigação fonético-forense. Em termos mais abrangentes, a fonética forense poderia ser também descrita como a aplicação do conhecimento fonético geral a problemas de natureza legal (SAN SEGUNDO, 2014).

Para além da “diferenciação” ou “caracterização” forense de um locutor, a fonética forense inclui tarefas como a descrição do perfil de falantes, a partir da análise de características da voz e da fala, adjacente à análise de indexadores sociolinguísticos, a exemplo do sexo, idade, região de origem, nível social, interpretados a partir do comportamento linguístico dos locutores. Inclui também a identificação do conteúdo linguístico, processo denominado de “decodificação” do sinal de fala – a fim de determinar o que foi dito quando as gravações são de má qualidade ou apresentam algum tipo de degradação, e a autenticação de gravações, de modo a determinar se um registro foi ou não intencionalmente adulterado (ROSE, 2002; HOLLIEN, 2012). Outras tarefas podem compreender o tratamento do sinal de fala, como na aplicação de filtros de ruído, de modo a tornar a informação acústica inteligível e avaliável do ponto de vista fonético, tarefa que requer um conhecimento bastante aprofundado por parte do perito, de modo a evitar que informações linguísticas relevantes sejam eliminadas junto com o conteúdo indesejável (HOLLIEN, 2012).

Conforme menciona Rose (2002), dentre as atividades desempenhadas por um foneticista forense, a tarefa mais comum é aquela que envolve a comparação de uma ou mais amostras de fala desconhecida, denominada de amostra de “fala questionada”, com uma ou mais amostras de origem conhecida, a amostra de “fala padrão”.

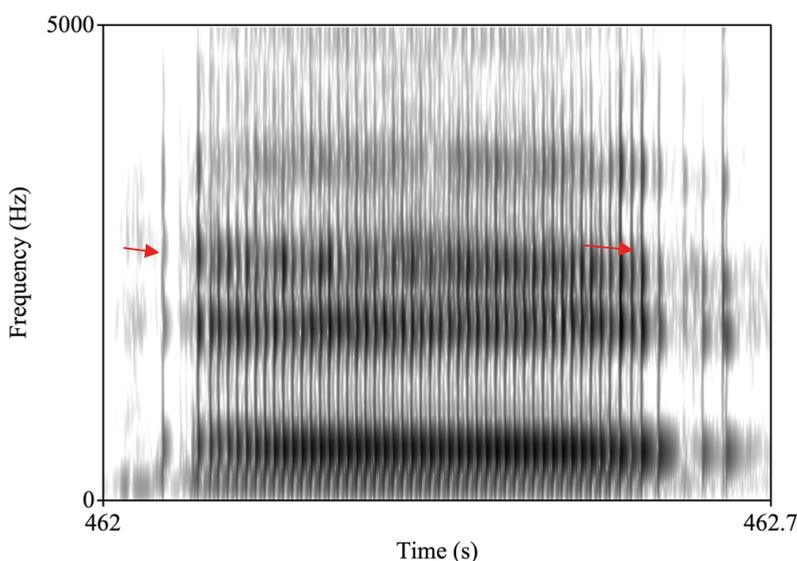
De acordo com Morrison (2009:229), o principal objetivo da comparação entre essas amostras é responder à importante questão: “quão semelhantes são as propriedades observadas nas amostras de fala padrão e questionada, sob a hipótese de que a amostra de fala questionada tenha a mesma origem que a amostra de fala padrão, em contraposição à hipótese de que estas tenham uma origem diferente?”.

A fim de responder a essa pergunta os foneticistas forenses lançam mão de análises baseadas em características distintivas ou idiossincráticas de um locutor, como a análise da qualidade vocal, padrões articulatórios na produção de vogais,

ditongos e consoantes, padrões temporais, a exemplo da duração dos segmentos, taxas de fala, entre outros parâmetros potencialmente discriminatórios.

Conforme menciona Rose (2002, p. 5), as análises comparativas são sempre realizadas em relação às unidades linguísticas, como as citadas anteriormente. Pode-se observar numa análise, por exemplo, que duas vogais idênticas realizadas em um mesmo contexto são produzidas de forma distinta por dois locutores, ou que as consoantes /k/ e /g/ são idiossincraticamente produzidas por um desses indivíduos, em decorrência de algum processo fonológico. Neste caso, o perito ou pesquisador deve ter um conhecimento abrangente sobre o sistema fonológico e inventário fonético da língua em questão, sobre a dinâmica de produção dos gestos articulatórios e as possíveis formas de se descrever as diferentes realizações. Entender a estrutura da língua possibilita a verificação de quando as realizações são ou não previsíveis do ponto de vista linguístico ou de quando essas estão relacionadas ao comportamento particular de um locutor, sendo, portanto, variáveis. Exemplo de um fenômeno prosódico idiossincrático é apresentado na Figura 1.

Figura 1 – Efeitos no espectrograma do evento de laringalização no início e final de uma pausa preenchida (indicado pelas setas vermelhas) com a vogal /ε/ no Português Brasileiro, por parte de um falante do sexo masculino durante a fala espontânea. O fenômeno é considerado como de caráter idiossincrático no PB, não apresentando *status* contrastivo no sistema fonológico da língua, e portanto, com frequência de ocorrência passível de variação entre os locutores. Ver Cavalcanti; Lucente; Barbosa (2018).



Em se tratando da produção de segmentos na fala, um dos fenômenos de grande relevância para a fonética forense é o fenômeno da “coarticulação”. Conforme comenta Hollien (1990:68-69), nenhum gesto articulatório é independente daqueles que o precedem ou sucedem, ao contrário disso, há uma sobreposição entre movimentos adjacentes em qualquer sequência de sons falados. De acordo com o autor, esse fenômeno é resultado de pelo menos dois de muitos fatores relacionados à produção oral humana: a taxa com que falamos e a interação entre as estruturas móveis (músculos) no processo de produção. Essas dimensões têm relação direta com aquilo que denominamos de plasticidade vocal, ou em termos mais amplos, plasticidade fonoarticulatória.

