

Dados de referência de F0 em corpus de falantes do Português Brasileiro na variedade falada na Capital Paulista

G.A. da Silva ^{a,*}

^a *Laboratório de Áudio, Acústica e Ciência Forense da Fala do Núcleo de Identificação Criminal
Instituto de Criminalística do Estado de São Paulo - Instituição Científica e Tecnológica
Superintendência da Polícia Técnico Científica de São Paulo*

*Endereço de e-mail para correspondência: gerson.gas@policiacientifica.sp.gov.br Tel.: +55-11-3811-7102.

Recebido em 16/09/2021; Revisado em 14/03/2022; Aceito em 13/07/2022

Resumo

Em que pese sua limitação, a frequência fundamental sempre foi amplamente utilizada como parâmetro forense e, de acordo com os resultados da primeira pesquisa internacional sobre práticas de Comparação Forense de Locutor, é utilizada por quase todos os foneticistas forenses que realizam algumas análises que não são puramente auditivas. Essa importância se reflete em inúmeras produções científicas sobre distribuições de F0 de longo termo descritas nas mais diversas línguas. Na medida em que inexistem tais publicações para falantes do português brasileiro, este artigo visa a preencher tal lacuna, descrevendo estatisticamente as medidas mais utilizadas por peritos em uma distribuição de frequência fundamental de longo termo em um *corpus* de 100 falantes do Português Brasileiro da variedade subpadrão, em registro modal falada na capital do Estado de São Paulo.

Palavras-Chave: População Relevante; Creaky Voice; Frequência Fundamental; Linguística Forense; Ciência Forense da Fala.

Abstract

Despite its limitation, fundamental frequency is widely used as a forensic phonetic parameter. According to the results of the first international survey on forensic speaker comparison practices, it is used by nearly all forensic phonetic experts who perform non-auditory analyses. This importance is reflected in numerous scientific productions on long-term F0 and described in diverse languages. However, no such publications exist for Brazilian Portuguese speakers. To overcome this research gap, the current study aims to provide a statistical description of measures most frequently used by experts in a long-term fundamental frequency distribution. The study employs a corpus of 100 Brazilian Portuguese speakers of the variety substandard in a modal register spoken in the capital of the State of São Paulo.

Keywords: Relevant Population; Creaky Voice; Fundamental Frequency; Forensic Linguistics; Forensic Speech Science.

1. INTRODUÇÃO

A ação da glote para produzir um som - a fonação - depende da maneira pela qual o fluxo de ar interage com as forças musculares e elásticas das pregas vocais. Essas interações resultam na fonte de produção dos sons vocálicos e de todos os outros sons vozeados. Os conceitos básicos da produção da voz são atualmente descritos segundo a universalmente aceita Teoria Mioelástica Aerodinâmica. A Teoria Mioelástica Aerodinâmica combina aspectos das teorias de Van Den Berg e de Perello para estabelecer que a fonação é o resultado de uma interação entre forças respiratórias de natureza aerodinâmica com as forças elásticas dos tecidos

laríngeos, tendo a contração dos músculos da laringe e os movimentos ondulatórios da túnica mucosa especial importância nesse processo [1-3]. A voz¹ é produzida por movimentos das pregas vocais que interrompem a saída do fluxo aéreo gerado nos pulmões [4-7]. O ar exalado pelos pulmões atravessa toda a estrutura do aparelho

¹ É de interesse pontuar que “voz” foi acima definida em seu sentido estrito, ou seja, como um fenômeno físico e fisiológico [4] - como produto único e exclusivo da atividade laríngea - , o que é suficiente para nosso propósito de definir F0. Definições mais abrangentes de voz a consideram de forma multidimensional, isto é, como o produto da ação conjunta dos subsistemas laríngeo e supralaríngeo [5]. Similarmente, “Qualidade de voz”, entendida como a impressão perceptiva da voz pelo ouvinte, em sua definição ampla, refere-se ao resultado final de uma sequência complexa de eventos cognitivos, fisiológicos e acústicos [6].

tráqueo-pulmonar e coloca em vibração as mucosas das pregas vocais. Tais movimentos são controlados por características biomecânicas das próprias pregas vocais, pela magnitude da pressão subglotal e pelo controle neural. Diz-se, portanto, que a vibração das pregas vocais é resultado da interação passiva das forças mioelásticas e aerodinâmicas. As primeiras se referem à resistência da musculatura laríngea à passagem do fluxo de ar subglotal. As forças aerodinâmicas estão relacionadas ao efeito Bernoulli², responsável pela pressão negativa que aproximará as mucosas - desde que estas sejam suficientemente flexíveis, a glote esteja suficientemente estreita e o fluxo subglotal seja suficientemente alto. A frequência básica de vibração das pregas vocais, dada pelo Máximo Divisor Comum entre os componentes da onda periódica complexa, é referida como frequência fundamental (doravante F0) e é percebida como *pitch*. Alguns autores, a fim de indicar o uso de F0 em situações concretas de fala, utilizam-se do termo Frequência Fundamental de Fala [8-9]. Em que pese sua limitação como parâmetro fonético forense, sobretudo devido à sua alta intravariabilidade³ [10], a F0 sempre foi bastante difundida entre pesquisadores e peritos, tendo se mostrado promissora já nos primeiros estudos de reconhecimento automático de locutor [11]. Embora F0 seja um parâmetro fonético muito bem documentado do ponto de vista forense e de tal forma amplamente utilizada, conforme apontam os estudos de Gold & French [12,13], no português brasileiro (doravante PB) há poucas análises descritivas documentadas de tal parâmetro envolvendo grandes grupos homogêneos para fins forenses [14]. Alguns desses estudos são dirigidos a outros propósitos que não os diretamente relacionados à fonética forense, e focados, sobretudo, em medidas de curto termo, como, por exemplo, em Behlau [15], em que medidas de F0 foram feitas em pontos médios de segmentos vocálicos. Na tarefa de Comparação Forense de Locutor, entretanto, medidas associadas a distribuições de longo termo de F0 são, em geral, bem mais apropriadas, uma vez que parâmetros estatísticos associados à F0 em longo intervalo de tempo, diferentemente de suas medidas de curto termo, não variam muito em função de fatores contextuais (fonéticos, entoacionais, discursivos, e assim por diante) [16]. Os aspectos oriundos de parâmetros de longo termo, ao contrário, são essencialmente independentes do conteúdo

da mensagem falada, sendo, idealmente, traços invariantes no tempo⁴ que refletem traços estáveis do falante.

De acordo com a primeira pesquisa internacional sobre práticas de Comparação Forense de Locutor, conduzida por Gold e French [12], o valor médio de F0 é a medida de tendência central mais amplamente utilizada pelos peritos entrevistados. Dentre os peritos que conduzem análise não puramente auditiva, 94% relataram se utilizar da média, 41% disseram avaliar a mediana e 34% disseram se utilizar da moda como parâmetro fonético forense. A mesma pesquisa aponta também que o desvio padrão é muito utilizado como medida de variabilidade (dispersão em torno da média): 72% dos peritos reportaram se utilizar de tal medida. Por fim, os peritos relataram que o valor médio de F0 - como medida central - mais o desvio padrão - como medida de variabilidade - é a combinação mais usada para análise fonética forense. Adicionalmente, a pesquisa mostrou que não é raro o uso do valor de base de F0, como medida de tendência central (reportado por 25% dos entrevistados).

Este artigo se pauta numa perspectiva fonético acústico estatística ao proceder à descrição estatística de algumas das medidas de F0 de longo termo. Elencamos o valor médio e o valor de base, como medidas de tendência central (ou localização), e o desvio padrão como medida de dispersão. Também descrevemos a distribuição do coeficiente de variação. Tal descrição foi feita a partir de um corpus forense composto por 100 falantes do PB em fala semiespontânea na norma popular e na variedade utilizada na capital do Estado de São Paulo. Os resultados obtidos foram comparados a estudos sobre o tema conduzidos em corpora de outras línguas.

2. A FREQUÊNCIA FUNDAMENTAL COMO PARÂMETRO FONÉTICO FORENSE

De acordo com Nolan [10] e Rose [16], as propriedades ideais de um parâmetro a ser utilizado em um exame de Comparação Forense de Locutor, são:

- Alta intervariabilidade;
- Baixa intravariabilidade;
- Resistência à tentativa de disfarces ou imitações;
- Possuir alta frequência de ocorrência em material relevante;
- Ser robusta na transmissão

² A chamada pressão de Bernoulli é a diferença em pressão que deve existir para manter a energia invariante em um duto com seção transversal variada. O fluxo total de energia depende da pressão e da velocidade da partícula de ar. A velocidade da partícula de ar em um estreitamento é maior do que a pressão de ar em uma área adjacente mais larga. Assim, pelo fato de o fluxo total de energia se manter constante, a pressão em uma área estreita é menor do que a pressão em áreas adjacentes.

³ A questão da intravariabilidade será aprofundada na seção 2 deste artigo "A Frequência Fundamental Como Parâmetro Fonético Forense".

