Revista de Estudos da Linguagem, v. 26, n. 2, p. 535-570, 2018

Investigando a robustez de uma metodologia para determinação do valor de base da frequência fundamental

Probing the robustness of a methodology to determine the base value of fundamental frequency

Pablo Arantes
Universidade Federal de São Carlos, São Carlos, São Paulo / Brasil
pabloarantes@gmail.com

Maria Érica Nascimento Linhares
Universidade Federal de São Carlos, São Carlos, São Paulo / Brasil
erica_linhares@hotmail.com

Resumo: Este trabalho testa a robustez de uma metodologia proposta pelos foneticistas suecos Traunmüller e Eriksson para determinar o valor de base, um estimador estatístico do valor típico da frequência fundamental (F0) de um falante com base na média e no desvio-padrão da F0. A metodologia consiste em estimar uma constante, *k*, que indica quantos desvios-padrão abaixo da média de F0 do falante o valor de base está. O método para estimar a constante foi criado e testado em amostras de fala atuada. Verificamos neste trabalho se a aplicação da mesma técnica a amostras de fala não atuada produz resultados comparáveis aos reportados por Traunmüller e Eriksson. A investigação usou amostras de fala produzidas por falantes nativos de alemão, estoniano, francês, inglês britânico, italiano, português brasileiro e sueco, em três estilos de elocução: entrevista, leitura de frases e leitura de palavras. Os resultados indicam que a variabilidade causada pelos estilos de enunciação na F0 possibilita a aplicação da metodologia a amostras de fala não atuada. Os valores da constante derivados dos dados não atuados são próximos aos reportados pelos autores suecos, o que indica que ela é robusta tanto do ponto de vista dos falantes quanto das línguas.

Palavras-chave: entoação; valor de base; frequência fundamental.

eISSN: 2237-2083
DOI: 10.17851/2237-2083.26.2.535-570

Abstract: This paper probes the robustness of Traunmüller and Eriksson's methodology to determine the base value of the fundamental frequency of speech, an estimator of a speaker's typical F0 value. The methodology entails the estimation of a constant, *k*, indicating where the base value for a speaker lies in relation to F0 standard deviations below the F0 mean. The methodology was originally developed from acted speech samples. Here we test if *k* values can be successfully obtained from non-acted samples and how they compare to the ones reported by Traunmüller and Eriksson. A speech corpus of speech samples differing in speaking styles (spontaneous interview, sentence reading, word list reading) from seven languages (English, Estonian, French, German, Italian, Brazilian Portuguese, Swedish) was used. Results show that *k* values estimated from non-acted speech are roughly the same as those reported in Traunmüller and Eriksson's original paper. We speculate that deviations can be explained by the fact that some speakers make extensive use of non-modal register.

Keywords: intonation; base value; fundamental frequency.

Recebido em 10 de dezembro de 2016

Aceito em 6 de junho de 2017

1 Introdução

Ao longo de um enunciado, a frequência fundamental da voz (F_0) varia em razão de fatores de diferentes naturezas: fatores linguísticos de escopo amplo, como a modalidade do enunciado, ou locais, como a composição fonética dos segmentos que formam o enunciado – cf. a distinção entre micro e macromelodia em Hirst (2005); fatores paralinguísticos, como o estado emocional do falante no momento da enunciação e, ainda, fatores orgânicos, como o sexo e as idiossincrasias do trato vocal do falante – principalmente massa e comprimento das pregas vocais (TITZE, 1994). Essa multiplicidade de fatores dificulta a estimativa do valor médio ou típico da F0 que um falante emprega em suas produções faladas.

Em alguns cenários, é interessante que a estimativa de valor típico da F_0 reflita fatores orgânicos mais do que fatores linguísticos. Duas situações desse tipo são, por exemplo, a comparação de vozes

com finalidade forense (JESSEN, 2008) e o uso da voz como meio para autenticação da identidade de um usuário em aplicações de segurança (SCHULTZ, 2007). Em aplicações como essas, a influência que o conteúdo linguístico de enunciados específicos possa vir a exercer sobre os contornos de F_0 produzidos por um indivíduo não está no centro das atenções. O que se busca, ao contrário, é minimizar essas influências de forma a fazer os fatores orgânicos/biológicos ressaltarem no estimador estatístico de valor típico.