A plasticidade vocal implica a suposição de que um traço orgânico pode ser impresso de forma variável no sinal acústico, no que diz respeito à produção interna de um mesmo indivíduo e entre diferentes locutores, dado que estes normalmente diferem entre si no que diz respeito a características fonoarticulatórias, como tamanho e massa das pregas vocais, dimensões da laringe, língua, lábios etc., o que pode implicar modificações substanciais nas medidas acústicas. Existem, contudo, limites fisiológicos para os graus de variação, assim como tendências e padrões sistematicamente adotados pelos sujeitos (NOLAN, 1983). A análise de tais padrões ou daquilo que seria um comportamento particular de um locutor é um dos principais focos da fonética forense.

Rose (2012) chama a atenção para o fato de que os diferentes tipos de informações transmitidas por meio da voz/fala humana, a exemplo de demarcadores como sexo, idade, nível socioeconômico, escolaridade, não são sinalizadas em canais separados, mas são transmitidas juntamente com o conteúdo linguístico. O conhecimento de como esse processo se dá é fundamentalmente necessário para entender e interpretar a variação na fala.

### **3. QUESTÕES TERMINOLÓGICAS**

A fim de fornecer aos leitores um panorama geral sobre a prática de comparação forense de locutor, alguns comentários acerca da terminologia empregada na área se fazem necessários. Uma distinção clássica que merece atenção diz respeito à diferença entre as nomenclaturas “identificação” e “verificação” de locutor.

Conforme menciona Hollien (2012), a identificação de locutor é um processo em que são feitas tentativas de se identificar um indivíduo por meio da fala e da voz quando a identidade dessa pessoa não é conhecida, dentro de um

conjunto de vozes/falas numa determinada população. Essa tarefa, embora de grande importância, é bastante complexa, uma vez que todos os tipos de distorções e interferências possíveis em uma gravação podem estar presentes.

A verificação de locutor, por sua vez, se dá quando um locutor conhecido e cooperativo deseja ser reconhecido pela sua fala. Assim, amostras da fala/voz desse locutor são coletadas a fim de se construir um banco de referência. A identidade do indivíduo é então verificada quando uma nova amostra de fala é comparada aos dados obtidos previamente e a decisão é tomada: mesmo ou diferentes locutores? Enquanto as técnicas de verificação de falante são bastante úteis e exploradas pela indústria de tecnologia (como em sistemas de reconhecimento de fala, por exemplo), elas são utilizadas apenas ocasionalmente no contexto forense (HOLLIEN, 2012).

Há algum tempo, o emprego do termo “identificação de locutor” tem sido foco de discussões no campo da fonética forense. Atualmente, o termo vem sendo amplamente substituído pela nomenclatura “comparação” forense de locutor, para se referir à análise da autoria de registros de voz e de fala. Tal discussão tem por base razões propriamente teórico-terminológicas.

Para Morrison (2009), numa abordagem probabilística (*likelihood-ratio framework*<sup>2</sup>), o cientista forense não realiza a “identificação” de um locutor, porque esse termo implica determinar uma probabilidade posterior. Segundo o autor, um termo neutro e mais apropriado seria “comparação”. Nesse sentido, a comparação forense entre as amostras padrão e questionada não é uma tarefa que se resume a uma tomada de decisão binária (mesmo locutor vs. locutores diferentes), mas sim uma tarefa que busca determinar a força da evidência em relação a hipóteses de que as gravações têm ou não a mesma origem. Nesse caso, o que o cientista/perito forense tem como resposta é um valor de probabilidade e não a “verdade” expressa em termos de uma resposta categórica.

## 4. FATORES E VARIÁVEIS

A prática de comparação forense de locutor pode apresentar características bastante distintas quando realizada em condições controladas e em contextos sem controle. De acordo com Rose (2002), a falta de condições ideais no mundo

---

<sup>2</sup> *Likelihood-ratio framework* ou a abordagem baseada na razão de verossimilhança é um modelo conceitual que pode ser aplicado a suposições subjetivas baseadas na experiência, com vista à probabilidade de se ter observado uma determinada evidência dadas as hipóteses concorrentes (MORRISON, 2009).

real dificulta as análises forenses em decorrência de dois importantes fatores: a falta de controle sobre a variação e a redução nas dimensões de comparação entre amostras.

Conforme menciona Rose (2002), como geralmente não temos controle sobre as situações em que os dados forenses são coletados, esse se torna um grande desafio na avaliação das diferenças entre as amostras. Na grande maioria das vezes, o perito é levado a confrontar uma gravação de referência coletada num contexto quase ou completamente distinto do contexto de aquisição da amostra padrão, em decorrência da presença e/ou da ausência de variáveis típicas de cada condição.

Mais adiante comentaremos sobre alguns dos principais fatores e variáveis relacionadas à redução da dimensionalidade comparativa entre amostras de fala num contexto de comparação forense de locutor.

#### **4.1 Transmissão e qualidade da gravação**

De acordo com Rose (2002), uma das condições em que a dimensionalidade do contexto real na comparação de falantes é substancialmente reduzida diz respeito ao registro de fala telefônico. Isso se deve comumente a que aspectos da fala que poderiam ser úteis à comparação, características com considerável potencial distintivo, simplesmente não estão disponíveis porque foram filtradas, ou mesmo em decorrência de fatores práticos, como o ruído de fundo, má transmissão, falas sobrepostas, todos aspectos relacionados à degradação do sinal de áudio.