⁴ "Longo termo" não deve ser tomado no sentido diacrônico de amostras não contemporâneas e sim como uma categoria usada para classificar um determinado parâmetro da fala. Nesse sentido, temos traços de curto termo e traços de longo termo. Um traço de curto termo, por exemplo, é aquele cuja análise exige o isolamento de trechos de sinal de fala com duração limitada e se associam a unidades abstratas como fonemas, sílabas, unidades entoacionais, e assim por diante. Traços de longo termo, por outro lado são tomados em porções maiores do enunciado ou em todo o enunciado (Espectro de Longo Termo, Formante de Longo Termo e "Qualidade de Voz", por exemplo).

- Ser relativamente fácil de extrair e de se medir.

A F0 contempla satisfatoriamente muitas das *desiderata* propriedades acima delineadas, a saber, robustez, mensurabilidade e disponibilidade, podendo ser útil mesmo em algumas estratégias de disfarce, como nos mostra Künzel [17]. A F0 é um parâmetro robusto e pode ser extraída com relativa facilidade mesmo de gravações que apresentam baixa qualidade acústica⁵. Medidas de F0 também são menos afetadas pela transmissão telefônica comparada a outros parâmetros segmentais [18]. Uma vez que a F0 está associada ao vozeamento - e tendo em vista que a maior parte dos sons da fala é vozeada- trata-se de um parâmetro bastante disponível. Além disso, a F0 apresenta uma vantagem adicional: a de não ser afetada pelo conteúdo lexical do enunciado (nas línguas não-tonais). Vários estudos têm apontado que medidas de F0 possuem alta intervariabilidade⁶, o que se explica, em parte, por sua correlação com a anatomia das pregas vocais. Essa relação está representada na Figura 1, reproduzida de Titze [19], em que o autor, partindo de algumas suposições razoáveis, - como a de que a tensão longitudinal nas pregas vocais é a mesma para todos os falantes - mostra que, *ceteris paribus*, a F0 é inversamente proporcional ao efetivo comprimento das pregas vocais (mais precisamente, da porção mioelástica que efetivamente vibra). Assim, quanto mais longo o comprimento das pregas vocais, menor de F0 a ele associado. Essa porção é 60% maior nos homens do que nas mulheres, o que explica o fato da F0 ser, em média, mais alta nas mulheres do que nos homens. A Figura 1 explicita tal relação inversa entre a F0 e o comprimento das pregas vocais com a tensão e massa fixa por unidade de comprimento (aqui representada por L_m). Observe-se que a hipérbole tem a forma $1700/L_m$ e que o valor $F_0 = 170\text{Hz}$ corresponde à frequência produzida por um falante adulto do sexo feminino cujas pregas possuem 10mm de comprimento.

Em sua meta-análise sobre intervariabilidade de F0 entre falantes do sexo masculino e feminino conduzida em diversas línguas - em sua maioria europeias-em texto lido e fala espontânea, Traummüller e Eriksson [20] obtiveram valores médios de 124 Hz e 211 Hz para falantes masculinos e femininos, respectivamente. O gráfico de dispersão da Figura 2, construído com base na meta-análise de dados de fala espontânea levada a efeito pelos autores, mostra que há variabilidade não só entre locutores de sexo diferentes (locutores do sexo feminino - em azul - são facilmente discriminados, por meio da

linha de regressão - em pontilhado -, dos falantes masculinos - em cinza), como também entre línguas, isto é, o estudo suporta a hipótese de que F0 é um parâmetro específico da língua, conforme sugerido por Rose [16].

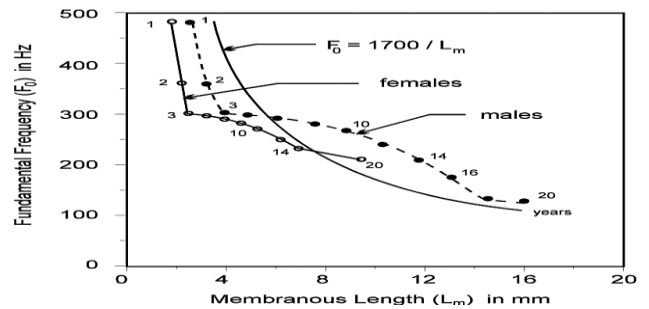


Figura 1. Média da frequência fundamental de fala como função do comprimento das pregas vocais. Fonte: Titze [19].

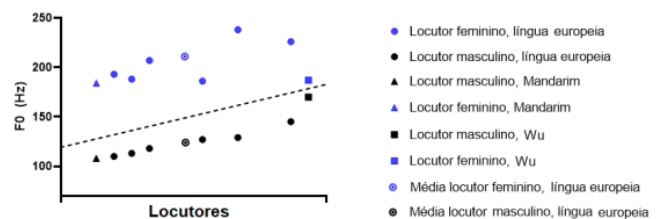


Figura 2. Gráfico de dispersão de F0 de longo termo construído com base na meta-análise de Traummüller e Eriksson [20].

É interessante notar que o gráfico da Figura 2 mostra que, de forma geral, há significativa intervariabilidade entre os sexos, com exceção de falantes do dialeto Wú. De fato, falantes masculinos nesse dialeto chinês apresentam, em média, um alto valor de F0 e sua diferença em relação aos falantes femininos é bem menor do que o restante das comunidades linguísticas observadas. Os autores argumentam que tal valor se deve mais à convenção social do que a diferenças anatômicas propriamente ditas. Embora afirmem que tal fenômeno não é único - ocorrendo também no dialeto falado na província sueca de Småland, por exemplo -, os autores pontuam que os falantes, em geral, utilizam-se, predominantemente, das regiões mais baixas de sua tessitura em conformidade com o princípio de minimização do esforço fisiológico.

A intervariabilidade é condição necessária para que uma dada medida acústica possa ser utilizada como parâmetro fonético forense. A F0 é amplamente utilizada como parâmetro forense justamente por manifestar intervariabilidade também entre indivíduos de um mesmo sexo - isso ocorre por conta não só das diferenças anatômicas entre os indivíduos como também em relação à forma pela qual estes manipulam seus ajustes do trato vocal. À guisa de ilustração, a Tabela 1 apresenta descritores relativos a distribuições de F0 de Longo termo

⁵ Uma dentre as medidas de inteligibilidade de áudio é a razão sinal-ruído (em inglês *signal-to-noise ratio* (SNR)).

⁶ Interviabilidade - variabilidade extrínseca ou variabilidade interfalante- refere-se ao fato de que diferentes falantes, produzindo o mesmo conteúdo acústico, apresentarão valores diferentes do mesmo parâmetro. A intervariabilidade garante que diferentes locutores não de diferir quanto às suas características fonéticas.

de seis locutores masculinos.

Locutor	Média	Moda	Mediana	Assimetria	Curtose	Normalidade
1	111,6	113	115,6	-0,881	0,656	<0,001
2	162,5	149	155,4	1,047	0,973	<0,001
3	126,4	81	121,4	0,611	-0,582	<0,001
4	143,6	148	147,4	-0,520	-0,175	<0,001
5	138,0	122	135,3	0,193	-0,538	<0,001
6	114,8	98	111,3	0,309	-0,219	<0,001

Tabela 1. Inter- variabilidade nas características da distribuição de F0 de seis locutores.

Note-se que os falantes não só diferem em termos de medidas de tendência central e curtose, como também em termos de assimetria. Observando os *p-values* do teste paramétrico de Kolmogorov-Smirnov, notamos que nenhuma das distribuições segue uma distribuição normal e que são, portanto, assimétricas. Deve-se atentar, porém para o fato de que, embora uma distribuição de F0 de longo termo possua, tipicamente, uma assimetria positiva⁷ [12,16,20,21], há locutores (locutor 1 e locutor 4) que apresentam uma assimetria negativa. Outra informação de relevância forense que a tabela nos fornece é que os indivíduos se distinguem não só em função dos estimadores de localização, como também em função da assimetria e do grau de achatamento de suas respectivas distribuições de F0 de longo termo. Kinoshita, Ishihara e Rose [22] já haviam notado que, para além das medidas de tendência central, outros parâmetros da distribuição talvez possam ser usados para descrever F0 fornecendo evidências mais fortes.

A F0 seria um parâmetro ideal se não possuísse a desvantagem de também variar num mesmo indivíduo⁸. Essa variação pode se manifestar, não só em função de fatores puramente linguísticos - como variações em função do contexto fonético ou do nível de proeminência prosódica, por exemplo - mas também em função de uma plethora de variáveis exógenas agrupadas, segundo Braun [23], em três grandes categorias de variáveis exógenas, a saber, as de natureza fisiológica [24-27], psicológica [28] e técnica [29-31].

Na medida em que a intravariabilidade introduz complicações para a análise de F0 é importante

⁷ Assimetria positiva é típica para um dado falante [16,20] o que se deve, muito provavelmente, à relação exponencial entre a alongação e a tensão do músculo tireovocal (*vocalis*) nas pregas vocais [21].

