ARANTES, Pablo. Estimativas de longo termo da frequência fundamental: implicações para a fonética forense. *ReVEL*, vol. 12, n. 23, 2014. [www.revel.inf.br].

ESTIMATIVAS DE LONGO TERMO DA FREQUÊNCIA FUNDAMENTAL: IMPLICAÇÕES PARA A FONÉTICA FORENSE

Pablo Arantes[1]

pabloarantes@ufscar.br

RESUMO: Investigamos a média, a mediana e o valor de base cumulativos para estimar quanto tempo leva para variabilidade atingir estabilidade. Usamos a técnica estatística *changepoint analysis* para localizar os pontos de estabilização. Em um experimento, pontos de estabilização foram estimados em gravações de 26 línguas. O ponto de estabilização médio ocorreu em 5s para o valor de base em 10s para a média e a mediana. A variância após o ponto de estabilização foi reduzida por um fator de 40 vezes no caso da média e mediana e 120 vezes no caso do valor de base. No segundo experimento, quatro falantes leram dois textos diferentes. Os pontos de estabilização não coincidiram nos dois textos. O deslocamento mediano foi de 2,5s no caso do valor de base, 3,4s no caso na mediana e 9,5s no caso da média. A diferença mediana no valor estimado de F_0 no ponto de estabilização nos dois textos é de 2%. No terceiro experimento, foram analisadas leituras do mesmo texto feitas por um falante do sexo masculino e um do sexo feminino, gravadas em um intervalo de 10 anos. O deslocamento mediano do ponto de estabilização é 0,4s no caso da falante feminina e 5,6s no caso do falante masculino. As diferenças nas estimativas de F_0 estão em torno de 5% para a falante feminina e 12% para o falante masculino. Os resultados sugerem que as estimativas de F_0 atingem a estabilidade mais cedo do que sugere a literatura.
PALAVRAS-CHAVE: Frequência fundamental; Fonética forense; Estatística descritiva.

INTRODUÇÃO

Neste trabalho discutiremos algumas propriedades temporais de diferentes medidas estatísticas de longo termo da frequência fundamental da voz (F_0) e algumas implicações dos resultados obtidos para a prática da fonética forense.

Duas questões importantes que devem ser respondidas a respeito da estimação do valor típico de F_0 de uma amostra são: (a) que medida estatística é a mais apropriada e (b) qual deve ser a duração mínima da amostra de fala a ser usada para obter uma estimativa representativa de um determinado falante.

[1] Universidade Federal do São Carlos (UFSCar). Parte dos resultados apresentados nas seções 3.1 e 3.2 foram reportados em Arantes e Eriksson (2014). As seções 1 e 3.3 são inteiramente originais.

No que diz respeito à escolha do estimador estatístico mais adequado, a média aritmética é a medida mais citada na literatura e a mais usada na prática forense, embora a mediana algumas vezes também seja mencionada (ROSE, 1991; TRAUNMÜLLER; ERIKSSON, s.d.; ERIKSSON, 2011). Uma possível inconveniência relacionada ao uso da média é o fato de que amostras de F_0 tipicamente têm assimetria positiva, isto é, histogramas de amostras de F_0 tendem a apresentar uma cauda direita mais pesada do que a esperada se a amostra se conformasse a uma distribuição normal (JASSEM, 1975). É bastante sabido que o cálculo da média é sensível à presença de assimetrias, e o valor produzido nestes casos pode ser enviesado. Um dos objetivos do presente trabalho é testar, além da média, outros dois estimadores de localização: a mediana e o valor de base. A seção 1 explicará em mais detalhes o que é o valor de base e como ele é calculado.

A outra questão importante é a duração da amostra. Os resultados que discutiremos aqui e outros reportados na literatura (ROSE, 1991, por exemplo) mostram que, se a média ou outro estimador de valor central for calculada de forma cumulativa, a variabilidade dos valores obtidos inicialmente é alta e diminui conforme aumenta o número de valores que compõem a amostra. Uma pergunta com a qual lidaremos é "que duração é necessária para que a variabilidade diminua?". Em uma revisão da literatura a respeito do assunto, Eriksson (2011) cita cinco diferentes estimativas, que variam entre 14 segundos e dois minutos, isto é, uma diferença de quase uma ordem de magnitude. Outro aspecto que discutiremos é "que método usar para decidir em que momento o grau de variabilidade atingiu alguma estabilidade?". Pouca informação sobre os métodos usados para chegar a estas estimativas está disponível na literatura. Rose (1991), por exemplo, usa o critério da inspeção visual de gráficos que mostram a evolução temporal da média cumulativa ao longo de uma amostra de fala, procedimento que pode ser afetado pela subjetividade do observador. Uma das contribuições deste trabalho é o uso de uma técnica estatística para a detecção destes pontos de estabilização.[2]

As duas questões que serão investigadas têm relevância para a fonética forense. Saber que estimador é capaz de produzir a estimativa menos variável no menor tempo possível é importante porque, em casos reais, as amostras disponíveis podem ter durações curtas (ERIKSSON, 2005). Em cenários assim, é fundamental

[2] Ver seção 2.6 para informações adicionais sobre a técnica utilizada.

que peritos disponham de valores de referência nos quais possam se basear para garantir maior objetividade e replicabilidade em suas análises.

1 ESTIMADORES ESTATÍSTICOS DA FREQUÊNCIA FUNDAMENTAL

Na estatística descritiva, são chamados de medidas de tendência central ou localização os estimadores que caracterizam a variabilidade de uma amostra de dados, sintetizando-a em um número, que representa o valor típico ou mais representativo daquela amostra. Diferentes estimadores desse tipo existem com méritos e limitações próprias (KEENEY; KEEPING, 1962). A média e a mediana são estimadores de localização genéricos no sentido de que podem ser aplicados a amostras de qualquer natureza, desde que a variável observada possa ser medida em uma escala não nominal. Como dissemos na introdução, a média é a medida mais usada para estimar o valor típico de F_0 de uma amostra de fala, apesar dos potenciais problemas causados pela assimetria típica das amostras de F_0. A mediana, cujo cálculo leva em conta não o valor das amostras mas seu *rank* ou posto dentro da amostra, é uma alternativa comum à média por ser robusta à presença de valores extremos.