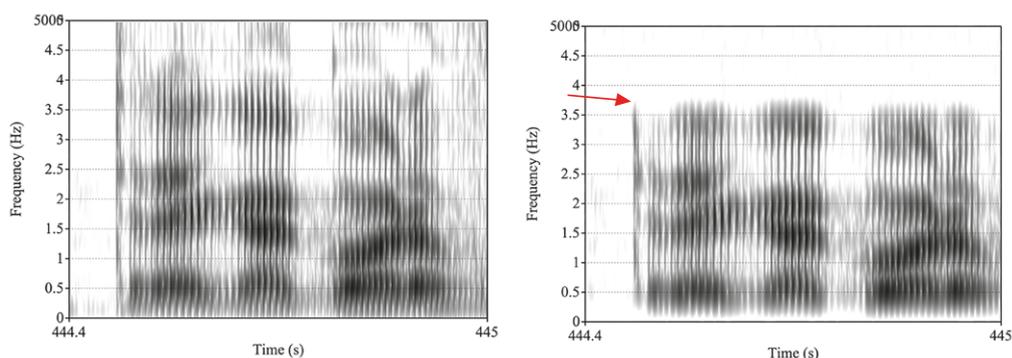
Conforme comenta Hollien (2012), as degradações podem ser resultantes da redução da largura de banda de frequência, ruído aditivo, redução do nível de energia acústica, distorção espectral ou harmônica, transmissão inadequada, transdutores de captação inadequados (telefones, gravadores), fatores ambientais, tais como o vento, atrito da roupa, interferências de outros falantes, música, entre outros.

Como mencionado anteriormente, esses problemas ocorrem basicamente porque, em geral, não podemos ter controle das situações nas quais as amostras de fala são registradas, o que se torna inevitavelmente um grande problema para os foneticistas forenses: tentar comparar amostras coletadas em condições distintas e que de algum modo incorporam marcas e efeitos diferentes no sinal acústico (ROSE, 2002).

De acordo com Künzel (2001), embora alguns dos distúrbios típicos, como altos níveis de ruído de fundo ou distorções em decorrência do uso de microfones

de carbono tenham se tornado um menor problema com o advento da telefonia digital moderna, outros fatores persistiram, como por exemplo os efeitos de passa-banda do canal de transmissão telefônico (350-3.400 Hz). A Figura 2 ilustra o efeito isolado do filtro de telefone sobre o espectrograma da fala. Na imagem é possível constatar a perda da informação acústica em torno da frequência de 3.500 Hz, como indicado pela seta.

Figura 2 – Efeito artificial do filtro de telefone na fala encadeada durante a frase: “(en)trega agora”, por um falante do sexo masculino. Espectrograma manipulado no *software* Praat (BOERSMA; WEENINK, 2021).



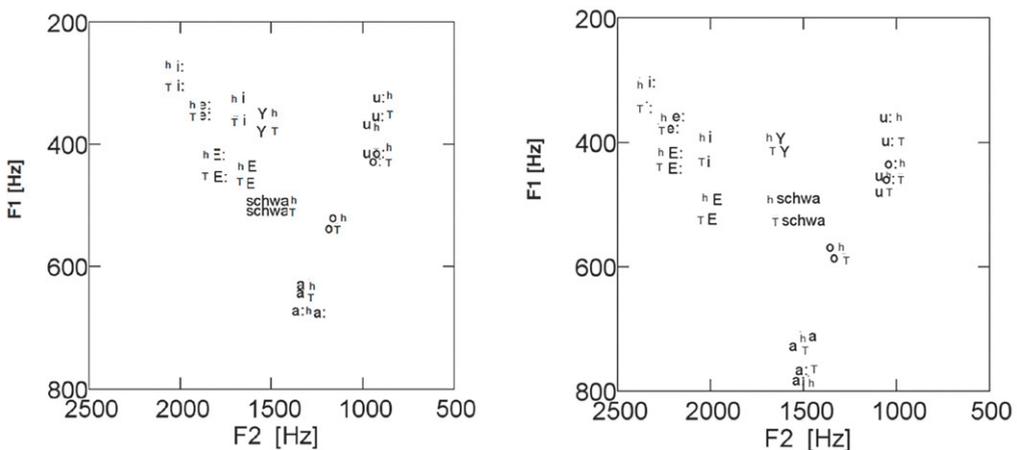
Um exemplo prático dos efeitos de telefones móveis no sinal de fala pode ser constatado no trabalho de Passetti (2015). Em um estudo com vogais orais do Português Brasileiro, a autora buscou determinar o grau de modificação fonético-acústica causado pelo filtro de banda do canal telefônico. Nesse estudo foram analisados dez locutores do sexo masculino, em gravações simultâneas nas condições de fala via celular e via direta. Os resultados desse trabalho revelaram que houve alterações nas frequências do primeiro (F1) e do terceiro formante (F3), em torno de 14% entre as duas condições testadas. No que diz respeito às frequências do segundo formante (F2), a transmissão telefônica influenciou de forma a aumentar as frequências de vogais com baixos valores de F2 e a diminuir as frequências de vogais com altos valores de F2, desencadeando o abaixamento global do espaço vocálico na gravação telefônica, pelo aumento nas frequências de F1, e o estreitamento do mesmo pela diminuição dos valores de F2. Apenas a *baseline*<sup>3</sup> e a duração interpicos de F0 não apresentaram diferenças significativas

<sup>3</sup> Caracterizado pelo valor de F0 mais típico de um locutor, correspondente aproximadamente ao sétimo quantil dos valores F0 da amostra (ARANTES; ERIKSSON, 2014).

no estudo. De acordo com a autora, as modificações nas disposições das vogais têm implicações perceptuais, uma vez que o abaixamento e redução do espaço vocálico fizeram com que as vogais se situassem proximamente a regiões centrais, podendo soar como mais abertas no telefone celular.

Numa investigação com vogais do alemão, a partir de gravações diretas e via telefone com dez indivíduos do sexo masculino e dez do sexo feminino, Künzel (2001), por sua vez, verificou que a medida de F1 foi a mais afetada dentre os parâmetros acústicos testados. As medidas centrais de F1 de cada vogal apresentaram valores mais altos nos dados transmitidos por telefone em comparação com os dados registrados de forma direta, corroborando a hipótese inicial do autor de que a atenuação dos componentes inferiores de um formante aumentaria o peso relativo dos componentes superiores, provocando um deslocamento artificial ascendente das frequências centrais. Nesse estudo, o efeito previsto em F1 ocorreu para todas as vogais testadas, exceto para a vogal /a/, para a qual, conforme menciona o autor, a medida apresenta um valor muito alto para ser afetado pelo efeito de passa-banda. Ver Figura 3.