Pode-se pensar em uma situação em que o inverso seja verdadeiro, isto é, em que o interesse se volta para os efeitos de um contraste linguístico sobre o comportamento da F_0, independentemente dos falantes que expressam esse contraste. Pode-se estar interessado, por exemplo, em estabelecer o efeito da modalidade interrogativa sobre o contorno de F_0. Não interessam, nesse caso, diferenças mais ou menos esperadas entre falantes, como o fato da F_0 de homens ser em geral menor do que a de mulheres. O importante é tentar neutralizar essas características idiossincráticas e pôr em relevo o modo pelo qual a variação de F_0 expressa o contraste linguístico em questão. Procedimentos de normalização da curva de F_0 podem ser usados para essa finalidade e são geralmente empregados em cenários nos quais diferentes falantes produzem repetições de enunciados em que existe algum contraste linguístico sob investigação. Esses procedimentos, de forma geral, requerem o uso de uma estimativa do valor típico da F_0 (em geral a média aritmética) de cada um dos falantes que contribuíram com enunciados para um determinado *corpus* (JASSEM, 1975; MAIDMENT; LECUMBERRI, 1996; ROSE, 1991).

Ambos os cenários discutidos anteriormente deixam clara a importância e a utilidade de estudar as características estatísticas das curvas de F_0, em especial a adequação das diferentes maneiras de obter uma estimativa do valor típico ou tendência central dessas amostras. No contexto da estatística descritiva, há diversos procedimentos para determinar o valor mais representativo de uma amostra de dados, cada qual com vantagens e limitações próprias (KENNEY; KEEPING, 1962). A média e mediana são estimadores de localização versáteis, no sentido de que podem ser aplicados a amostras de qualquer natureza, desde que a variável observada possa ser medida em uma escala intervalar ou proporcional (STEVENS, 1946). A média aritmética é o estimador de tendência central de F_0 cujo uso é mais

prevalente na literatura, apesar de sua sensibilidade à presença de assimetria na amostra, que é bastante comum em dados de F0 (JASSEM, 1975). A mediana, mais robusta à presença de assimetrias e valores extremos, é uma alternativa à média – cf. a proposta de De Looze e Hirst (2014) para o uso da mediana como valor de referência para um procedimento de normalização de contornos de F_0.

O valor de base (*base value* ou *base line* em inglês) é um estimador estatístico de localização proposto pelos foneticistas suecos Traunmüller e Eriksson [s.d.] especialmente para amostras de F_0 e leva em conta as especificidades típicas desse tipo de amostra. Uma dessas especificidades é que a variação de F0 em geral não é simétrica, como se viu no parágrafo anterior. Quando os falantes fazem excursões entoacionais, o movimento, na grande maioria das vezes, é ascendente, fato que se revela nos histogramas de distribuições de F0 como uma assimetria positiva. Eriksson (2011) sugere que o nível de F0 que pode ser considerado típico para um falante é aquele logo acima do mínimo necessário para manter a fonação modal. Movimentos abaixo desse nível seriam, segundo o autor, menos comuns porque poderiam resultar em vozeamento não modal. Em situações que fazem a variabilidade da F_0 aumentar, como, por exemplo, falar com maior envolvimento emocional, essa tendência à assimetria se mostra ainda mais claramente. O gráfico da figura 1 mostra o contorno de F0 normalizado temporalmente da mesma frase[1] lida pelo mesmo falante – um ator, simulando três níveis de envolvimento emocional, com níveis de vivacidade crescentes. Em verde, o contorno da elocução com um nível neutro ou típico de envolvimento; em vermelho, baixo envolvimento e, em azul, alto grau de envolvimento. É bastante evidente no gráfico que quanto maior é o envolvimento, maior a gama de valores explorados pelas excursões de F_0. As excursões, no entanto, têm uma direção preferencial: as curvas, independentemente do nível de envolvimento, raramente descem abaixo de um ponto em torno de 100 Hz, que funciona como um piso a partir do qual o falante expande a gama tonal. Esse ponto seria o valor de base para esse falante.