O valor de base (*base value* ou *base line*, no original) é um estimador estatístico de localização proposto por Traunmüller e Eriksson (s.d.) especialmente para amostras de F_0 e leva em conta as especificidades típicas dessa amostra. Uma das especificidades consideradas é que a variação de F_0 geralmente não é simétrica. Quando os falantes fazem excursões entoacionais, o movimento, na grande maioria das vezes, é ascendente, fato que se revela nos histogramas de distribuições de F_0 como uma assimetria positiva. O nível de F_0 típico que os falantes parecem manter é aquele logo acima do mínimo necessário para manter a fonação. Movimentos abaixo desse nível seriam, assim, menos comuns porque poderiam resultar em vozeamento não modal (ERIKSSON, 2011). Mesmo em situações que fazem a variabilidade da F_0 aumentar, como, por exemplo, falar com maior envolvimento emocional, essa tendência à assimetria se mantém e colabora para aumentar ainda mais o grau de assimetria da distribuição de longo termo de F_0.

Traunmüller e Eriksson (s.d.) propuseram o valor de base como uma alternativa a medidas como a média ou a mediana. O valor de base (F_b) de uma determinada amostra de F_0 é obtido pela aplicação da fórmula $F_b = F_{média} - k\sigma$, em

que $F_{média}$ e σ são, respectivamente, o valor da média aritmética e o desvio-padrão de F_0 da amostra, e k é uma constante determinada empiricamente. No trabalho citado, o valor sugerido para a constante é 1,5. Posteriormente, Lindh e Eriksson (2007) revisaram o valor de k para 1,43 e sugeriram uma formulação alternativa para o cálculo do valor de base: assumindo uma distribuição normal para os dados de F_0, o ponto 1,43·σ abaixo da média corresponde aproximadamente ao 7º percentil da distribuição. Os autores mostraram que essa formulação é mais robusta do que a formulação original e, por essa razão, foi ela que usamos para o cálculo do valor de base neste estudo.

1.1 EFEITO DA LÍNGUA

O uso da F_0 para a expressão de contrastes relevantes do ponto de vista linguístico e paralinguístico é bastante variado entre as línguas naturais. Para ilustrar a importância central da língua no uso que os falantes fazem da F_0, podemos citar a observação feita por Traunmüller e Eriksson (s.d.) de que mesmo a diferença típica entre a F_0 média de homens e mulheres, causada por diferenças fisiológicas entre os sexos, pode estar sob controle de convenções linguísticas. Em sua meta-análise dos dados sobre a variabilidade da F_0 disponíveis na literatura, os autores citam dois exemplos de línguas – o dialeto chinês Wú e o dialeto sueco da província de Småland – em que a diferença entre a F_0 média de homens e mulheres é muito menor do que a normalmente observada em outras comunidades linguísticas. Esse comentário serve para mostrar que, mesmo quando estamos interessados naquilo que é próprio da voz de um indivíduo particular (e o valor típico de F_0 usado por ele pode ser uma dessas características individualizantes), é importante levar em conta a língua que ele fala.

Neste estudo, estamos interessados em observar como varia em diferentes línguas o comportamento temporal de três diferentes estimadores de localização de F_0 quando computados de forma cumulativa ao longo de uma amostra de fala. Há línguas em que a variabilidade no valor dos estimadores leva mais tempo para atingir um patamar que poderia ser considerado estável? Os estimadores mostram comportamento parecido entre si ou algum deles tende a atingir a estabilidade mais cedo?

Investigamos o efeito da língua sobre a variabilidade nas medidas de tendência central aqui através da análise de um conjunto de gravações de falantes de 26 línguas

lendo a passagem "O vento norte e o sol". As gravações estão disponíveis no site da International Phonetic Association (IPA)[3]. Dezesseis falantes são do sexo masculino. A amostra inclui línguas de oito diferentes famílias linguísticas: afro-asiática (amárico, árabe, hebraico), sino-tibetana (cantonês), indo-europeia (inglês americano, búlgaro, catalão, croata, tcheco, holandês, francês, galego, alemão, hindi, gaélico, farsi, português europeu, sindi, esloveno e sueco), urálica (húngaro), nigero-congolesa (igbo), altaica (japonês, coreano), tai-kadai (tailandês), túrquica (turco). Destas, o cantonês, o igbo e o tailandês são tonais, e o sueco (GÅRDING, 1998) e o japonês (ABE, 1998) têm acentos tonais (*pitch accents*) em palavras com mais de duas sílabas.

1.2 EFEITO DO TEXTO

É um fato reconhecido pela literatura que o conteúdo segmental da fala influencia o contorno da F_0 dos enunciados (LEHISTE, 1970). Do ponto de vista da fonética forense, essa relação entre os segmentos e a F_0 é relevante na medida em que normalmente o conteúdo linguístico da amostra questionada, aquela cuja identidade do falante é desconhecida, não é idêntico ao das amostras de referência, aquelas cuja identidade dos falantes é conhecida. É importante, então, saber, entre outras coisas, o quanto a estimativa da F_0 típica de um falante é dependente do conteúdo segmental de uma determinada amostra e quanto ela varia quando comparada a estimativas feitas a partir de outras amostras com conteúdo segmental que não seja idêntico. Neste estudo, observaremos se textos diferentes lidos pelos mesmos falantes afetam: a) o valor de F_0 estimado por diferentes medidas estatísticas; e b) a quantidade de fala necessária para que o valor dos estimadores atinja estabilidade.