Figura 3 – Espaço vocálico para falantes do sexo masculino (esquerda) e feminino (direita) para as condições de gravação direta (h) e telefônica (T).



Fonte: Künzel (2001:91)

Outra variável de considerável relevância e que geralmente acompanha a degradação causada pelo filtro telefônico é a presença de ruído. O ruído pode ser de diferentes naturezas, a exemplo do ruído ambiental, nos quais se enquadram os ruídos de vento, automóvel, trânsito, motores, equipamentos elétricos.

Eles podem ainda ser de natureza “linguageira”, quando há a sobreposição de falas, como uma conversa no fundo de uma gravação ou ligação telefônica. Nesse último caso, o impacto acústico é substancialmente elevado, uma vez que o ruído de fala pode representar modificações significativas nos parâmetros fonético-acústicos do sujeito foco da análise. O ruído pode ainda ser proveniente de mecanismos internos do sistema de captação de áudio, como o ruído interno de microfones, computadores, gravadores e celulares.

O trabalho de Constantini e Barbosa (2013) a partir da análise de diferentes parâmetros prosódicos em diferentes relação sinal-ruído, evidenciou que alguns parâmetros acústicos apresentaram mudanças em suas medidas após a adição de ruído gaussiano, dos quais: a ênfase espectral, mediana de frequência fundamental, taxa de saliência duracional, média, desvio-padrão e assimetria de z-score de duração de unidade VV. Destes, a ênfase espectral e a mediana da frequência fundamental foram os parâmetros mais afetados, tendo sido observado uma diferença mínima para os demais parâmetros. De acordo com os autores, a presença do ruído aditivo torna a extração dos parâmetros mais difícil por parte do software de análise (*praat*), repercutindo em erro nas medidas. Ver Tabela 1.

Tabela 1 – Médias de ênfase espectral e de mediana de frequência fundamental e magnitudes de mudança após adição de ruído gaussiano. Dados do trabalho de Constantini e Barbosa (2013:3)

	Gravação original	Ruído 0,01	Magnitude de mudança	Ruído 0,02	Magnitude de mudança
<b>M<sub>mediana f0</sub></b>	125 Hz	126 Hz	0,8%	128 Hz	2,4%
<b>M<sub>ênfase espectral</sub></b>	1.18 dB	1.8 dB	55%	3 dB	154%

De acordo com os autores, a análise da medida de ênfase espectral em gravações ruidosas é desencorajada, uma vez que a medida é considerada bastante sensível ao componente de ruído. A análise da estrutura rítmica do enunciado, no entanto, pode ser considerada mais consistente em situações em que se é preciso comparar amostras de fala em diferentes relações sinal-ruído. Isso se deve ao fato de que, por apresentarem maior energia acústica em relação aos demais segmentos da fala, os segmentos vocálicos tendem a ser preservados no espectrograma.

## 4.2 Estilo de fala

É um fato inegável o de que os falantes tendem a modificar os seus comportamentos linguísticos de acordo com os diferentes contextos comunicativos ou em função do interlocutor, o que reforça a natureza dinâmica da comunicação humana. Esses ajustes podem ocorrer tanto no nível lexical, a partir da seleção cuidadosa de palavras ou de um vocabulário específico, no nível sintático, refletindo uma maior ou menor produção de concordâncias, e até mesmo no plano fonético, a partir da modificação de parâmetros fonético-acústicos, a depender do estilo de fala adotado.

Um dos contextos que podem implicar modificações no comportamento fonético-acústico de um locutor diz respeito à condição de fala por telefone. O trabalho de Passetti (2018) sobre o estilo de fala telefônico com falantes do PB, verificou que, entre locutores do sexo masculino, o estilo de fala telefônico apresentou-se caracterizado pela modificação de descritores da frequência fundamental e de parâmetros temporais, a exemplo da distribuição dos picos de duração normalizada de sílabas fonéticas<sup>4</sup>, duração dos intervalos entre duas pausas silenciosas consecutivas e a duração de grupos acentuais, sendo essas modificações potencializadas em decorrência da presença de ruídos ambientais, com exceção da duração dos intervalos entre pausas silenciosas.

De acordo com a autora, em geral, o estilo de fala telefônico está associado a valores globais da frequência fundamental mais elevados, a uma maior variação entoacional, um maior grau de saliência duracional e uma maior quantidade de pausas silenciosas. Os intervalos entre pausas silenciosas tendem, por sua vez, a serem menores, sinalizando trechos de fala mais curtos na interação por telefone.

Outros fatores importantes dizem respeito ao grau de monitoramento dos falantes e aos efeitos das emoções na fala. Ao falar em público (exemplo de um registro de fala mais monitorado), articulações mais precisas tendem a ser produzidas em contraste com a fala espontânea. É comum que quando o falante se sente mais à vontade e confortável em uma conversa, o estilo de fala se torne mais casual, com palavras e frases simplificadas, ou mesmo utilizando um outro vocabulário. É nesse tipo de contexto de fala que reduções, assimilações e coarticulações são mais frequentemente e claramente produzidas. A taxa de fala tende a ser aumentada em comparação com a fala clara e, portanto, a quantidade de reduções e apagamentos também tende a aumentar (BJURSÄTER, 2004).

---

<sup>4</sup> Unidade constituída por uma vogal e todos os segmentos consonantais que a seguem, independentemente da fronteira silábica, até o *onset* da vogal seguinte, o qual determina o início da próxima unidade VV (BARBOSA, 2006).

Emoções podem resultar no aumento ou diminuição dos valores dos parâmetros acústicos, como na expressão de raiva, por exemplo, frequentemente relacionada a maiores taxas de fala, maiores valores de F0 e de intensidade, uma maior gama de variação de F0 e uma articulação mais tensa (MURRAY & ARNOTT, 1993). É importante estar atento, portanto, a essas possíveis modificações na fala relacionadas ao estado emocional dos sujeitos, especialmente em situações com altos níveis de estresse.