Traunmüller e Eriksson ([S.d.]) desenvolvem uma metodologia para estimar o valor de base (*base value*, em inglês), F_b, de uma amostra de F0, e propõem a fórmula $F_b = F_{média} - k\sigma$, em que $F_{média}$ e σ são, respectivamente, o valor da média aritmética e do desvio padrão de F_0

[1] As gravações foram cedidas por Anders Eriksson.

de uma amostra de F_0, e k é uma constante determinada empiricamente. Em um experimento com fala atuada emulando diferentes funções paralinguísticas, Traunmüller e Eriksson ([S.d.]) obtiveram o valor de 1,5 para a constante, mas indicaram que esse valor não é fixo e pode apresentar uma variação entre 1,1 e 2 – valores obtidos com base em conjuntos de dados diferentes e replicações subsequentes da análise original.

Lindh e Eriksson (2007), em um estudo posterior, revisaram o valor de k para 1,43 e sugeriram uma formulação alternativa para o cálculo do valor de base, que se mostrou mais robusta do que a original. Nessa formulação, chamada por eles de *alternative base value*, assumindo uma distribuição normal para os dados de F0, o ponto $1{,}43 \cdot \sigma$ abaixo da média corresponde, aproximadamente, ao 7º percentil da distribuição empírica de F_0.

No presente trabalho, testamos a robustez da metodologia apresentada por Traunmüller e Eriksson ([S.d.]) para a determinação do valor de base da frequência fundamental da voz. Para isso, ela será aplicada a amostras de fala não atuada, produzidas por falantes de sete línguas: alemão, estoniano, francês, inglês britânico, italiano, português brasileiro e sueco, a fim de observar se os valores da constante k estimados pela fala não atuada são comparáveis aos obtidos pelos autores por meio da fala atuada. Além disso, uma vez que a estimativa de k pode variar, investigaremos o grau de sensibilidade do valor de base em função das variações de k, que também consideramos ser uma forma de avaliar a robustez da proposta de Traunmüller e Eriksson para determinar o valor de base.

Outros modelos presentes na literatura propõem conceitos comparáveis ao valor de base de Traunmüller e Eriksson. Embora o propósito central do presente trabalho seja testar a robustez do modelo proposto por eles, discutiremos brevemente as semelhanças e diferenças entre eles. Destacamos dois modelos em particular: Gårding (1983) e Fujisaki e Hirose (1984). Ambos propõem modelar o contorno de F_0 de frases individuais pela sobreposição de componentes que atuam em diferentes níveis.

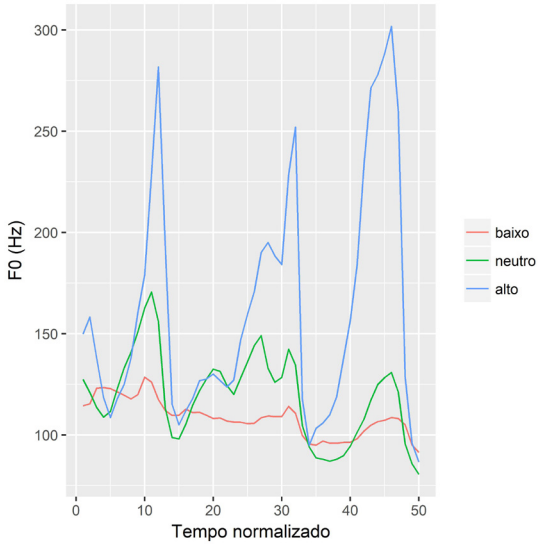
Em Gårding (1983), os componentes são lexicais e frasais. No componente frasal estabelece-se a grade tonal, que funciona como um quadro global para a frase ao definir continuamente valores mínimos e máximos para a variação de F0 aos quais os tons locais serão sobrepostos. A linha inferior da grade tonal poderia ser posta em comparação ao valor de base de Traunmüller e Eriksson. Na referência citada, não

há informações detalhadas a respeito do procedimento adotado para a definição dos valores da grade para cada frase a ser analisada. Nos casos em que a linha inferior da grade tonal tem uma inclinação negativa, no entanto, ela seria mais bem comparada à tendência de declinação (VAISSIÈRE, 1983) do que ao valor de base.