Para testar o efeito o do conteúdo específico de um texto sobre a variabilidade das medidas de F_0 foram usadas gravações de quatro diferentes falantes de português brasileiro lendo dois textos diferentes. Os textos são a passagem "O vento norte e o sol" adaptada e traduzida para o português brasileiro (BARBOSA; ALBANO 2004) e uma passagem de "A menina do narizinho arrebitado", do escritor Monteiro Lobato. O primeiro texto será referido como texto 1 e o segundo como texto 2. O texto 2 é foneticamente balanceado (todos os fonemas do português brasileiro ocorrem no texto), ao passo que na tradução do texto 1 esse critério não foi observado. Cada um

[3] http://web.uvic.ca/ling/resources/ipa/handbook_downloads.htm.

dos dois textos foi lido por um falante masculino e um feminino das variedades linguísticas dos estados de São Paulo e Minas Gerais

1.3 GRAVAÇÕES NÃO CONTEMPORÂNEAS

Eriksson (2005) menciona que, na prática forense, são comuns as situações em que gravações obtidas com algum intervalo são usadas em processos de comparação de vozes ou identificação auditiva de falantes. Intervalos de um ano ou pouco mais são relativamente comuns segundo o autor. O fato de que a voz humana muda com a passagem do tempo pode prejudicar a comparação de gravações feitas em momentos diferentes. A F_0 é uma das características que sofre alteração ao longo da vida de um falante e, como a estimativa da F_0 típica de um falante é um parâmetro bastante usado na comparação de voz com finalidade forense, é relevante que se estude o efeito da passagem de tempo em diferentes estimadores da F_0 de longo termo.

O efeito da passagem de tempo sobre a variabilidade das medidas de F_0 foi testado neste estudo por meio da comparação de gravações da leitura da passagem "O vento sul e o sol" (BARBOSA; ALBANO, 2004) feitas com aproximadamente uma década de diferença pelo par de falantes de Minas Gerais mencionado na seção anterior. A primeira gravação foi realizada em 2003, e a segunda, em 2013. Na ocasião da primeira gravação, os falantes tinham em torno de 20-25 anos. Nenhum dos dois fumou regularmente durante o período entre as duas gravações.

O interesse é saber qual é a influência do intervalo de tempo entre duas gravações de um falante lendo o mesmo texto sobre o tempo necessário para os três estimadores de localização de F_0 atingirem a estabilidade. Mantendo constantes o falante e o conteúdo fonético/linguístico do texto lido, as eventuais diferenças observadas no tempo de estabilização poderiam ser atribuídas às possíveis mudanças na voz, em especial as que afetam a produção da F_0. Do ponto de vista da fonética forense, esta análise pode ajudar a estabelecer qual dos três estimadores estudados aqui é mais resistente à influência dos fatores que causam mudanças na F_0 de um indivíduo com o passar do tempo.

2. MATERIAIS E MÉTODOS

O contorno de F_0 de cada gravação analisada foi extraído com a ajuda de um *script* do programa de análise acústica Praat criado pelo autor[4] que implementa uma heurística sugerida por Hirst (2011) que procura minimizar erros de extração por meio da otimização da escolha dos parâmetros *floor* e *ceiling* usados pelo algoritmo de extração de F_0. Os erros remanescentes foram corrigidos manualmente. O processamento adicional dos contornos para obter as medidas cumulativas de localização foi feito por outro *script* do Praat escrito para essa finalidade.

2.1 MEDIDAS DE LOCALIZAÇÃO

As seguintes medidas estatísticas de localização foram investigadas:

- média aritmética: soma dos valores da amostra de F_0 dividida pelo tamanho da amostra;
- mediana: 50º percentil da amostra de valores de F_0; e
- valor de base: 7º percentil da amostra de valores de F_0.

Todas as medidas foram calculadas de forma cumulativa do instante em que o vozeamento se inicia até o último ponto de F_0 do contorno em incrementos consecutivos de 200 ms. O número de valores de F_0 acrescentados a cada passo de 200 ms depende do valor do parâmetro *floor* passado para o algoritmo de extração do programa Praat. No conjunto de dados do IPA, o valor médio desse parâmetro foi 70 Hz para os falantes masculinos e 120 Hz para os femininos, o que corresponde, respectivamente, a 20 e 32 valores de F_0 a cada 200 ms.

No conjunto de dados do IPA, a duração mediana das gravações é de 38s, com valores entre 25s (galego) e 66s (tailandês). As gravações usadas para testar o efeito do texto têm duração mediana de 32s (texto 1) e 41s (texto 2).

Os valores da média e da mediana de um contorno de F_0 em geral serão parecidos, mas o valor de base, por definição, será menor do que o das outras medidas. Como o que interessa aqui não é o valor absoluto das medidas, mas sua variabilidade em função do aumento da amostra, adotou-se um procedimento de

[4] Disponível em http://code.google.com/p/praat-tools/.

normalização que ajustou a escala da série temporal das três medidas de localização ao intervalo [0, 1]. Para fazer a transformação, foi usada a fórmula $(f_i - f_{min}) / (f_{max} - f_{min})$, em que f_i é o i-ésimo valor de F_0 em um contorno, e f_{min} e f_{max} são os valores mínimo e máximo. Os valores normalizados são usados apenas para facilitar a análise visual dos contornos. A análise estatística, descrita em maiores detalhes na seção seguinte, foi feita apenas nas curvas não transformadas.

2.2 ANÁLISE ESTATÍSTICA

O interesse principal do trabalho é determinar quanto tempo leva para que a variabilidade da série temporal definida pelo cálculo cumulativo das medidas de localização de F_0 seja reduzida a um valor que pudesse ser considerado estável. Na literatura sobre o tema, aquilo que estamos chamando aqui de ponto de estabilização é determinado por meio da inspeção visual dos traçados das séries temporais. Embora o recurso à inspeção visual tenha sua utilidade, seria desejável desenvolver um método menos suscetível à subjetividade inerente a uma análise puramente visual.

Com o objetivo de atingir um maior patamar de objetividade na determinação dos pontos de estabilização, aplicou-se a técnica estatística chamada *changepoint analysis* (KILLICK; ECKELEY, 2013), implementada na forma de uma biblioteca de funções do ambiente de computação estatística R (R CORE TEAM, 2014). Na modalidade usada para a análise apresentada aqui, a técnica produz uma estimativa do ponto no tempo em que a variância[5] da série temporal analisada sofre mudança e testa a hipótese de que os valores da variância antes e depois do ponto são estatisticamente diferentes. Instruiu-se o algoritmo a buscar apenas um ponto de mudança nas amostras de média, mediana e valor de base cumulativas e a não assumir que os valores das séries analisadas sigam uma distribuição normal.