### 4.3 Sobre a espontaneidade na fala

Uma questão que interessa a linguistas, foneticistas forenses e foneticistas de um modo geral diz respeito ao impacto do grau de controle do experimentador no comportamento de fala dos sujeitos. Esse tópico também é tratado com bastante atenção por pesquisas em sociolinguística, as quais buscam elicitar o comportamento linguístico do sujeito o mais próximo daquilo que configura o uso real, de modo a reduzir ao máximo a intervenção do pesquisador sobre os dados obtidos – efeito classicamente referido no campo da sociolinguística como o “paradoxo do observador” (ver LABOV, 1972).

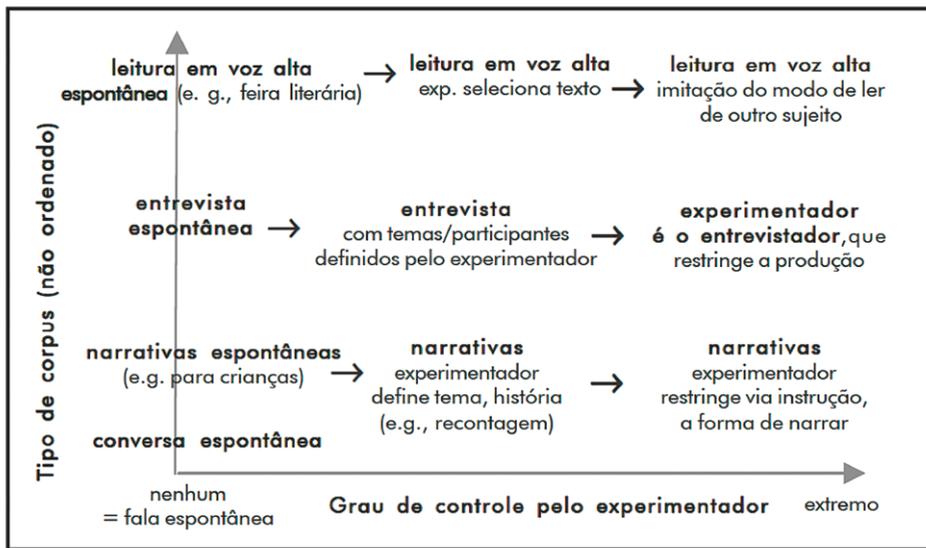
A ideia bastante recorrente de que a amostra de fala padrão, aquela coletada em um contexto de gravação controlado, apresenta uma redução significativa na dimensão de espontaneidade em relação à amostra de fala questionada, pode e deve ser discutida. Para tanto, alguns comentários acerca daquilo que se entende por fala espontânea, em detrimento de uma instância de fala dita não espontânea, devem ser tecidos.

De acordo com Barbosa e Madureira (2015), se considerarmos a questão da espontaneidade da fala, podemos pensar em um contínuo que vai da fala espontânea, sem intervenção ou controle por parte do pesquisador, à fala de laboratório, contexto no qual o pesquisador exerce algum tipo de controle sobre aquilo que deve ser produzido pelos sujeitos, como a seleção do gênero e do conteúdo a ser falado.

A Figura 4 representa uma proposta de classificação dos *corpora* de fala desenvolvida por Barbosa (2012), a partir de duas dimensões: o grau de controle do experimentador e o gênero (não ordenado). Neste modelo, é possível observar que conversas espontâneas, sem tópicos predefinidos, configuram o contexto com menor ou nenhum controle do experimentador correspondente ao eixo de espontaneidade, enquanto uma tarefa de imitação representa o oposto, maior nível de controle experimental por parte do pesquisador.

O gênero do tipo “entrevista”, com temas definidos, seria o gênero conversacional correspondente àquele de um interrogatório em um contexto forense, somado, é claro, a características particulares de uma situação investigativa, com uma possível acentuação do fator “estresse”.

Figura 4 – Proposta de classificação de corpora segundo eixo “grau de controle do experimentador” e “gênero ou tipo de corpus”.



Fonte: Barbosa (2012:16).

Como menciona Xu (2010), existe uma visão generalizada de que somente a partir da análise da fala espontânea é que podemos entender a natureza da fala cotidiana. Acompanhando essa visão está a ideia, cada vez mais difundida, de que a chamada fala de laboratório<sup>5</sup> é inadequada e não expressa ou esclarece a riqueza da fala espontânea, sendo, portanto, um registro pouco representativo do falante. De acordo com o autor, essa crença é baseada numa consideração limitada da literatura da área e uma apreciação inadequada de princípios científicos básicos. Em defesa da naturalidade da fala de laboratório o autor argumenta (XU, 2010:3-4):

This may be one of the most readily conceived characteristics of lab speech, because it

<sup>5</sup> Em uma definição ampla, a partir de Xu (2010), a fala de laboratório refere-se à fala gravada em laboratório, geralmente obedecendo um roteiro definido previamente. Contudo, o termo “fala de laboratório” é frequentemente usado para se referir a um discurso estereotipado

seems to contain an element of truth, i.e., scripted speech, by definition, is non-spontaneous. And non-spontaneous seems to be the opposite of natural. But it is important to first determine what is meant by “natural”. If naturalness is taken to mean reflecting human capabilities, all speech must be natural by definition. Indeed, even the most stereotyped lab speech is produced by real speakers. Regardless of what the speakers are asked to do in an experiment, their performance is based on their naturally acquired ability to speak, and is therefore a reflection of what they do everyday. It is not the case, for example, that they learn from the experimenter how to produce a vowel or a consonant or a tone, or they learn from the experimenter how to make an emphasis or ask questions. They already know how to do those things, and that’s why they are invited into the laboratory in the first place.

Conforme menciona o autor, se naturalidade significa refletir as capacidades humanas, todo registro de fala é por essência natural, uma vez que mesmo um registro de fala controlado é produzido por falantes reais e os seus desempenhos são resultado da capacidade de falar adquirida naturalmente. São por essas razões, inclusive, que os falantes são convidados para participar em pesquisas de laboratório.