⁸ A voz não é um dado constante para um determinado falante. Cada som da fala é um evento único porque é produzido com diferente energia a cada tempo. Ninguém diz uma mesma palavra exatamente da mesma maneira duas vezes e, ao menos que se esteja lidando com gravações de uma mesma emissão, é certo que se deparará com o fato de que a variabilidade da fala se manifesta em um mesmo indivíduo, ou, em outras palavras, a par da variação entre diferentes locutores, há uma intravariabilidade da fala [82]. A intra-variabilidade pode ser referida como variabilidade intrínseca, ou variabilidade intra - falante.

considerar seus efeitos juntamente com aqueles relacionados com a intervariabilidade [32] sempre que se for analisar o potencial discriminativo do referido parâmetro. Em estatística inferencial, o poder discriminativo de uma certa distribuição é referido como *F-ratio* [10] e é dado pela razão entre intervariabilidade/ e intravariabilidade. Alta variabilidade intra-sujeito resulta em baixo *F-ratio* e, portanto, baixo potencial discriminativo. Dessa forma, por possuir uma alta intravariabilidade, a F0 deixa de ser um bom parâmetro em muitas situações com que normalmente nos deparamos num cenário forense real, em que a tarefa comparativa envolve, via de regra, diferentes canais, diferentes sessões de gravação ou fatores de degradação, além de outras fontes de variabilidade. Boss [33], por exemplo, relata uma diferença de 88 Hz para um mesmo locutor, entre F0 produzida em entrevista policial e F0 encontrada na gravação questionada. Em casos reais, entretanto, não é raro encontrar discrepâncias de tais magnitudes, referentes a um mesmo locutor, normalmente devido à variabilidade normalmente causada por uma complexa combinação dos fatores elencados por Braun.

3. BIMODALIDADE EM FUNÇÃO DE OCORRÊNCIAS DE CREAKY VOICE

Embora distribuições da média da F0 de longo termo para ajuste modal e esforço vocal normal sejam tipicamente unimodais, indivíduos podem apresentar distribuições bimodais de F0 de longo termo. Isso pode acontecer, por exemplo, quando o locutor se vale de um esforço vocal acima do normal, como ocorre num debate parlamentar [34]. Na tarefa de Comparação Forense de Locutor a bimodalidade se dá, geralmente, em função de ocorrências de *creaky voice* [35,36].

Creaky voice funciona como parâmetro contrastivo em muitas línguas, como, por exemplo, nas línguas Wakashan [37] ou em Jalapa Mazatec, uma língua falada no México [38]. No PB - e em outras línguas também - o ajuste *creaky voice* é também empregado para sinalizar fronteiras terminais e não terminais de enunciado, por exemplo [39-41]. Embora *creaky voice* também tenha sido objeto de vários estudos variacionistas da terceira onda [42-43], focados na variação estilística de modo geral, há vários outros estudos que apontam que *creaky voice* é um aspecto normal da fala na maioria dos falantes, não importa língua, gênero ou quaisquer outros fatores [44-45]. De fato, *creaky voice* é o ajuste não neutro mais utilizado em uma conversação [46].

Sob a perspectiva da Fonética Articulatória, *creaky voice* ocorre quando as pregas vocais estão firmemente comprimidas, mas abertas o suficiente para permitir o vozeamento. Há alta tensão adutiva, alta tensão medial e baixa tensão longitudinal, o que contrasta, por exemplo,

com o ajuste modal no qual há uma moderada tensão nas três áreas [47]. Como resultado, temos uma maior duração adutiva das pregas vocais durante o ciclo fonatório do que durante o ajuste modal. Além disso, as pregas vocais se apresentam relativamente curtas e espessas, o que contribui para o baixo F0. A intensidade é muito débil e a observação laringoscópica mostra pregas vocais curtas e grossas. O espectrograma da Figura 3 mostra uma *creaky voice* prototípica [48], referente a um dos locutores excluídos da população relevante. Esse tipo de ajuste não neutro é caracterizado por uma F0 baixa e irregular (o que, acusticamente, traduz-se por uma baixa relação harmônico/ruído) e constrição glotal. Essa vibração irregular se dá tanto em termos de duração quanto em amplitude, resultando, respectivamente, em um maior *jitter* e *shimmer*⁹. Ainda do ponto de vista acústico, *creaky voice* se caracteriza, também, por possuir um baixo quociente de abertura - o que implica, acusticamente, em baixo H1-H2¹⁰, na medida em que há um maior reforço de H2 quanto menor a abertura glotal [49-52]. A forma de onda que pode ser vista na parte superior da Figura 3 mostra a existência de pulsos irregulares.

Além disso, o espectrograma da parte inferior da figura mostra que tais irregularidades podem levar a falhas no algoritmo de rastreamento de F0.

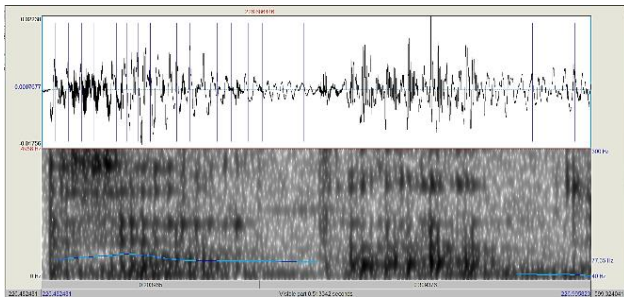


Figura 3. Sinal de fala com ocorrência de *creaky voice*.

A curva de densidade à esquerda na Figura 4 é referente ao mesmo locutor cujo espectrograma é mostrado na Figura 3. Note-se que, além de um pico de frequência em 115Hz, há também valores frequentes próximos dos 50 Hz, os quais se referem a ocorrências de *creaky voice*. A curva que pode ser vista à direita mostra outro locutor cujo registro é, predominantemente, modal.

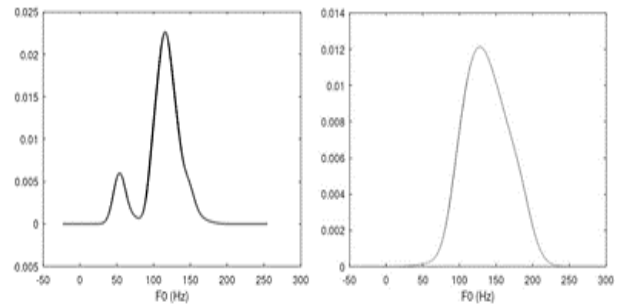


Figura 4. Curvas de densidade por método Kernel de dois locutores.

As Figuras 3 e 4, combinadas, ilustram dois problemas relacionados ao fato de se incluir *creaky voice* em bancos de dados de fala. O primeiro diz respeito à falta de acurácia dos algoritmos de rastreamento de F0 (*pitch tracking*). O segundo problema é que parâmetros estatísticos de localização (como a média) não são precisos para descrever distribuições multimodais. Na literatura fonético-forense não há uma metodologia única para se lidar com *creaky voice*. Alguns autores, por exemplo, tratam a distribuição inteira como se fosse um único objeto [35,36] e, outros, simplesmente desconsideram os falantes que apresentam muitas ocorrências de *creaky voice*¹¹ ou mesmo omitem as ocorrências de *creaky voice*, concentrando-se apenas na distribuição modal [53]. Há ainda autores que tentaram incluir as amostras que continham *creaky voice* como uma característica do falante a ser analisada separadamente em uma análise global [54].

4. INFORMAÇÕES METODOLÓGICAS

O uso de bases de referência para avaliação da tipicidade em casos forenses reais é um procedimento bastante difundido entre peritos [13] e integra todo um processo epistemológico de mudança de paradigma, em curso em várias disciplinas criminalísticas [55]. Para a construção de tais bases de referência, os peritos se servem dos métodos da sociolinguística a fim de se obter a população relevante para o caso em tela. A noção de população relevante foi pela primeira vez estabelecida por Coleman e Walls [56] e, desde então, tem sido objeto de várias reformulações [57,58]. Frequentemente, população relevante pode ser definida como o conjunto de falantes de mesmo sexo e língua do locutor de interesse no material questionado. “Língua”, muitas vezes, pode ser definida de modo mais restrito, de forma a incluir o

⁹ *Jitter* e *shimmer* são índices de perturbação de curto termo de F0. Perturbação se refere à irregularidade de vibração das pregas vocais. Se a irregularidade estiver relacionada com o tempo de vibração, é chamada de *jitter* e se estiver relacionada com a amplitude da vibração, é chamada de *shimmer*.

¹⁰ H1 e H2 se referem às amplitudes do primeiro e do segundo harmônico, respectivamente. A diferença entre H1 e H2 está possivelmente ligada à velocidade das forças aerodinâmicas das pregas vocais. Quanto maior o valor do primeiro harmônico, em relação ao segundo, menor é o coeficiente de contato e maior a velocidade de fechamento das pregas vocais.

¹¹ Conforme explicamos na seção anterior, locutores que apresentam significativas ocorrências de *creaky voice* apresentam distribuições bimodais de F0 de longo termo. O pico em baixas frequências representa o ajuste fonatório *creaky voice*. O pico maior, mais à direita, representa a fonação modal. No cálculo de F0 médio levado a efeito por Hudson [36], considerou-se a média dos valores médios dos dois picos, abaixando-se, destarte, o valor médio de F0.

conjunto de realizações concretas, consagradas pelo uso e que liga, de forma mais imediata, o material de fala encontrado na cena do crime a um determinado grupo social, uma região e um determinado estilo, por exemplo, o que corresponde ao conceito de norma [59].

Nesse estudo, descrevemos uma base de referência composta por 100 amostras de fala em produção semi-espontânea, de indivíduos falantes do PB, especificamente da variedade urbana da capital do estado de SP, em norma não culta. Para constituição de tal base de referência, foram selecionados falantes do sexo masculino cujas idades variavam de 18 a 62 anos (média 28.6; desvio padrão 6.2 anos). As gravações foram realizadas em uma sessão apenas, durante o período da tarde. Nos parágrafos seguintes, detalhamos de que modo a seleção dos 100 sujeitos foi feita de forma a atender para os requisitos de variação estilística, (diafásica), de variantes coletivas (diatópicas e diastráticas), de duração da amostra de fala requerida para caracterizar genuinamente o falante e do ajuste fonatório de qualidade de voz.