[5] Podem-se também identificar pontos de transição causados por mudanças na média e na média e variância conjuntamente.

3. RESULTADOS

3.1 EFEITO DA LÍNGUA

A Figura 1 mostra a evolução temporal das três medidas de localização cumulativas normalizadas para as 26 línguas da amostra analisada. A presença de flutuações de grande amplitude é uma tendência observada em basicamente todas as línguas, especialmente durante os primeiros segundos de cada série. À medida que o valor dos estimadores é computado em trechos mais longos, a amplitude das flutuações diminui progressivamente, embora seja possível observar diferenças entre as línguas. Em alguns casos, a variabilidade decresce de maneira bastante rápida, como no caso do árabe, do alemão, do esloveno, do galego, do húngaro e do sueco, enquanto em outros a redução parece se dar de forma mais gradual, como no catalão, no coreano, no farsi, no francês e no gaélico. Com a possível exceção do turco, o valor cumulativo dos três estimadores tende, visualmente, a atingir um patamar estável em algum momento, em geral, no primeiro quarto da duração da gravação.

A Tabela 1 lista a localização temporal dos pontos de estabilização encontrados através da aplicação da técnica estatística descrita na seção 2.2 nas séries temporais dos três estimadores. Em todos os casos, a aplicação da técnica permitiu encontrar um ponto de estabilização. A variância no trecho após o ponto de estabilização é menor do que a observada no trecho anterior nas amostras de todas as línguas. A Tabela 1 lista entre parênteses os fatores de redução da variância, isto é, a razão entre o valor da variância antes e depois do ponto de estabilização. A Figura 2 mostra através de *boxplots* a variabilidade do ponto de estabilização da média, da mediana e do valor de base das 26 línguas analisadas.

Um dos achados principais é que, de forma geral, os pontos de estabilização estimados pela técnica *changepoint analysis*, nesta amostra de línguas, estão na parte inferior da gama de valores sugeridos pela literatura e mencionados na introdução. Outro achado relevante é que o valor de base tende a estabilizar mais cedo (em torno de 5s) do que a média e a mediana (em torno de 10s). Os pontos de estabilização do valor de base são também menos variáveis (desvio mediano absoluto de 2,2s) do que a média e a mediana (desvios de 6,2s e 7,6s, respectivamente). Os valores altos dos fatores de redução da variância sugerem que os pontos identificados pela análise estatística podem ser entendidos como pontos de estabilização. Nesse

quesito, o valor de base também tem vantagens em relação aos outros dois estimadores: o fator de redução médio do valor de base é 120, enquanto o fator da média e da mediana são ligeiramente inferiores a 50.