Em Fujisaki e Hirose (1984), o contorno observável de F0 em um enunciado é considerado o resultado da sobreposição de dois componentes – um frasal e um acentual – que modulam uma frequência de base, valor que é considerado específico para cada falante. Nesse modelo, a frequência de base é comparável ao valor de base de Traunmüller e Eriksson. Mixdorff (2015) discute diferentes abordagens para a determinação da frequência de base. Do ponto de vista conceitual, faria sentido considerar a frequência de base um valor relativamente fixo para cada falante. Mixdorff, no entanto, determina a F_b do modelo de Fujisaki e Hirose com base em informação sobre os componentes de baixa frequência de cada curva de F0 da frase a ser modelada. Como esse procedimento é aplicado em frases relativamente curtas, ele tem o inconveniente de ser suscetível a variações locais, conforme se esteja modelando frases de diferentes modalidades, por exemplo (ver figura 3.3 em MIXDORFF, 2015, p. 39). Nesse exemplo, o valor da frequência de base da frase interrogativa extraído de forma automática não coincide com o menor valor do contorno. Além disso, é quase 30 Hz mais alto do que a frequência de base de uma declarativa produzida pelo mesmo falante.

Essa breve discussão mostra a importância de discutir procedimentos de determinação de valores que podem ser postos em equivalência tanto com valor de base de Traunmüller e Eriksson quanto com a frequência de base de Fujisaki e Hirose. No caso dos procedimentos apresentados em Mixdorff (2015), a determinação do valor da frequência de base não é guiada por um critério motivado em princípios fortemente articulados ao próprio modelo de Fujisaki ou a outra teoria de produção da fala. No caso de Traunmüller e Eriksson, o modelo teórico mais geral que embasa sua proposta é a teoria da modulação (TRAUNMÜLLER, 1994), que é mais ampla em escopo do que o modelo de Fujisaki e Hirose.

FIGURA 1 – Contornos normalizados temporalmente de uma mesma frase interpretada em três níveis de envolvimento emocional por um ator sueco[2]



Fonte: Elaborado pelos autores.

2 A metodologia de Traunmüller e Eriksson

Utilizamos, neste trabalho, a metodologia descrita em Traunmüller e Eriksson ([s.d.]) para derivação da fórmula do valor de base. Os autores sugerem que o valor de base pode ser entendido intuitivamente como o valor de F_0 que corresponderia à situação em que o falante produzisse fala sem nenhuma variação entoacional, isto é, com variabilidade de F0 nula, condição que corresponde ao conceito de carreador na teoria da modulação, proposta por Traunmüller (1994). O desvio-padrão da F_0 dessa situação idealizada seria zero, e a média observada refletiria a F_0 típica ou preferida daquele falante. A fórmula proposta pelos autores, mencionada na seção anterior, calcula o valor de base por meio de duas incógnitas, $F_{média}$ e σ, que podem ser facilmente estimadas com base em amostras de F_0, e uma constante, k. A metodologia apresentada pelos autores no trabalho citado apresenta uma maneira empírica de chegar

[2] A frase, em sueco, é "Nån av mammorna hann lämna honom", e uma tradução aproximada seria "Algumas das mães puderam deixá-lo".

a um valor para *k* utilizando um *corpus* de gravações. A presente seção apresenta os princípios fundamentais dessa metodologia.