Durante o procedimento de seleção dos locutores, procurou-se obter o vernáculo [60-62]. Sabe-se, porém, que nenhuma entrevista é isenta de algum grau de formalidade e isso decorre do paradoxo do observador: o analista, ao realizar uma entrevista, almeja a obtenção do vernáculo, que se espera que ocorra quando o falante não está sendo observado¹². Dessa forma, diz-se “fala semi-espontânea”, pois na presença do entrevistador, entrevistado nunca terá uma fala espontânea¹³. Além disso, na medida em que, em casos reais, a emotividade está sempre presente, não procuramos, durante a entrevista, neutralizar impactos na emissão em virtude de eventual emotividade envolvida. Ao contrário, durante a entrevista procurou-se fazer com que os entrevistados se sentissem à vontade e, sempre que o quisessem, podiam evocar lembranças que envolvessem emoções.

A população descrita nesse estudo foi definida de tal forma a incluir apenas o falar urbano de falantes nativos, socializados na cidade de São Paulo e que habitam a capital do estado. Além da variação diatópica, considerou-se também, a variação diastrática representada pelas diferenças entre o português falado pelas partes mais escolarizadas da população e entre as partes menos escolarizadas. Sabe-se que tais variações, ligadas a estratificação social, podem ser subdivididas em norma culta padrão, norma coloquial (tensa ou distensa) e norma

popular (também chamada de vulgar). A população utilizada neste estudo é composta por falantes das classes populares semi-escolarizadas, que apresentaram, durante a entrevista, uma fala distensa e despreocupada com a norma culta- em relação à qual apresenta grande desvio - e emprego, não raro, de vocabulário gírio como língua especial¹⁴.

Uma questão que sempre se tem colocado quando se fala em distribuições de F0 é qual seria a duração mínima necessária para se caracterizar o falante. Embora afirme que a duração se estabilize por volta de 60s, Nolan [10] reconhece que há estimativas que apontam para durações de estabilização bem menores. Rose [63], por exemplo, descobriu que os valores de F0 de longo termo para sete falantes do dialeto chinês se estabilizariam bem antes de 60s. Resultados semelhantes foram obtidos em Arantes [14] onde, por meio da técnica estatística *change point analysis*, foram investigados os efeitos de três fatores (língua, texto e contemporaneidade) sobre propriedades temporais de três estimadores de localização de F0 de longo termo (Média, Valor de Base e Moda). Os achados em Arantes [14] e Rose [63] têm grande contribuição prática na medida em que, em casos reais, não raro, amostras questionadas são de curta duração. Deve-se pontuar, entretanto, que, diferentemente das amostras questionadas, as amostras padrão, quando coletadas pelos próprios peritos em ambiente controlado, têm a duração que os analistas acham necessária para o exame. Em um experimento com F0 de longo termo acerca da relação entre EER (*Equal Error Rate*) e quantidade de amostras vozeadas, Kinoshita [64] mostrou que, para o japonês, é necessário um mínimo de dois minutos para se obter algo próximo ao desempenho ótimo - correspondendo ao menor valor de EER possível. Na medida em que não havia limitação na duração de amostras coletadas para construção de nossa população e que todas as amostras possuem, no mínimo, 5 minutos de fala líquida, foi-nos possível tomar como base os resultados de Kinoshita [64] para estabelecer uma duração fixa de 2 minutos de fala líquida para cada falante que compõe a base de referência.

É sempre importante lembrar que alguns estimadores de localização, como a média, só são confiáveis quando calculados em distribuições unimodais. Desta forma, em decorrência do que foi exposto na seção 3, a fim de se obter medições mais confiáveis, falantes que apresentavam distribuições bimodais, por conta de várias ocorrências de *creaky voice*, foram descartados da base de referência. Assim, dentre os dois locutores cujas respectivas curvas de densidade estão representadas na **Figura 4**, aquele situado à esquerda representa um bom

¹² O vernáculo - estilo em que é mínima a atenção prestada ao controle do discurso- é , segundo Labov [60], o “estilo de base” do locutor . Tal conceito é oriundo dos estudos sociolinguísticos da primeira onda [61], em que o estilo é analisado a partir do monitoramento da fala e controlado a partir do binômio prestígio/ estigma.

¹³ Num cenário de entrevista, a fala tende a ser menos espontânea ainda se o entrevistado desconfia que tal entrevista tenha, como objeto declarado, a linguagem . Num procedimento de obtenção de material de referência, o suspeito geralmente sabe que seu material de fala será utilizado em uma análise fonética.

¹⁴ No contexto da população em estudo, o conceito se refere ao uso de gírias. O termo “ língua especial” também pode se referir aos jargões profissionais e à linguagem técnica, por exemplo.

exemplo de locutor que não foi incluído na base obtida para este estudo.

4.1. As condições em que as gravações foram obtidas

O material foi obtido com qualidade de estúdio, em uma sala tratada acusticamente a uma taxa de amostragem de 44.100Hz, 16 bits de resolução e em formato wav/PCM, estéreo. Para a captação das amostras, foram utilizados um gravador da marca Zoom, modelo H4n acoplado a uma placa *Edirol UA 25*. Os locutores foram posicionados sentados defronte da cápsula estéreo X/Y do gravador a uma distância fixa de, aproximadamente, 30 cm, evitando efeitos da acústica de campo distante. Utilizando-se do software *Praat*, medidas foram extraídas automaticamente por meio do algoritmo de autocorrelação, com janela de 10 msec e intervalo de 40 a 300 Hz, como sugerido pelos estudos de Hudson *et al.* [45] e os dados de saída foram rodados num script em *Matlab* para obtenção de medidas estatísticas. Amostras que contivessem medidas não confiáveis de F0 (quebras de frequência anormais, valores espúrios e outros artefatos indesejáveis) não foram incorporadas na população utilizada neste estudo.

4.2. Os descritores estatísticos de F0

A estatística descritiva foi utilizada para explorar e resumir os dados coletados. Como medidas de tendência central, elencamos a F0 médio e seu valor de base. Como medida de variabilidade (dispersão em torno da média), o desvio padrão foi descrito tanto em Hz como em semitom. Além dessas medidas avaliamos o coeficiente de variação da distribuição da F0.

5. DESCRIÇÃO ESTATÍSTICA DA POPULAÇÃO

A **Tabela 2** resume os descritores de F0 selecionados, seus valores médios, de assimetria e de curtose.

Descritor	Valor Médio	Assimetria	Curtose
F0 de Longo Termo (Hz)	133,2	0,07	0,14
F0 valor de Base (Hz)	94,6	-0,20	0,11
Desvio Padrão (Hz)	26,9	0,69	0,47
Desvio Padrão (semitom)	2,42	0,09	0,14
Coeficiente de Variação (%)	20,3	0,74	0,66

Tabela 2. Descritores extraídos.

Na tarefa de Comparação Forense de Locutor, é muito importante a expressão do valor central de F0 para um

dados falante, na medida em que essa caracteriza sua configuração fonatória habitual [65] sendo a razão pela qual tal medida seja amplamente empregada nas análises de cunho fonético forense. Na **Figura 5**, apresentamos o histograma referente à distribuição de F0 de longo termo obtido no presente estudo.

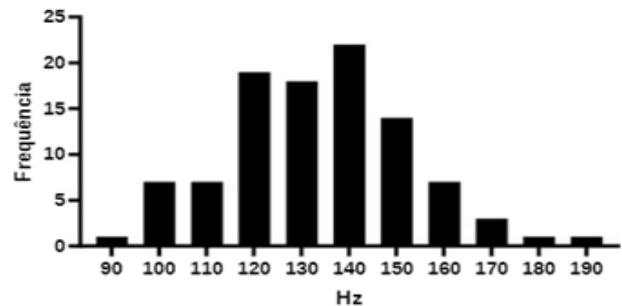


Figura 5. Histograma mostrando distribuição das médias de F0 para 100 falantes masculinos do PB.

O gráfico de frequência acumulada da distribuição de F0 médio mostrado na **Figura 6** nos permite constatar que cerca de 65% das frequências médias se encontram entre 115 e 150 Hz. Trata-se de um intervalo ligeiramente maior dos que os 27Hz para os 60% de falantes alemães, de acordo com os estudos de Künzel [66].

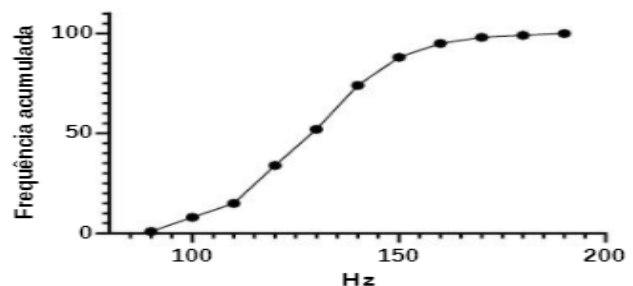


Figura 6. Frequência cumulativa da média de F0.