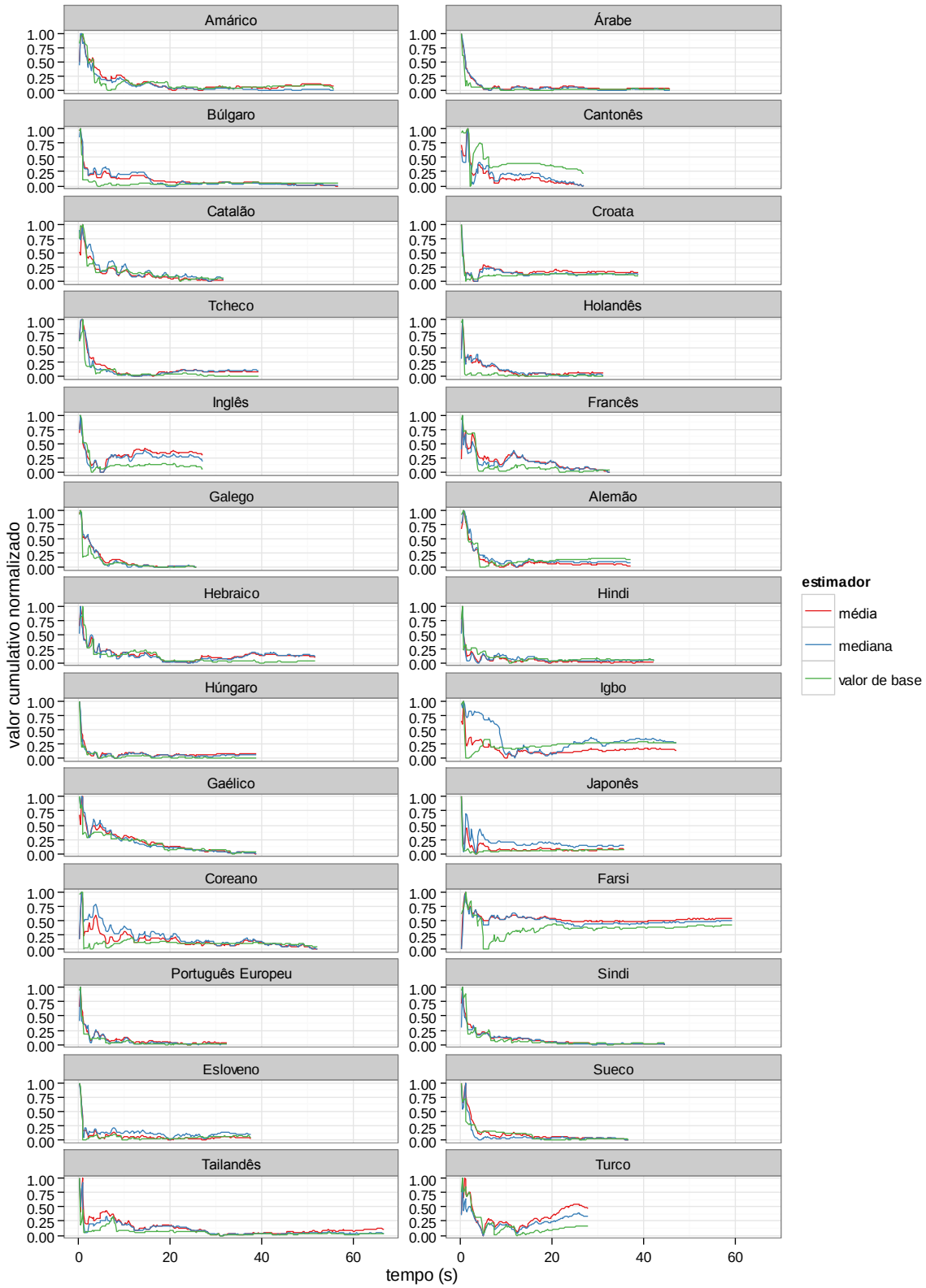

Figura 1: Valor normalizado da média, da mediana e do valor de base computados cumulativamente nas gravações da passagem "The North Wind and the Sun" em 26 línguas. O eixo horizontal indica a duração das amostras.

Língua	Média	Mediana	Valor de base
Amárico	11 (44)	16,2 (137)	4,8 (69)
Árabe	4,2 (242)	4,6 (226)	4,6 (619)
Búlgaro	16,2 (48)	15,6 (79)	3,8 (440)
Cantonês	6.2 (26)	7 (8)	6 (48)
Catalão	11 (40)	10,6 (33)	11,2 (31)
Croata	11 (40)	10,6 (33)	11,2 (31)
Tcheco	6,8 (86)	4,8 (77)	8 (204)
Holandês	8,2 (50)	10,2 (36)	4,4 (778)
Inglês americano	11,6 (48)	1.6 (5)	2,4 (63)
Francês	15 (7)	16 (8)	3,4 (22)
Galego	5,2 (28)	4,8 (67)	5,2 (100)
Alemão	5,2 (141)	5,8 (135)	3,8 (26)
Hindi	10 (112)	16,4 (322)	10 (84)
Húngaro	2,4 (194)	3,8 (171)	4 (217)
Hebraico	15,2 (13)	7,8 (11)	17,6 (170)
Igbo	7,8 (19)	8,8 (2)	21 (222)
Gaélico	14,4 (7)	12,6 (17)	15 (10)
Japonês	6 (183)	11.6 (58)	0,4 (1772)
Coreano	21,4 (18)	21,6 (13)	1 (3)
Farsi	20,6 (27)	20,6 (21)	4,6 (3)
Português europeu	11(116)	6,4 (81)	5,8 (219)
Sindi	15 (66)	15.4 (44)	6.4 (172)
Esloveno	9 (141)	14,8 (21)	0,8 (232)
Sueco	6,2 (108)	2,8 (504)	15,4 (271)
Tailandês	22,2 (23)	24,2 (52)	27,6 (137)
Turco	20 (11)	3,2 (3)	4,6 (16)

Tabela 1: Pontos de estabilização (em segundos) da média, da mediana e do valor de base para as 26 línguas da amostra da International Phonetic Association. O fator de redução da variância após o ponto de estabilização é informado entre parênteses.

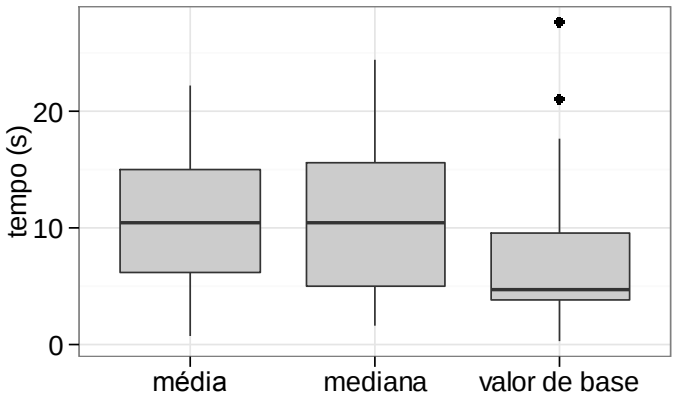

Figura 2: Boxplots mostrando a distribuição dos pontos de estabilização da média, da mediana e do valor de base computados cumulativamente na amostra de 26 línguas da International Phonetic Association.

3.2 EFEITO DO TEXTO

A Figura 3 mostra os valores das três medidas de localização cumulativas normalizadas para os dois textos lidos pelos falantes femininos e masculinos de São Paulo e Minas Gerais. A Tabela 2 mostra os pontos de estabilização encontrados pela análise estatística e os fatores de redução da variância.

Supondo que o conteúdo fonético/linguístico do texto não exercesse efeito relevante sobre os estimadores de localização de F_0, o ponto de estabilização e o valor do estimador neste ponto deveriam idealmente ser idênticos para o mesmo falante lendo dois textos diferentes. Não foi este o caso para os falantes da amostra investigada. As diferenças absolutas médias entre os pontos de estabilização dos textos 1 e 2 é de 4,9 segundos, com um mínimo de 0,4 e máximo de 21,1 segundos. 75% das diferenças são menores do que 6 segundos. Levando em consideração o fato de que o desvio-padrão dos pontos de estabilização na amostra de língua da IPA é 6 segundos, a diferença típica entre os pontos de estabilização observada entre os textos 1 e 2 é ligeiramente inferior àquelas que seriam esperadas quando diferentes línguas são comparadas. A comparação das diferenças entre os valores brutos (em Hertz) da média, da mediana e do valor de base cumulativos dos textos 1 e 2 no ponto de estabilização mostra que, em média, a diferença, considerando os dados dos quatro falantes, é de 2%, variando entre um mínimo de 0% e um máximo de 9%. 90% das diferenças são menores do que 4%, isto é, menores do que um semitom.