De acordo com Barbosa (2012: 14), é preciso desvincular completamente os conceitos de fala espontânea e conversa livre, pois, segundo o autor, uma leitura, uma narrativa ou uma entrevista são de igual maneira espontâneas se ocorrerem em uma situação de comunicação dita “natural”, cabendo ao pesquisador apenas a seleção do material.

A redução nas dimensionalidades entre as amostras de fala padrão e questionada num contexto forense pode estar, muito mais provavelmente, relacionada ao quanto essas amostras são ou não representativas do sujeito em questão, ou seja, o quanto elas refletem do uso linguístico do falante, levando-se em conta fatores como os tamanhos e durações das amostras, os fatores de degradação mencionados anteriormente, a familiaridade entre os interlocutores, e os estilos de fala comparados, normalmente distintos no contexto de fala “padrão” e “questionada”.

Nesse sentido, a representatividade das amostras pode apresentar um impacto considerável na tarefa de comparação de locutor e na extração de medidas acústicas repercutindo na comparabilidade dos materiais coletados, como comentado a seguir.

---

como aquele em que o locutor apenas realiza a repetição de frases.

## 4.4 Sobre a representatividade dos dados

Por representatividade, pode-se entender o quão representativas são as observações reais do registro de voz e de fala em questão, levando-se em conta a sua origem. Nesse sentido, quanto mais representativos forem os dados, mais consistentes serão as estimativas, a favor ou contra a hipótese de uma origem comum (ROSE, 2012).

Num contexto forense real, por exemplo, é bastante comum que a amostra questionada apresente discrepâncias em termos de duração e extensão em comparação à amostra padrão, assim como diferenças no que diz respeito ao número de registros disponíveis, sendo essa última em geral mais extensa e numerosa, dada a possibilidade da realização de múltiplas coletas. Esse fator pode apresentar repercussões tanto na representatividade quanto na comparabilidade entre amostras ou mesmo limitar o potencial explicativo da análise. Nesse sentido, a depender do grau de assimetria entre as amostras, maiores ou menores são as chances de se realizar generalizações e de se chegar a conclusões gerais a respeito do perfil de um locutor.

Uma das principais questões relacionadas à tarefa de comparação de locutor diz respeito à quantidade de material de fala necessário e à duração mínima dos registros, de modo que esses incorporem dados suficientes de um falante e expressem uma gama de variabilidade significativa de parâmetros fonético-acústicos relacionados ao perfil locucional.

Algumas pesquisas têm se voltado para essa questão, como é o caso do trabalho de Arantes e Eriksson (2014). Os autores buscaram determinar, a partir de um método estatístico denominado “*change point analysis*”, o ponto de estabilização da variabilidade de medidas da frequência fundamental numa análise com diversas línguas. A razão para uma análise dessa natureza, com um considerável número de sistemas prosódicos (26 línguas, tonais e não tonais), tem como objetivo evitar que os dados obtidos estejam fortemente condicionados à influência de um conjunto de línguas em particular.

Os resultados desse experimento revelaram que a média e a mediana da frequência fundamental tendem a atingir a estabilidade com cerca de dez segundos de duração, enquanto o valor de base (*baseline*) leva em torno de cinco segundos, um intervalo de tempo ainda menor do que sugerido por trabalhos anteriores.

No que diz respeito a taxas de fala, a exemplo da taxa de elocução e taxa de articulação, os experimentos realizados por Arantes, Eriksson e Lima (2018) evidenciaram que, em geral para essas medidas, o tempo médio de estabilidade

é de 12,1 segundos, com a maior parte da distribuição entre 7,9 e 16,2 s. Os autores verificaram, no entanto, que houve um efeito significativo do tipo de taxa no tempo de estabilização, com a taxa de articulação apresentando um tempo de estabilização mais rápido. De acordo com Arantes, Eriksson e Lima (2018), é possível que os intervalos de estabilização ligeiramente mais longos na condição de taxa de elocução devam-se à inclusão das durações das pausas silenciosas, o que repercute no aumento da variabilidade de duração dos segmentos. Variáveis como o nível das taxas de fala (se rápido, normal ou lento) e o critério de ponto de estabilização (se levando em conta ou não o limiar perceptivo), também demonstraram influenciar significativamente o tempo de estabilização das medidas.

O tamanho e extensão das amostras de fala também podem apresentar impacto na capacidade de discriminação por parte dos ouvintes, como, por exemplo, em tarefas de reconhecimento de locutor, nas quais o ouvinte é solicitado a identificar o registro de um falante específico a partir de uma sequência de vozes (*vocal line-up*).

Os resultados obtidos por Decoster et al. (2001), por exemplo, revelam que há uma relação direta entre o tipo e duração do estímulo empregado e o desempenho dos ouvintes na tarefa de discriminação. Nesse caso, a performance dos sujeitos tende a ser tanto melhor quanto mais informações acústicas e prosódicas sobre o locutor estiverem contidas no estímulo, com um melhor desempenho para unidades linguísticas maiores, como sentenças, em detrimento de unidades linguísticas menores em extensão e duração, a exemplo de vogais. Tanto experimentos fonéticos gerais (KREIMAN; SIDTIS, 2011) quanto estudos que tratam de dados de fala/voz de locutores geneticamente relacionados (SAN SEGUNDO, 2014) confirmam a variação na performance dos ouvintes em decorrência desses fatores.

Ainda em se tratando da representatividade dos dados, é preciso levar em consideração a variável “contemporaneidade” das amostras, que diz respeito à janela de tempo que separa os registros das falas padrão e questionada. A não contemporaneidade entre amostras pode implicar diferenças nos parâmetros fonético-acústicos em função da variação intralocutor. Nesse sentido, é importante saber o que esperar em termos de modificação nos padrões de voz e de fala, especialmente se as amostras foram coletadas em janelas de tempo consideravelmente grandes.