Na **Figura 7**, observa-se a curva de frequência cumulativa da distribuição dos valores médios de F0 (em azul) sobreposta à curva cumulativa de uma distribuição normal (em vermelho). A sobreposição quase perfeita de ambas as curvas indica que a distribuição dos dados de F0 médio é bem simétrica (o que também pode ser constatado por meio de *Q-Q plot* ou através da observação do valor de *skew* de F0 médio fornecido pela **Tabela 2**).

É importante notar que o valor médio auferido para a distribuição de F0 média, a saber, 133,2 Hz, difere dos valores obtidos em outras distribuições encontradas na literatura. Tanto Rose [16] quanto Traunmüller e Eriksson [20] sugerem que F0 seja específico da língua e que a comparação entre os valores de F0 nas diversas línguas possa fornecer suporte a essa hipótese. É, portanto, interessante a comparação da distribuição de F0 médio

aferida no presente com outras distribuições encontradas em outros estudos, como os indicados na **Tabela 3**.

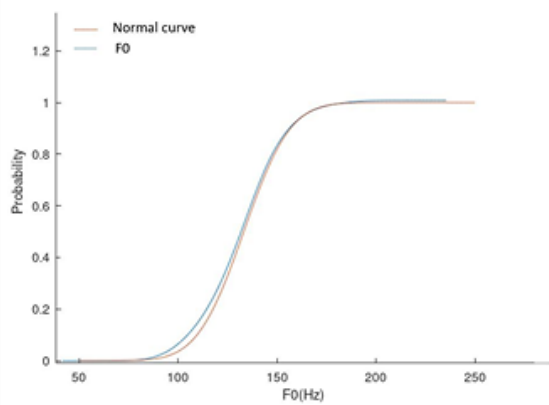


Figura 7. Frequência cumulativa da média de F0 da distribuição normal (em vermelho) e da distribuição de F0 (em azul) do corpus obtido no presente estudo.

Corpus	Idioma	F0 (Hz)
Gold (2014) [35]	Inglês (Reino Unido)	103,6
Hudson (2007) [36]	Inglês (Reino Unido)	106,0
Kinoshita (2005) [64]	Japonês [62]	135,7
Lindh (2006) [67]	Sueco	120,8
Künzel (1989) [68]	Alemão	115,8
Loakes (2006) [69]	Inglês (Austrália)	105,2
Skamitzl & Vaňková (2017) [70]	Tcheco	117,0
Brixen (2018) [71]	Dinamarquês	119,0

Tabela 3. Valores médios de F0 em diversas línguas.

Dentre os valores médios apresentados na **Tabela 3**, os obtidas por Gold [35] para 100 falantes ingleses, e por Lindh [67] - para 109 jovens falantes suecos - merecem alguns comentários, por apresentarem valores bastante distintos e extremos em relação aos estudos auferidos em outras línguas europeias e por apresentarem particularidades nos critérios adotados para a formação da população de referência. Gold [35] obteve um valor médio de 103,6 Hz em uma distribuição de F0 para 100 locutores ingleses. A autora argumenta que o valor relativamente baixo, comparado com valores obtidos em outras línguas seja devido ao fato de F0 talvez ser específico da língua, como sugerido por Rose [16], ou que talvez possa ser explicado pelo fato de que muitos dos locutores que compõem o banco de dados usado - o DyViS (*Dynamic Variability in Speech*) - apresentem significativas ocorrências de *creaky voice*¹⁵. Já o valor

médio de F0 na distribuição obtida por Lindh [67] foi de 120,8 Hz. O próprio atribui tal valor a erros de medição e ao próprio algoritmo que calcula F0, sugerindo que o valor da média das medianas (115,8 Hz) é mais representativo da distribuição do que o valor obtido para a média. É de importância se notar, por outro lado, que a população de referência utilizada no estudo levado a termo por Lindh [67] era formada por falantes cujas idades compreendiam um intervalo bastante estreito - entre 20 e 30 anos. Sabe-se que a idade afeta as aferições de F0 de longo termo, na medida em que, para homens, espera-se que F0 se reduza no intervalo compreendido entre a puberdade e os 35 anos, aproximadamente [20].

A **Figura 8** mostra a sobreposição do histograma referente à distribuição obtida no corrente estudo (histograma da **Figura 5**, em preto) ao histograma referente à distribuição de F0 de longo termo obtido por Gold [35] (em cinza). É importante notar que as distribuições não só diferem em termos de média como também em relação à suas respectivas amplitudes, o que, aparentemente, se deva, dentre outros componentes, a fatores atitudinais ou emocionais decorrentes da forma como se deu a entrevista para a construção do corpus utilizado no presente estudo.

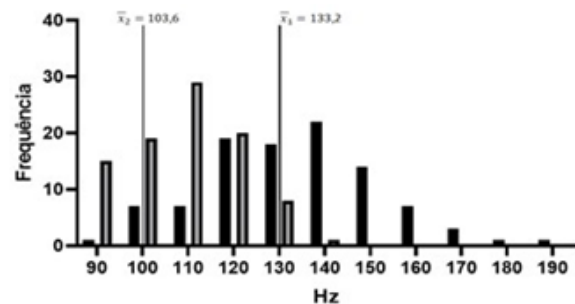


Figura 8. Histograma de F0 médio relativo aos dados obtidos no presente estudo (em preto) sobreposto ao histograma obtido por Gold [35].

Na **Figura 9**, o histograma de F0 médio obtido no presente estudo (histograma da **Figura 5**, em preto) foi sobreposto ao histograma obtido a partir de dados de Lindh [67] (em cinza).

Jessen *et al.* [24] conduziram um importante estudo sobre a variação de F0 em função do esforço vocal. Em tal investigação, os autores compararam os valores e a distribuição de F0 em duas situações: fala espontânea com esforço vocal normal em laboratório e fala espontânea em falantes expostos sujeitos a um ruído gaussiano de 80dB, induzindo-os a um aumento do esforço vocal devido ao efeito Lombard¹⁶ [70-75]. Os autores mostraram que há

¹⁵ O fato de o DyViS ser composto por muitos falantes com significativas ocorrências de *creaky voice* também foi reportado por Hudson et al. [36].

¹⁶ Em 1909, no Hospital de Lariboisière, em Paris, o otorrinolaringologista Etienne Lombard utilizava, em um de seus pacientes, um aparelho que ensurdecia um dos ouvidos, através de um ruído intenso, a fim de examinar a acuidade do outro. Notou, então, que

um deslocamento em direção aos valores mais altos de F0 em função do esforço vocal.

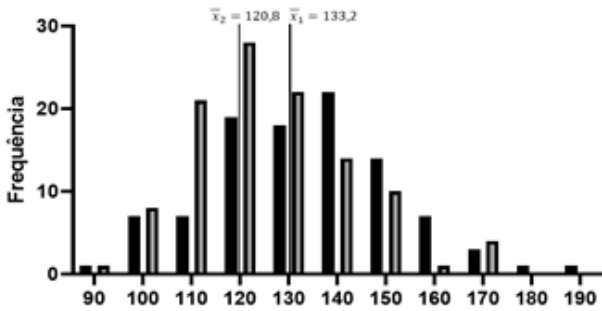


Figura 9. Histograma da figura 6 sobreposto ao histograma obtido por Lindh [67].

Na **Figura 10**, sobrepomos o histograma de F0 médio referente à distribuição obtida neste estudo ao histograma que os autores obtiveram para esforço vocal normal. Comparando a distribuição de F0 obtida nesse estudo (histograma da **Figura 5**, em preto) com os valores de Jessen et al. [24] (histograma em cinza), observam-se semelhanças não só em relação à extensão como também à tipicidade de F0. Em ambas as distribuições, um valor de 80 Hz, por um lado, ou um valor de 165 Hz são instâncias atípicas.

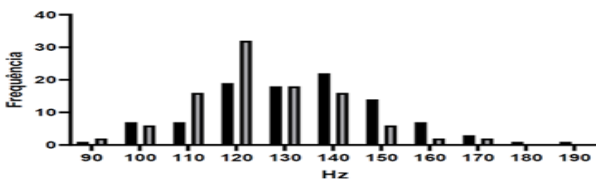


Figura 10. Histograma de F0 médio obtido no presente estudo (em preto) sobreposto ao histograma obtido por Jessen et al. [24].

Conforme mencionado, o desvio padrão é a medida de variabilidade de F0 mais comumente usada segundo Gold e French [12]. O histograma da **Figura 11** mostra a distribuição do desvio em Hz obtida no presente estudo.

A distribuição do desvio é bastante semelhante à obtida por Jessen *et al.* [24] e à obtida por Lindh [67], conforme mostrado nas **Figuras 12 e 13**, respectivamente.

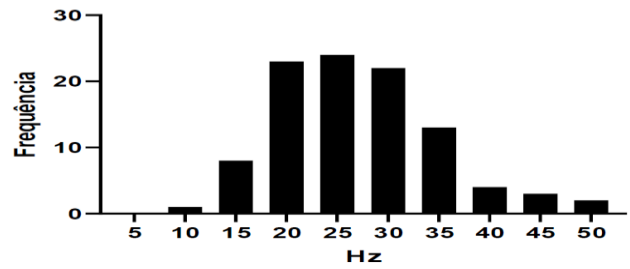


Figura 11. Histograma do desvio padrão em Hz.