Nenhum dos textos parece ter pontos de estabilização sistematicamente menores ou fatores de redução de variância maiores do que o outro. A única exceção parece ser que os pontos de estabilização da mediana no texto 2 tendem a ser mais precoces do que os do texto 1 para todos os falantes. Não está claro se esse comportamento pode ser atribuído ao fato de que o texto 2 é foneticamente balanceado e porque a mediana é o único estimador afetado. Os fatores de redução da variância após o ponto de estabilização dos três estimadores são parecidos – o valor mediano para a média e a mediana é em torno de 20 e o do valor de base é 36.

Falante	Texto	Média	Mediana	Valor de base
SP-Fem.	1	9,8 (18)	9,6 (21)	9,6 (6)
	2	15,2 (28)	5,2 (13)	5,6 (83)
SP-Masc.	1	10,2 (92)	10,2 (58)	6,2 (33)
	2	4,8 (22)	4,8 (42)	5,2 (7)
MG-Fem.	1	7,2 (11)	7,2 (4)	9,4 (9)
	2	28,2 (8)	5,6 (6)	9,8 (39)
MG-Masc.	1	10,4 (33)	10,8 (33)	5,2 (41)
	2	24 (20)	8 (14)	13,8 (39)

Tabela 2: Pontos de estabilização (em segundos) da média, da mediana e do valor de base para os dois textos lidos por falantes masculinos e femininos de São Paulo e Minas Gerais. O fator de redução da variância após o ponto de estabilização é informado entre parênteses.

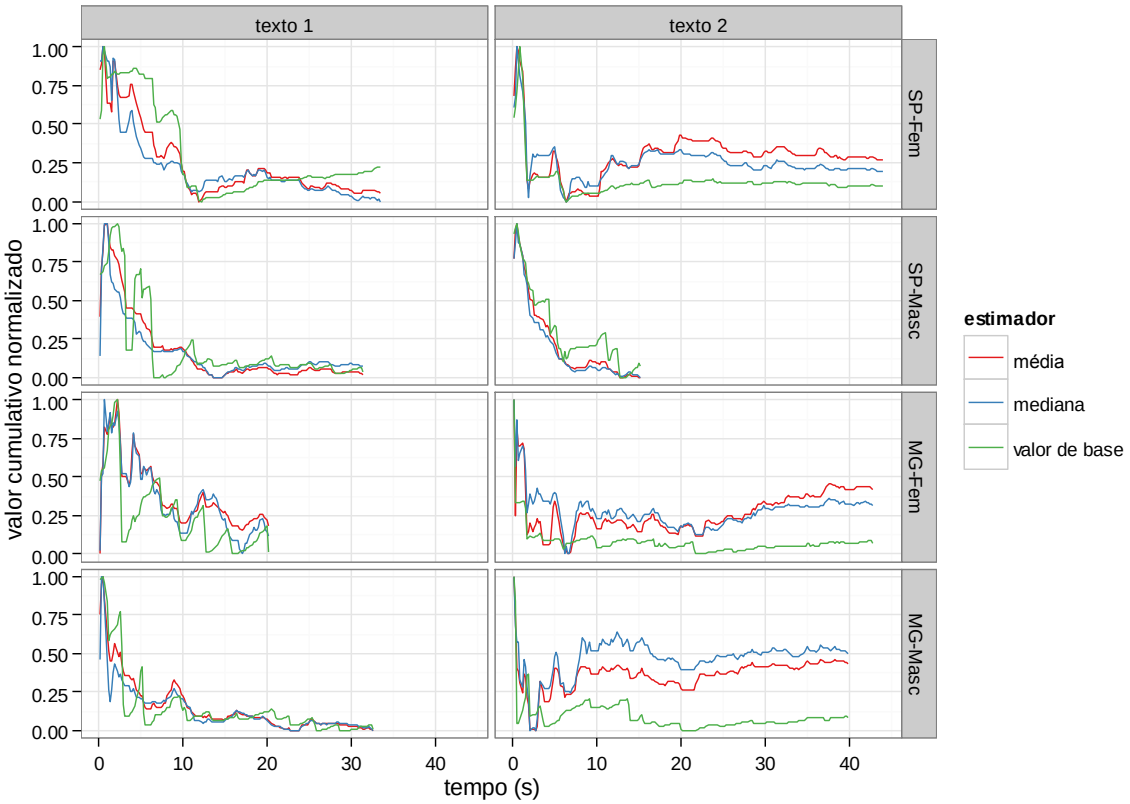

Figura 3: Valor normalizado da média, da mediana e do valor de base computados cumulativamente no texto "O vento sul e o sol" (texto 1) e no texto "A menina do narizinho arrebitado" (texto 2), lidos por quatro falantes. O eixo horizontal indica a duração das amostras.

3.3 GRAVAÇÕES NÃO CONTEMPORÂNEAS

A Figura 4 mostra os valores das três medidas de localização cumulativas normalizadas para duas gravações do mesmo texto lido por um falante masculino e um feminino de Minas Gerais com um intervalo de aproximadamente 10 anos (2003 e 2013). A Tabela 3 lista os pontos de estabilização e os fatores de redução da variância.

O exame visual da Figura 4 sugere uma forte semelhança entre o traçado das séries temporais das gravações, indicando que o intervalo de tempo não teve um efeito forte no comportamento dos três estimadores. Especialmente no caso da falante feminina, é possível observar uma forte diminuição na variabilidade em torno dos 10 segundos tanto na primeira quanto na segunda gravação.

Os resultados da análise estatística mostram que o deslocamento mediano absoluto do ponto de estabilização é de 5,2 segundos para o falante masculino e de 0,4 segundos para a falante feminina. A comparação do valor de F_0 estimado pelas três medidas estatísticas no ponto de estabilização é sistematicamente menor na

segunda gravação para os dois falantes – diferença média de 2 semitons (relativos ao menor dos valores do par) para o falante masculino e de 0,85 semitons para a falante feminina. Esses resultados também sugerem que o efeito da passagem de tempo não afeta de maneira drástica a localização dos pontos de estabilização nem a estimativa do valor típico de F_0. A comparação do comportamento dos três estimadores não sugere nenhuma diferença importante.