Mudanças adotadas ou sofridas pelos locutores, como migrar para uma outra região (com diferente dialeto/língua) ou mesmo hábitos adquiridos ao longo da vida, como o tabagismo por exemplo – capaz de afetar clinicamente a qualidade

vocal, resultando numa diminuição da frequência fundamental e aumento dos valores da relação ruído-harmônico (NHR), *jitter* e *shimmer* (PINHO; CRESPO; MOURÃO, 2014) – são exemplos de fatores que devem ser considerados numa análise comparativa, além de mudanças na qualidade da voz em decorrência do envelhecimento (presbifonia).

## 5. CRITÉRIOS PARA A SELEÇÃO DE PARÂMETROS FONÉTICO-ACÚSTICOS

Tendo em vista todos os fatores comentados anteriormente e a suas implicações para a prática de comparação forense de locutor, faz-se necessário tecer alguns comentários sobre os critérios de seleção de parâmetros fonético-acústicos para a análise do perfil de falantes. Existem alguns requisitos teóricos básicos que precisam ser respeitados, de modo que a seleção de um determinado conjunto de medidas seja tanto coerente do ponto de vista da análise quanto sensível à variação entre sujeitos.

De acordo com Nolan (1983), um parâmetro candidato ao uso forense deve idealmente satisfazer seis premissas básicas, conforme apresentadas na Tabela 2. Os critérios propostos pelo autor representam uma tentativa de se estabelecer quais informações contidas no sinal acústico são altamente dependentes e representativas de um indivíduo, de modo a restringir aquelas de baixo poder discriminatório.

É preciso considerar no entanto que, na realidade prática a seleção de parâmetros é um processo altamente dependente das técnicas de análise disponíveis, o que não implica que critérios teóricos e práticos não possam ser propostos e discutidos. Além do fato de que é preciso ter em mente que os parâmetros elencados como discriminatórios para uma dupla de locutores podem não ser contrastivos para outra. Nesse sentido, a relevância de cada parâmetro precisa ser estudada caso a caso, conforme menciona San Segundo, (2014).

Tabela 2 – Critério para a seleção de parâmetros para a tarefa de comparação forense de locutores (CFL) de acordo com Nolan (1983:11)

<i>Alta variabilidade entre locutores</i>	O parâmetro deve apresentar um alto grau de variação entre locutores.
<i>Baixa variabilidade intralocutor</i>	Deve ser consistente em relação às produções de um mesmo indivíduo e preferencialmente resistente ao seu estado de saúde, condição emocional e ao contexto comunicativo.
<i>Resiste à tentativa de disfarce ou imitação</i>	Deve ser resistente às tentativas de disfarce ou imitação, seja por se tratar de uma característica fisiológica difícil de ser modificada ou por refletir um aspecto da fala que esteja abaixo do nível de percepção do falante.
<i>Resistente ao canal de transmissão</i>	A aplicabilidade de um parâmetro será limitada se a sua informação for passível de ser perdida ou reduzida em decorrência do efeito do canal de transmissão.
<i>Disponibilidade</i>	O parâmetro ou fenômeno deve apresentar uma frequência de ocorrência considerável, de forma a não requerer uma grande quantidade de material de fala, nem sempre disponível num contexto forense.
<i>Mensurabilidade</i>	A extração do parâmetro ou análise do fenômeno não devem ser demasiadamente complexas, de modo que inviabilize a mensuração.

Encontrar parâmetros que satisfaçam a esses critérios é um objetivo e um grande desafio para as pesquisas em fonética forense. Contudo, algumas estratégias podem ser adotadas no contexto experimental de modo a se avaliar o grau de consistência dos parâmetros fonético-acústicos para a tarefa de comparação de locutor e as suas resistências às variáveis presentes no contexto prático. Dentre algumas das estratégias mais comuns podemos citar:

- Gravações não contemporâneas: as gravações são realizadas respeitando-se um determinado intervalo de tempo entre si e posteriormente submetidas à comparação. Na prática forense a amostra de fala padrão é quase sempre coletada algum tempo após o incidente, sendo neste caso, posterior à amostra de fala padrão. Nesse sentido, pesquisas com coletas de caráter transversal são encorajadas.
- Análise de diferentes estilos de fala: diferentes estilos de fala são comparados, como no caso de uma gravação no contexto de um diálogo e um material de fala coletado a partir de entrevista. Em contextos forenses, o material de fala questionado é quase sempre confrontado com um material padrão coletado num estilo diferente, como no contexto de um interrogatório, por exemplo.

- Análise de diferentes estados/emoções: o estado psicológico de um falante pode implicar modificações nos parâmetros acústicos analisados. Sujeitos em situações espontâneas podem apresentar um diferente comportamento linguístico em comparação a situações em que estão sob efeito de um fator de estresse.
- Análise do efeito de diferentes tipos de degradação: a presença do ruído de fundo ambiental, um dos tipos mais comuns e frequentes de degradação, pode implicar erros no cálculo de medidas acústicas, principalmente os ruídos provenientes de outras falas (falas sobrepostas). O disfarce vocal, por sua vez, outro tipo de degradação, pode causar modificações substanciais nos parâmetros acústicos, como na qualidade vocal, no fluxo discursivo, no padrão entoacional, nas taxas e na intensidade dinâmica da fala, além de variações nos valores de frequência fundamental da voz (MATHUR, CHOUDHARY E VYAS, 2016). Entender o impacto desses tipos de degradações é fundamental para a tarefa de comparação de locutores num contexto de investigação. Ver Passetti; Barbosa (2018). Sobre os efeitos de outros fatores, a exemplo da compressão de áudio e do microfone nos parâmetros acústicos da voz, ver Cavalcanti et al., (2021).
- Gravações com diferentes canais de transmissão/captação: Nesse caso, diferentes tipos de gravação são realizadas, como a comparação de gravações via telefone, WhatsApp, a partir de gravadores e microfones não profissionais, sempre analisadas em relação à gravação em contextos ideais. Aqui, busca-se entender o impacto das diferentes formas de captação e transmissão do registro acústico na tarefa de comparação de locutor.
- Gravações com sujeitos relacionados: gravações realizadas com indivíduos que compartilham material genético em comum (gêmeos monozigóticos, gêmeos heterozigóticos, irmãos) ou compartilham o mesmo ambiente linguístico é uma das estratégias que pode ser adotada num contexto experimental para a análise do grau de variação entre locutores semelhantes e de seus possíveis limites. Outro fator importante a ser considerado diz respeito ao grau de familiaridade entre os locutores, que é quase sempre reduzido num contexto de interrogatório. Sobre estudos com locutores geneticamente relacionados ver Cavalcanti (2020), Cavalcanti, Eriksson, Barbosa (2021).