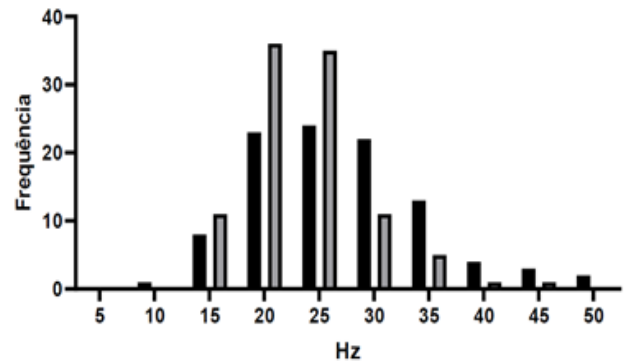


Figura 12. Histograma de desvio em Hz obtido no presente estudo (em preto) sobreposto ao histograma obtido por Jessen *et al.* [24] (em cinza).

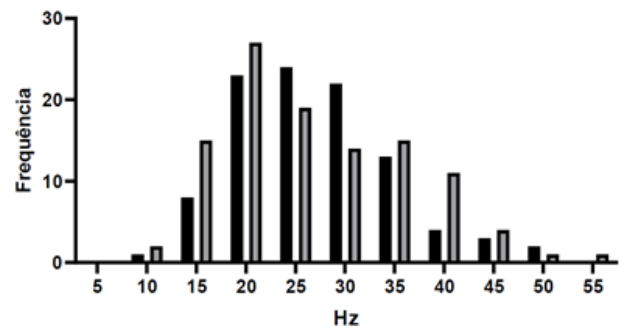


Figura 13. Histograma de desvio em Hz obtido no presente estudo (em preto) sobreposto ao histograma obtido por Lindh [67].

Tanto na análise pericial quanto na literatura o desvio padrão é comumente expresso em semitons. O gráfico da **Figura 14** mostra a distribuição de F0 em semitons.

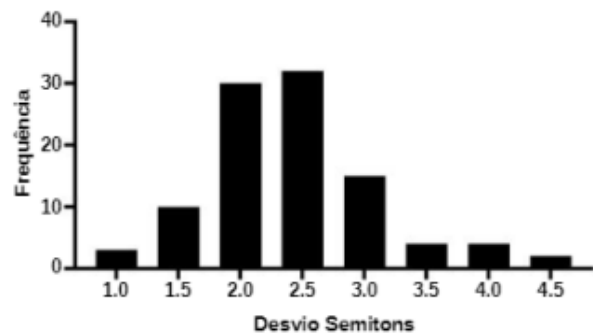


Figura 14. Histograma do desvio padrão em semitons.

a voz do paciente, com o qual ele conversava, aumentava de intensidade de maneira involuntária e inconsciente assim que o ruído começava e retomava seu nível anterior tão logo o som cessava. Em 1911, Lombard publica um artigo batizando essa sua descoberta como “sinal de Lombard”. Subsequentemente, alguns autores passaram a chamar o fenômeno de reflexo de Lombard, devido à natureza involuntária do fenômeno, mas hoje em dia o termo aceito é efeito Lombard, na medida em que o fenômeno não é verdadeiramente um reflexo [75].

Assim como ocorreu com F0 médio, a distribuição de F0 em semitons se aproxima de uma normal (ver [Tabela 2](#)). A comparação entre a distribuição obtida no presente estudo e os dados obtidos por Lindh [67] é mostrada na [Figura 15](#).

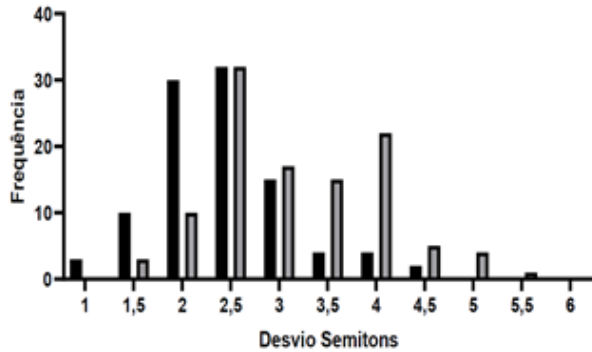


Figura 15. Histograma de desvio em semitom obtido no presente estudo (em preto) sobreposto ao histograma obtido por Lindh [67] (em cinza).

Não há muitos dados disponíveis na literatura para distribuição de desvio-padrão, conforme se mostra na [Tabela 4](#) em que o desvio padrão se expressa em semitons.

Corpus	Desvio-padrão (semitom)
Gold (2014) [35]	1,64
Kinoshita (2005) [64]	2,33
Loakes (2006) [69]	1,87
Skarnitzl & Vaňková (2017) [70]	1,54

Tabela 4. Valores de desvio- padrão expressos em semitom.

Não é difícil observar que o desvio padrão de F0 se correlaciona positivamente com seu respectivo valor médio. Sendo assim, a fim de normalizar essa relação, fazendo com que essa correlação desapareça totalmente [70], utiliza-se o coeficiente de variação - também chamado de fator de modulação de F0 [20], que é dado pela razão entre o desvio padrão e o valor médio e é, geralmente, expresso em porcentagem. Por conta da normalização que promove, o coeficiente de variação se revela bem mais informativo do que o desvio padrão. A [Figura 16](#) mostra o histograma que se refere ao coeficiente de variação da distribuição. Note que a assimetria positiva, indicada na [Tabela 2](#), pode ser visualizada no histograma.

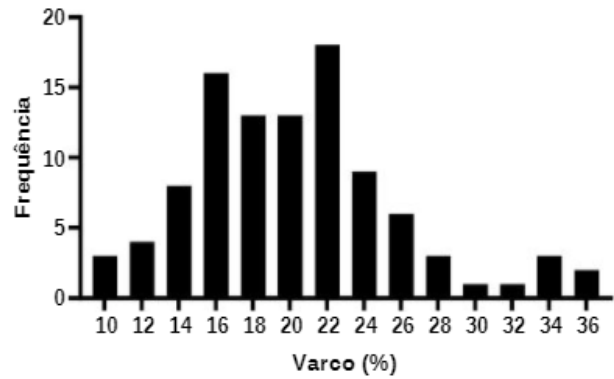


Figura 16. Histograma referente ao coeficiente de variação.

Além da mediana e da média, o artigo de Gold e French [35] aponta que um quarto dos entrevistados se vale de medidas de valor de base de F0 (F0 baseline) como medida de localização. O valor de base de F0 desempenha um importante papel na Teoria da Modulação da Fala [76]. A Teoria da Modulação considera os sinais da fala como o resultado de gestos puramente linguísticos (convencionais) e paralinguísticos os quais permitem modular, de um modo complexo, um sinal portador. Esse sinal, que transmite informações extralinguísticas sobre o falante (sexo, idade, esforço vocal, e assim por diante), é pensado como uma vogal neutra com frequência Fb (valor de base). Ao demodular o sinal de fala, o ouvinte recupera os diferentes tipos de informação. Lindh [77] define F0 baseline como a frequência inicial - ou o ponto inicial da fonação. Uma vez que F0, em nossa base de referência, apresenta distribuição próxima à normal, o valor de base está localizado a 1,43 desvios padrão abaixo da média, correspondente ao 7º percentil da distribuição - 7,64% abaixo de F0, aproximadamente. Do ponto de vista matemático, a distribuição do valor de base pode ser pensada como uma transformação do contorno da distribuição de F0 médio que desloca sua assimetria para a esquerda em função do desvio padrão. A distribuição referente à F0 baseline, mostrada na [Figura 17](#), apresenta assimetria negativa, o que pode ser constatado tanto visualmente quanto consultando a [Tabela 2](#).

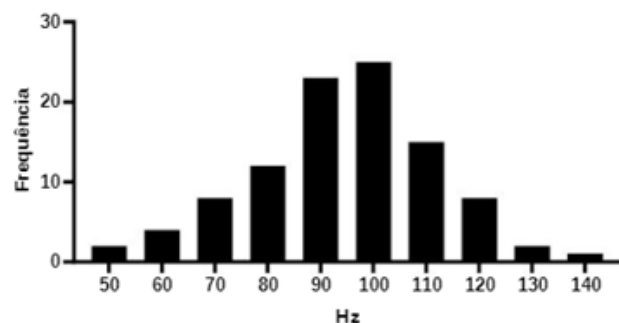


Figura 17. Histograma da frequência de base.

A curva de frequência cumulativa, mostrada na **Figura 18**, mostra que 65% dos valores se encontram entre 80 e 110 Hz.

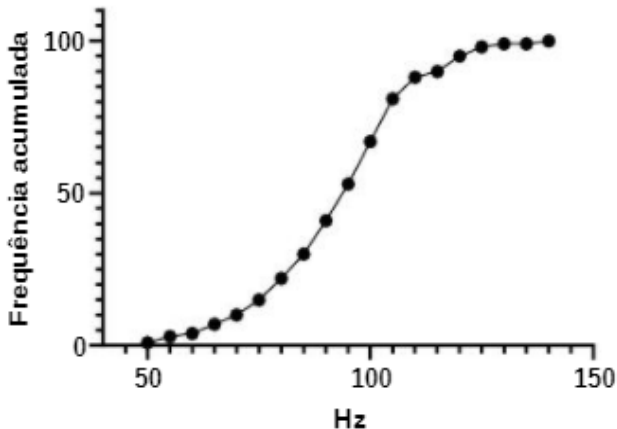


Figura 18. Curva de frequência cumulativa de F0 de base.