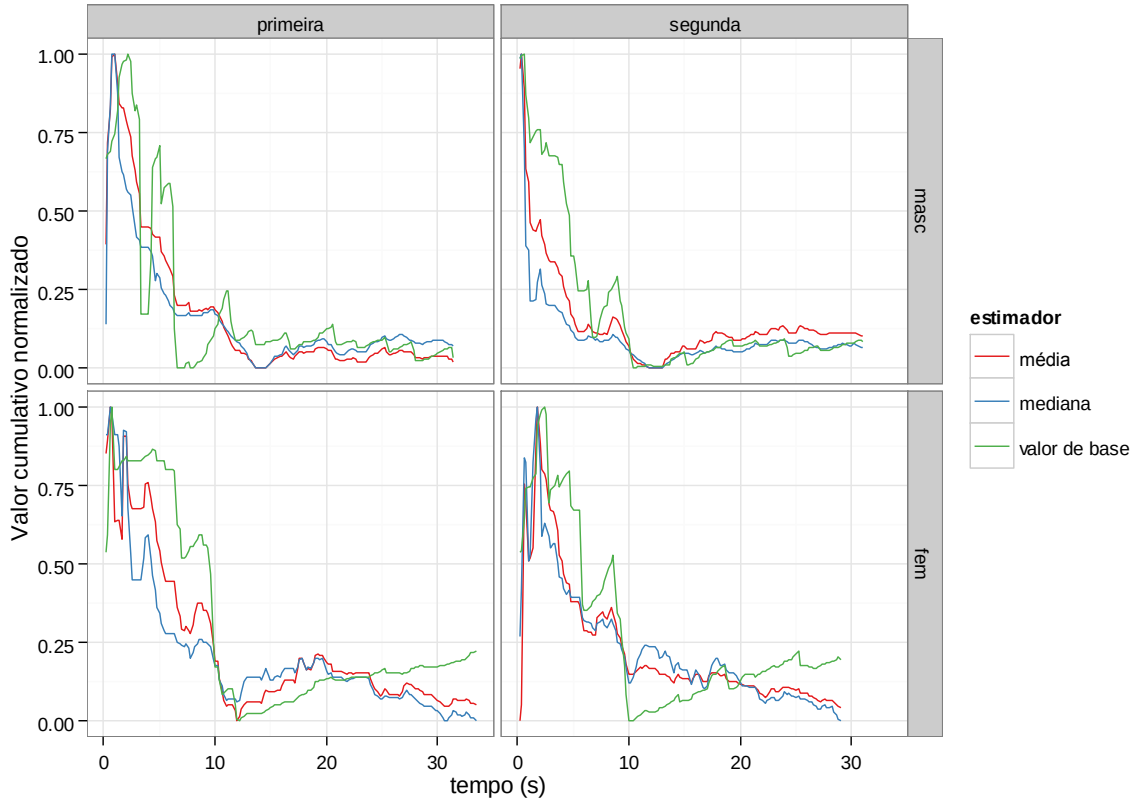

Figura 4: Valor normalizado da média, da mediana e do valor de base computados cumulativamente no texto "O vento sul e o sol", lido por dois falantes em gravações feitas em um intervalo de dez anos. O eixo horizontal indica a duração das amostras.

Falante	Gravação	Média	Mediana	Valor de base
Masc.	1	10,2 (92)	10,2 (58)	6,2 (33)
	2	5 (34)	5 (85)	9,4 (90)
Fem.	1	9,8 (18)	9,6 (21)	9,6 (6)
	2	9,4 (39)	9,2 (9)	9,2 (10)

Tabela 3: Pontos de estabilização (em segundos) da média, da mediana e do valor de base para as duas gravações do mesmo texto lidos por um falante masculino e um feminino em 2003 (gravação 1) e em 2013 (gravação 2). O fator de redução da variância após o ponto de estabilização é informado entre parênteses.

CONCLUSÃO

Neste estudo, investigamos o comportamento de três medidas estatísticas de localização que podem servir como indicadores do valor de F_0 típico usado por um falante: a média aritmética, a mediana e o valor de base. Em especial, nos interessamos em observar a variabilidade exibida pelas medidas quando elas são computadas de forma cumulativa ao longo de um trecho de fala com vistas a estabelecer o instante de tempo a partir do qual a variabilidade de cada medida atinge um nível de variabilidade que possa ser considerado estável. Este estudo importa para a prática da fonética forense na medida em que o nível típico de F_0 usado por um falante é uma das características usadas na comparação de vozes. As amostras usadas por peritos nesse processo de comparação podem ter durações curtas. Portanto, saber que medida estatística é mais adequada e em que condições as estimativas de F_0 típica que ela produz são confiáveis é uma questão importante na prática. Trabalhos anteriores já trataram da mesma questão, embora os resultados obtidos sejam muito variáveis em função de diferenças metodológicas. Introduzimos uma novidade na discussão ao investigar três parâmetros estatísticos e não só a média, além de lançar mão de um método estatístico para estabelecer de forma mais objetiva os pontos de estabilização das medidas estatísticas. Analisamos mais de perto o efeito de três fatores sobre o comportamento temporal das medidas estatísticas analisadas: a língua, o texto e o lapso de tempo entre duas gravações.

Em conjunto, os resultados mostraram que as três medidas atingem um patamar estável de variabilidade mais precocemente do que a maior parte das estimativas reportadas na literatura, que chegam a sugerir uma duração mínima de dois minutos. Em uma amostra de 26 línguas de oito famílias diferentes, o ponto de estabilização identificado pela técnica *changepoint analysis* nunca ocorreu mais tardiamente do que 30 segundos. Os valores médios estão em torno de 5 e 10 segundos, com desvios entre 2 e 7 segundos. Há alguma evidência de que o valor de base estabilize ligeiramente mais rápido e de forma um pouco menos variável. Os resultados também indicam que a técnica estatística empregada para a identificação dos pontos de estabilização parece fazer um bom trabalho, uma vez que a variância dos trechos após o ponto identificado é sempre menor do que a dos trechos antes dele – em torno de 40 vezes menor para a média e a mediana, e 120 vezes menor para o valor de base. O comportamento da amostra de línguas mostrou-se relativamente

bem homogêneo apesar da variabilidade de tipologias. As línguas tonais, por exemplo, não se destacaram das demais.