Algumas das estratégias supracitadas ajudam a aproximar o contexto experimental do contexto real em que se dão as análises forenses, minimizando possíveis vieses comuns e inerentes a essa primeira condição. Tais estratégias

podem ainda serem ou não combinadas, a depender do objetivo da investigação e do nível de proximidade ou distanciamento que se busca do contexto real em que se dão as comparações forenses de locutor.

## REFERÊNCIAS

ARANTES, P., ERIKSSON, A., LIMA, V. Minimum Sample Length for the Estimation of Long-term Speaking Rate. Proc. 9th International Conference on Speech Prosody, p. 661-665, 2018.

ARANTES, Pablo; ERIKSSON, Anders. Temporal stability of long-term measures of fundamental frequency. Dublin: ISCA, 2014. p. 1.149-1.152, 2014.

BARBOSA, P. A. Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. Revista de Estudos da Linguagem, v. 20, n. 1, p. 11-27, junho, 2012.

BARBOSA, P.; MADUREIRA, S. Manual de Fonética Acústica Experimental: aplicações a dados do Português. São Paulo: Cortez, 2015.

BJURSÄTER, U. Speaking styles and Phonetic variation. Term paper autumn, Department of Linguistics, Stockholm University, 2004.

BOERSMA, PAUL & WEENINK, DAVID. Praat: doing phonetics by computer [Computer program], 2021. Version 6.1.39, retrieved 8 February 2021 from <http://www.praat.org/>.

CAVALCANTI, J. C.; ERIKSSON, A.; BARBOSA, P. A. Acoustic analysis of vowel formant frequencies in genetically-related and non-genetically related speakers with implications for forensic speaker comparison. Plos one, v. 16, n. 2, p. e0246645, 2021.

CAVALCANTI, J. C., ENGLERT, M., OLIVEIRA JR., M., & CONSTANTINI, A. C. Microphone and Audio Compression Effects on Acoustic Voice Analysis: A Pilot Study. Journal of Voice, 2021.

CAVALCANTI, J. C. Análise fonético-acústica em gêmeos idênticos: os limites da variação entre locutores. (151-172). In: BARBOSA *et al.*, Análise Fonético-Forense em tarefa de Comparação de Locutor. Campinas, SP: Millennium Editora, 2020.

CAVALCANTI, J. C. LUCENTE, L. BARBOSA, P. Laryngealization, Gender and Speakers' Distinctiveness in Brazilian Portuguese. Proceedings of the 9th International Conference on Speech Prosody, Poznań, Poland, 2018.

CONSTANTINI, A. C. BARBOSA, P. A. Análise de parâmetros prosódicos em diferentes relações sinal-ruído. Anais do Colóquio Brasileiro de Prosódia da Fala, v. 2, Brasília, 2013.

DECOSTER, W. VAN GYSEL A. VERCAMMEN W.D, DEBRUYNE J, F. Voice similarity in identical twins. Acta OtoRhinoLaryngologica Belgica, 2001.

HOLLIEN, H. About forensic phonetics. Linguistica, Vol 52 No 1: Spoken discourse, 27-53, 2012.

HOLLIEN, H. The acoustics of crime. New York: Plenum Press, 1990.

JESSEN, M. Forensic phonetics. Language and Linguistics Compass, 2(4), 671-711, 2008.

KREIMAN, J. SIDTIS, D. Voices and listeners: Toward a model of voice perception. Acoustics Today, p.7-15, 2011.

KÜNZEL, H. J. Beware of the 'telephone effect': the influence of telephone transmissions on the measurement of formant frequencies, Forensic Linguistics, 8 (1), 80-99, 2001.

LABOV, William. Padrões sociolinguísticos. São Paulo: Parábola, [1972] 2008.

MATHUR S, CHOUDHARY BSK, VYAS CJM. Effect of Disguise on Fundamental Frequency of Voice. J Forensic Res 7:327. 2016.

MORRISON, G. S. Forensic voice comparison and the paradigm shift. Science & Justice, Volume 49, Issue 4, p. 298-308, 2009.

MURRAY, I.R. & ARNOTT, J.L. Toward the simulation of emotion in synthetic speech: A review of the literature on vocal emotion. Journal of the Acoustical Society of America 93 (2), 1097-1107, 1993.

NOLAN, F. The phonetic bases of speaker recognition. Cambridge: Cambridge University Press, 1983.

PASSETTI, R. R.; BARBOSA, P. A. Análise fonético-acústica de disfarces vocais com aplicações para a fonética forense. Cadernos de Estudos Linguísticos, v. 60, n. 3, p. 567-583, 22 nov. 2018.

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”. Dissertação de mestrado. Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas SP, 2015.

PASSETTI, R. R. Estudo acústico-perceptual do estilo de fala telefônico com implicações para a verificação de locutor em português brasileiro. Tese de doutorado, Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, SP, 2018.

PINTO, Aline Gomes Lustosa; CRESPO, Agrício Nubiato; MOURAO, Lucia Figueiredo. Influência do tabagismo isolado e associado a aspectos multifatoriais nos parâmetros acústicos vocais. *Braz. j. otorhinolaryngol.*, São Paulo, v. 80, n. 1, p. 60-67, Feb. 2014.

ROSE, P. *Forensic speaker identification*. London: Taylor & Francis, 2002.

SAN SEGUNDO, E. *Forensic speaker comparison of Spanish twins and non-twin siblings*, PhD dissertation, Menéndez Pelayo International University & CSIC, 2014.

XU, Y. “In defense of lab speech”, *Journal of Phonetics*, 38: 329-336, 2010.