Na **Figura 19**, sobreposamos o histograma da **Figura 17** aos dados referentes por Lindh [64] para F0 de base. Os histogramas apresentam semelhanças tanto em relação à extensão de F0 bem como em relação à frequência relativa de suas instâncias.

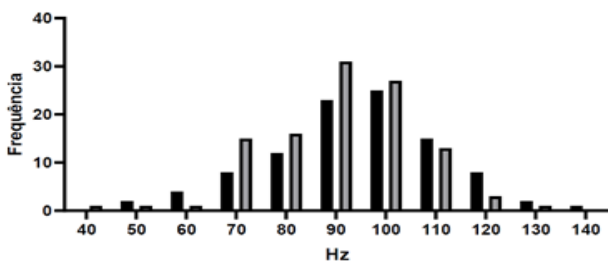


Figura 19. Histograma da **Figura 17** sobreposto ao histograma obtido por Lindh [67] para frequência de base.

Assim como ocorre com a distribuição de desvio – padrão, há poucos dados relativos aos valores de F0 de base disponíveis na literatura. A **Tabela 5** mostra os valores obtidos por Traunmüller & Eriksson [20] em sua meta-análise¹⁷, os obtidos por Lindh [64], para falantes suecos [67] e os valores auferidos por Skarnitzl & Vaňková [70] para falantes tchecos.

4. CONSIDERAÇÕES FINAIS

Embora bases construídas com amostras não contemporâneas, em diversos estilos de fala e através de diversos canais reflitam, mais apropriadamente, condições

forenses realísticas [78,79], a base descrita nesse artigo compõe uma população relevante para as instâncias de F0 de longo termo que podemos esperar para boa parte de falantes paulistanos que nos cedem amostras de seus respectivos materiais de fala em nossa sala de coleta. A base serve também para estudo comparativo com outras bases - construídas, em parte, sob as mesmas premissas-disponíveis na literatura.

Corpus	F0 de Base (Hz)
Traunmüller & Eriksson (1995) [20]	93,4
Lindh (2006) [67]	86,3
Skarnitzl & Vaňková (2017) [70]	95,4

Tabela 5. Valores de frequência de base disponíveis na literatura.

Nesse estudo, obteve-se uma distribuição próxima à normal para a F0 de longo termo, tal qual Hudson *et al.* [36], o que, por conseguinte, possibilitou-nos a obtenção de medidas confiáveis para a distribuição dos valores de base de F0, situada a 1,43 desvio padrão abaixo da média. Note-se que o valor médio da distribuição de F0 de longo termo obtida a partir dos dados desta pesquisa difere significativamente dos valores obtidos para línguas europeias. A comparação entre os histogramas mostra, por exemplo, que uma frequência de 120 Hz, bastante típica entre os falantes paulistanos, possui baixa tipicidade para falantes ingleses [80]. Diferenças em relação aos critérios utilizados para construção da população utilizada nesse estudo, distintas abordagens metodológicas e mesmo a possibilidade - aventada por muitos autores [16,20,35,36] - de que F0 seja dependente da língua, estão entre as plausíveis hipóteses que poderiam explicar tais divergências.

Embora muitos estudos tenham apontado para o fato de que a F0 seja um parâmetro bastante limitado [16,23,35,64,69], isso não a impede de ser utilizada como um fator de exclusão [12,69] muito útil e seguro quando, por exemplo, em um material questionado, envolvendo uma ligação telefônica, as instâncias da F0 estão distribuídas em torno de baixos valores, enquanto, no material padrão, coletado em ambiente controlado, elas se encontram em torno de altos valores. Isso porque fatores emocionais combinados com o próprio efeito telefônico resultam em valores mais altos de F0 no material questionado que no material padrão - e não o contrário [81].

De maneira geral, bases como a descrita nesse estudo permitem a quantificação do peso de F0 na avaliação da

¹⁷ A meta-análise levada a efeito por Traunmüller & Eriksson incluí dados de falantes europeus e chineses (em Mandarim e Wú). Para esse conjunto de valores, os autores obtiveram 119 Hz para o valor médio de F0 e desvio médio de 2,8 semitons.

evidência, segundo o arcabouço bayesiano¹⁸. Nesse arcabouço, o peso associado à F0 pode ser combinado com pesos associados a outros parâmetros, uma vez que as respectivas razões de verossimilhança dos diversos parâmetros analisados, desde que sejam estatisticamente independentes, podem ser combinadas entre si [16.82], gerando razões de verossimilhança de magnitudes bem maiores ou bem menores. Além disso, não é incomum o fato de o perito se deparar com cenários de baixa tipicidade [83], onde medidas associadas à fonte tenham um forte peso na avaliação da evidência, e, nesses casos, bases de referência são muito úteis.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Van Den Berg J. Myoelastic-aerodynamic theory of voice production. *J. Speech Hearing Res.* 1958;1(3):227
- [2] Perello J. La theorie muco-ondulatoire de la phonation. *Ann. D'otolaryngol.* 1962;79:722-725.
- [3] Hirano M. Morphological Structure of the Vocal Cord as a Vibrator and its Variations. *Folia Phoniatr Logop.* 1974;26:89-94.
- [4] Brackett IP. Parameters of voice quality. *Handbook of speech pathology and audiology.* Prentice-Hall: New York, 1971:441-464.
- [5] Kreiman J, Vanlancker-Sidtis D, Gerratt BR. Defining and measuring voice quality. ISCA. Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis; 2003.
- [6] Kreiman J, Van Lancker Sidtis D. *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception* (John Wiley and Sons, Chichester); 2011.
- [7] Colton RH, Casper JK. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment.* Baltimore, Md: Williams & Wilkins; 1996.
- [8] Hollien H. *The Acoustics of Crime —The New Science of Forensic Phonetics.* New York: Plenum; 1990.
- [9] Graddol, D.; Swann, J. Speaking fundamental frequency: Some physical and social correlates. *Language and Speech*, v. 26, n. 4, p. 351-366, 1983.11.
- [10] Nolan F. *The Phonetic Bases of Speaker Recognition.* Cambridge: Cambridge University Press; 1983.
- [11] Atal BS. Automatic speaker recognition based on pitch contour. *Journal of the Acoustical Society of America.* 1972; 52(6):1687-1697.
- [12] Gold E, French P. International practices in forensic speaker comparisons: first survey. *International Journal of Speech, Language & the Law.* 2011;18(2).
- [13] Gold E, French P. International practices in forensic speaker comparisons: second survey. *International Journal of Speech, Language & the Law.* 2019;26(1):1-20.
- [14] Arantes P. Estimativas de longo termo da frequência fundamental: implicações para a fonética forense. *ReVEL.* 2014;12(23).
- [15] Behlau MS. *Uma análise das vogais do Português Brasileiro falado em São Paulo: perceptual, espectrográfica de formantes e computadorizada de frequência fundamental.* Escola Paulista de Medicina; 1984.
- [16] Rose PJ. *Forensic Speaker Identification.* London: Taylor & Francis; 2002.
- [17] Künzel HJ. Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics.* 2000;7:149-179.
- [18] Passetti RR. O efeito do telefone celular no sinal da fala: uma análise fonético-acústica com implicações para a verificação de locutor em português brasileiro. Diss. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem; 2015.
- [19] Titze IR. Physiologic and acoustic differences between male and female voices. *J. Acoust. Soc. Am.* 1989;85:1699-1707.
- [20] Traunmüller H, Eriksson A. The frequency range of the voice fundamental in the speech of male and female adults. Unpublished manuscript; 1995.
- [21] Rose, P. (1991) 'How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency?', *SC 10: 229-47.*
- [22] Kinoshita Y, Ishihara S, Rose P. Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. *International Journal of Speech, Language and the Law.* 2009;16:91-111.
- [23] Braun A. Fundamental frequency – How speaker-specific is it? In: Braun A.; Köster JP. (eds.) *Studies in Forensic Phonetics.* Trier: Wissenschaftlicher Verlag. 1995:9-23.
- [24] Jessen M, Köster O, Gfroerer S. Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law.* 2005;12:174-213.
- [25] Künzel H, Braun A, Eysholdt U. *Einfluß von Alkohol auf Stimme und Sprache.* Heidelberg: Kriminalistik-Verlag; 1992.
- [26] Hollien, H.; Ship, T. Speaking fundamental frequency and chronological age in males. *J. Speech Hear. Res.* 1972; 15:155-159.
- [27] Chevrie-Muller C, Salomon D, Ferrey G. Contribution a l'établissement de quelques constantes physiologiques de la voix parlée de la orma adolescente, adulte et age. *Journal Français d'Oto-Rhino-Laryngologie*

¹⁸ A abordagem metodológica que emerge da combinação da análise fonética com a inferência bayesiana é conhecida como abordagem fonético - acústico estatística] ou abordagem semi-automática [84-86]. Essa conexão foi feita, pela primeira vez, por Philip Rose e está explicada em seu livro [16, 86].