A comparação da leitura de dois textos diferentes pelo mesmo falante mostra que o conteúdo fonético/linguístico causa diferenças tanto na localização quanto no valor de F_0 estimado pelas três medidas. Essa variação, no entanto, está em torno de 5 segundos, semelhante à variabilidade observada para a comparação entre línguas diferentes. As diferenças no valor estimado de F_0 são bastante pequenas, na média em torno de 2%. Do ponto de vista da fonética forense, isso significa que, mesmo duas amostras de fala com conteúdo não idêntico e de duração curta, de menos de 30 segundos, produzidas pelo mesmo falante, podem gerar estimativas de localização muito próximas.

A investigação das gravações não contemporâneas teve resultados diferentes conforme o sexo do falante. As diferenças observadas entre as gravações da falante feminina são muito pequenas, deslocamento do ponto de estabilização de menos de meio segundo para os três estimadores e diferenças no valor de F_0 no ponto de estabilização de menos de um semitom, diferenças negligenciáveis em ambos os casos. Para o falante masculino, essas diferenças foram de 5,6 segundos e 2 semitons. São diferenças maiores do que as observadas para a falante feminina, mas compatíveis com aquelas observadas quando o mesmo falante lê dois textos diferentes. Amostras maiores devem ser observadas para que se possa verificar o se o padrão mais comum é o observado no falante do sexo masculino ou no feminino.

O resultado de maior significância alcançado por este estudo foi mostrar que amostras de fala relativamente curtas são suficientes para que se obtenha uma estimativa estável da F_0 típica de um falante, quer se esteja usando a média, a mediana ou o valor de base para fazer a estimativa. O fato de que os pontos de estabilização foram obtidos por um método objetivo e passível de replicação é importante porque dá aos peritos a segurança necessária para aplicar os resultados relatados aqui no seu trabalho com casos reais de comparação de vozes.

REFERÊNCIAS

1. ABE, Isamu. Intonation in Japanese. In: HIRST, Daniel; DI CRISTO, Albert. *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1998.

2. ARANTES, Pablo; ERIKSSON, Anders. Temporal stability of long-term measures of fundamental frequency. In: INTERNACIONAL CONFERENCE ON SPEECH PROSODY, 7th, 2014, Dublin. Proceedings... Dublin: s.n., 2014.
3. BARBOSA, Plinio A.; ALBANO, Eleonora C. Brazilian Portuguese. *Journal of the International Phonetic Association*, v. 34, n. 2, 2004.
4. ERIKSSON, Anders. Tutorial on Forensic Phonetics. In: EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, 9th, 2005, Lisboa. Proceedings... Lisboa: s.n., 2005.
5. ERIKSSON, Anders. Aural/Acoustic vs. Automatic Methods in Forensic Phonetic Case Work. In: NEUSTEIN, A.; PATIL, H. A. *Forensic Speaker Recognition: Law Enforcement and Counter-terrorism*. S.l.: Springer-Verlag, 2011.
6. GÅRDING, Eva. Intonation in Swedish. In: HIRST, Daniel; DI CRISTO, Albert. *Intonation Systems. A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1998.
7. HIRST, Daniel. The Analysis by Synthesis of Speech Melody: from Data to Models. Journal of Speech Sciences, v. 1, n. 1, 2011.
8. JASSEM, W. Normalisation of F_0 curves, In: FANT, G.; TATHAM, M. *Auditory Analysis and Perception of Speech*. s.l.: Academic Press, 1975.
9. KENNEY, J. F.; KEEPING, E. S. Relative Merits of Mean, Median, and Mode. In: *Mathematics of Statistics*. Princeton, NJ: Van Nostrand, 1962.
10. KILLICK, Rebecca; ECKLEY, Idris. changepoint: An R package for changepoint alysis. R package version 1.1. http://CRAN.Rproject.org/package=changepoint, 2013.
11. LEHISTE, Ilse. *Suprasegmentals*. Cambridge, MA: MIT Press, 1970.
12. LINDH, Jonas; ERIKSSON, Anders. Robustness of Long Time Measures of Fundamental Frequency. In: EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, 10th, 207, Antwerp. Proceedings... Antwerp: s.n., 2007.
13. R CORE TEAM. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. URL http://www.R-project.org/, 2014.

14. ROSE, P. How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? *Speech Communication*, v. 10, 1991.
15. TRAUNMÜLLER, Hartmut; ERIKSSON, Anders. The frequency range of the voice fundamental in the speech of male and female adults. Disponível em: <http://www2.ling.su.se/staff/hartmut/f0_m&f.pdf>. Acesso em: 25 nov. 2011.

ABSTRACT: We investigated long-term mean, median and base value of F_0 to estimate how long it takes their variability to stabilize. Change point analysis was used to locate stabilization points. In one experiment, stabilization points were calculated in recordings of the same text spoken in 26 languages. Average stabilization points are 5s for base value and 10s for mean and median. Variance after the stabilization point was reduced around 40 times for mean and median and 120 times for the base value. In a second experiment, four speakers read two different texts each. Stabilization points for the same speaker across the texts do not exactly coincide. Average time shift of change point is 2.5 seconds for the base value, 3.4s for the median and 9.5s for the mean. After stabilization, individual differences in the three measures obtained from the two texts are 2% on average. In another experiment, recordings of a male and female speaker reading the same text taken a decade apart were analyzed. Average time shift of stabilization point is 0.4s for the female and 5.2s for male speaker. F_0 estimates at stabilization points are shifted by 5% for the female and by 12% for the male speaker. Overall, results show that stabilization points in long-term measures of F_0 occur earlier than suggested in the previous literature.
KEYWORDS: Fundamental frequency; Forensic phonetics; Descriptive statistics.

Recebido no dia 20 de junho de 2014.
Aceito para publicação no dia 07 de agosto de 2014.