- XVI. 197:433–455.
- [28] Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*. 1972;52:1238-1250.
- [29] Guillemin B, Watson C. Impact of the GSM mobile phone network on the speech signal: some preliminary findings. *International Journal of Speech, Language and the Law*. 2008;15(2):193-218.
- [30] Hirson A, French JP, Howard D. Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics. In: J. Windsor Lewis (ed.), *Studies in General and English Phonetics in Honor of Professor J. D. O'Connor*. London: Routledge; 1995.
- [31] Dieroff, H. G.; Siegert, C. Tonhöhenverschiebung unter Lärmbelastung. *Folia Phoniatica et Logopaedica*, v. 18, n. 4, p. 247-255, 1966.
- [32] Voigt, R.; Jurafsky, D.; Sumner, M.. Between-and Within-Speaker Effects of Bilingualism on F0 Variation. In: *Interspeech*. 2016. p. 1122-1126.
- [33] Boss D. The problem of F0 and real-life speaker identification: A case study. *International Journal of Speech Language and the Law*. 1996;3(1):155-159.
- [34] Rappaport W. Über Messungen der Tonhöhenverteilung in der deutschen Sprache. *Acustica* 8. 1958:220–225.
- [35] Gold E. Calculating likelihood ratios for forensic speaker comparisons using phonetic and linguistic parameters (unpublished doctoral thesis). University of York, York, U.K; 2014.
- [36] Hudson T, et al. F0 statistics for 100 young male speakers of Standard Southern British English. In: *16th Proceedings of the International Congress of Phonetic Sciences, Saarbrücken*. 2007:1809-1812.
- [37] Esling JH, Fraser KE, Harris JG. Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). *Journal of Phonetics*. 2005;33(4):383-410.
- [38] Ladefoged P, Disner SF. *Vowels and consonants*. John Wiley & Sons; 2012.
- [39] Dille SL, Shattuck-Hufnagel M, Ostendorf. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*. 1996;24:423-444.
- [40] Gordon M, Ladefoged P. Phonation types: a cross-linguistic overview. *Journal of Phonetics*. 2011;29:383-406.
- [41] Lima-Gregio AM. Oclusiva glotal e laringalização em sujeitos com fissura palatina: um estudo segundo abordagem dinamicista. Tese (doutorado), Universidade Estadual de Campinas; 2011.
- [42] Podesva RJ. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*. 2007;11(4):478–504.
- [43] Yuasa IP. Creaky voice: a new feminine voice quality for young urban-oriented upwardly mobile American women? *Am Speech*. 2010;85:315–337.
- [44] Abdelli-Beruh NB, Wolk L, Slavin D. Prevalence of vocal fry in young adult male American English speakers. *Journal of Voice*. 2014;28(2):185-190.
- [45] Wolk L, Abdelli-Beruh NB, Slavin D. Habitual use of vocal fry in young adult female speakers. *Journal of Voice*. 2012;26(3):e111-e116.
- [46] Podesva RJ. Gender and the social meaning of non-modal phonation types. *Annual Meeting of the Berkeley Linguistics Society*. 2013;37(1):427–448.
- [47] Gobl C, Chasaide A. Acoustic characteristics of voice quality. *Speech Communication*. 1992;11(4-5):481–490.
- [48] Keating P, Garellek M, Kreiman J. Acoustic properties of different kinds of creaky voice. *ICPhS*; 2015.
- [49] Kuang J. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *Journal of the Acoustical Society of America*. 2017;142:1693–1706.
- [50] Hammarberg B, et al. Perceptual and acoustics correlates of abnormal voice qualities. *Acta Otolaryngol*. 1980; 90(5-6):441-51.
- [51] Silva A. da. Explorando os dois lados do cérebro: a percepção holística da fala reversa (PHFR) sob a perspectiva forense. *Revista Brasileira de Criminalística*. 2019;8(1):62-79.
- [52] Hart DA, et al. Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia. *Eur Arch Otorinolaringol*. 2003;260(4):175-82.
- [53] Leemann A, et al. Speaker individuality in Fujisaki model f0 features: implications for forensic voice comparison. *International Journal of Speech, Language and the Law*. 2014;21(2):343-370.
- [54] Kinoshita, Y., Ishihara, S., and Rose, P. (2009). Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. *International Journal of Speech, Language and the Law*, 16, pp. 91-111.
- [55] Geoffrey Stewart Morrison, Ewald Enzinger, Vincent Hughes, Michael Jessen, Didier Meuwly, Cedric Neumann, S. Planting, William C. Thompson, David van der Vloed, Rolf J.F. Ypma, Cuiling Zhang, A. Anonymous, B. Anonymous, Consensus on validation of forensic voice comparison, *Science & Justice*, Volume 61, Issue 3, 2021, Pages 299-309.
- [56] Coleman RF, Walls HJ. The evaluation of scientific evidence. *Criminal Law Review*. 1974:276–287.
- [57] Morrison, et al. Consensus on validation of forensic voice comparison. *Science & Justice*. 2021.61(3):299-309.

- [58] Smith RL, Charrow RP. Upper and lower bounds for the probability of guilt based on circumstantial evidence. *Journal of the American Statistical Association*.1975;70:555–560.
- [59] Coseriu E. *Teoría del lenguaje y lingüística general*. 3. ed. Madrid: Gredos; 1973.
- [60] Labov W. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press; 1972b.
- [61] Eckert P. Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, Palo Alto. 2012;41:87-100.
- [62] Labov W. *Language in the inner city: Studies in Black English vernacular*. Philadelphia, PA: University of Pennsylvania Press; 1972a.
- [63] Rose PJ. How effective are long term mean and standard deviation as normalization parameters for tonal fundamental frequency?1991;10:229–47.
- [64] Kinoshita Y. Does Lindley’s LR estimation formula work for speech data? Investigation using long-term f0. *International Journal of Speech Language and the Law*.2005;12(2):235–254.
<https://doi.org/10.1558/sll.2005.12.2.235>.
- [65] Laver J. *The phonetic description of voice quality*. Cambridge University Press; 1980.
- [66] Künzel HJ. *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik Verlag; 1987.
- [67] Lindh J. Preliminary descriptive F0-statistics for young male speakers. *Lund Univ. Working Papers*. 2006;52:89-92.
- [68] Künzel HJ. How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*. 1989;46:117-125.
- [69] Loakes D. *A Forensic Phonetic Investigation into the Speech Patterns of Identical and Non-Identical Twins*. (Unpublished Ph.D. thesis), Melbourne University: Melbourne, Australia; 2006.
- [70] Skarnitzl R, Vaňková J. Fundamental frequency statistics for male speakers of Common Czech. *Acta Universitatis Carolinae. Philologica*. 2017; 3:7–17.
- [71] Brixen EB. A Study of f0 as a Function of Vocal Effort. *Audio Engineering Society Conference: 54th International Conference: Audio Forensics*. Audio Engineering Society; 2014.
- [72] Lombard E. Le signe de l’elevation de la voix. *Ann. Mal. de L’Oreille et du Larynx*. 1911:101-119.
- [73] Zollinger SA, Brumm H. The lombard effect. *Current Biology*. 2011;21(16):R614-R615.
- [74] Brumm H, Zollinger SA. The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* 148. 2011;11(13):1173-1198.
- [75] Garnier M, et al. The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement?. 7th international seminar on speech production; 2006.
- [76] Traunmüller H. Conventional, biological and environmental factors in speech communication: A modulation theory. *Phonetica*. 1994;51(1-3):170-183.
- [77] Lindh J. Fundamental frequency and the alternative baseline in forensic speaker identification. *Proceedings, IAFPA The College of St. Mark & St. John: Plymouth, UK; 2007*.
- [78] Morrison, Geoffrey Stewart. Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice*, v. 51, n. 3, p. 91-98, 2011.
- [79] Kinoshita, Yuko et al. Background population: how does it affect LR-based forensic voice comparison?. *The International Journal of Speech, Language and the Law*, v. 21, n. 2, p. 191-224, 2014.
- [80] Jessen, M. *Forensic speaker identification by experts*. *The Routledge Handbook of Forensic Linguistics*, Abingdon and New York: Routledge, 2010.
- [81] Hirson, A.; French, J.P.; Howard, D. Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics. In: J. Windsor Lewis (ed.), *Studies in General and English Phonetics in Honor of Professor J. D. O’Connor*. London: Routledge, 1995
- [82] Silva, Gerson Albuquerque. Proposta de Construção de um Banco de Dados de Amostras de Fala para Uso Forense em um Arcabouço Bayesiano. *Revista Brasileira de Criminalística*, v. 5, n. 1, p. 35-45, 2016.
- [83] Jensen, Michael. Two issues on the combination between automatic and auditory – acoustic methods in forensic voice comparison.29th Conference of The International Association for Forensic Phonetics and Acoustics, Marburg, 2021
- [84] Morrison, Geoffrey Stewart, Farhan Hyder Sahito, Gaëlle Jardine, Djordje Djokic, Sophie Clavet, Sabine Berghs & Caroline Goemans Dorny. 2016. INTERPOL survey of the use of speaker identification by law enforcement agencies. *Forensic Science International* 263. 92–100
- [85] Drygajlo, Andrzej, et al. *Methodological guidelines for best practice in forensic semiautomatic and automatic speaker recognition*. Verlag für Polizeiwissenschaft, 2015.
- [86] Jessen M. *Forensic Voice Comparison In: Visconti J.(ed.) Handbook of Communication in the Legal Sphere*. Mouton de Gruyter, Berlin/Boston/New York; 2018.