Uma vez que amostras de fala reais sempre apresentarão alguma variabilidade, é preciso estimar o valor de F_0 que corresponderia a um contorno perfeitamente monotônico considerando-se dados naturais. Essa estimativa é feita por meio da aplicação da técnica de regressão linear. Para tanto, é preciso dispor de uma série de pares de valores <média, desvio padrão>, extraídos de um *corpus* de fala natural. Traunmüller e Eriksson recorreram à fala atuada, em razão de esse estilo de enunciação possibilitar eliciar o mesmo conteúdo linguístico sob diferentes condições paralinguísticas, que induzem a produção de variabilidade nos contornos de F_0. Por meio da distribuição dos dados no plano cartesiano formado pelas dimensões média e desvio-padrão, a técnica de regressão linear possibilita estimar o valor que a média de F_0 teria se o desvio-padrão fosse nulo, o qual corresponderá ao valor de base para aquele falante.

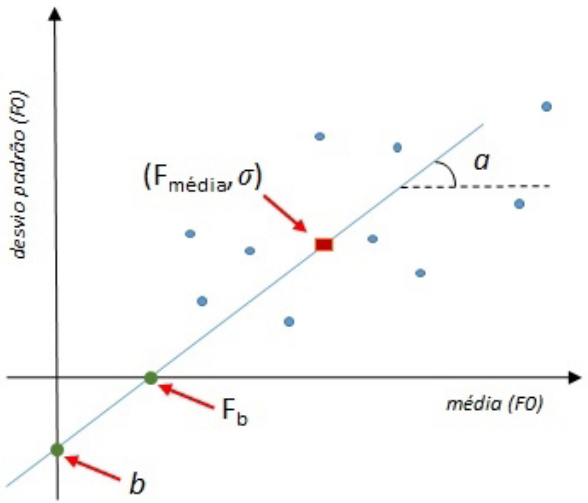
A aplicação da regressão linear estima a inclinação da reta que melhor descreve a relação linear entre os pontos presentes no plano. Se usarmos a equação $y = a \cdot x + b$ para descrever essa reta, então a inclinação corresponde ao parâmetro a, y corresponde aos valores de desvio padrão, e x, aos valores da média de F_0. O valor de base corresponderia ao valor médio de F_0 para o qual o desvio padrão seria nulo, o que corresponderia, linguisticamente, à F_0 que um falante produziria numa fala hipotética perfeitamente monotônica, não influenciada pelos diversos fatores que produzem variação em seu valor.

Dada a reta estimada pela análise de regressão linear, o valor de base (F_b), isto é, o ponto em que a linha de regressão cruza a linha horizontal $y = 0$ pode ser obtido pela expressão $F_b = -b/a$.

Assumindo que F_b é um valor que se aproxima do limite inferior da gama de valores de F_0 produzida pelo falante e que a distribuição dos valores de F_0 pode ser razoavelmente aproximada por uma distribuição normal, podemos propor a expressão $F_b = F_{média} - k\sigma$ para determinar o valor de F_b. Substituindo F_b, $F_{média}$ e σ pelos valores obtidos empiricamente na amostra analisada, obtém-se k. Esse valor de k pode ser usado na expressão proposta anteriormente para determinar o valor de base de qualquer amostra de F_0. Mesmo sendo derivado com base em dados de apenas um falante, os autores sugerem que o valor da constante k obtido dessa maneira deve, em princípio, funcionar bem para encontrar o valor de base em amostras de fala de qualquer falante em qualquer língua.

A figura 2 é uma representação esquemática das informações da regressão linear relevantes para a aplicação da metodologia de Traunmüller e Eriksson. Na figura, os pontos azuis são hipotéticos pares de valores <média, desvio-padrão> coletados em um *corpus*, o quadrado vermelho está localizado no ponto que corresponde à média das médias e à média dos desvios-padrão. A linha azul é a linha de regressão linear estimada a partir dos pontos, *a* indica o coeficiente de inclinação da reta, *b*, o ponto em que a reta intercepta o eixo *y*, e F_b é a localização do valor de base, isto é, o ponto no eixo *x* (média de F0) quando o desvio-padrão (eixo *y*) tem valor 0.

FIGURA 2 – Representação esquemática das informações da regressão linear relevantes para a aplicação da metodologia de Traunmüller e Eriksson



Fonte: Elaborado pelos autores.

3 Materiais e métodos

3.1 Materiais de fala

O material de fala usado no experimento vem do *corpus* coletado no âmbito do projeto internacional "A typology for word stress and speech rhythm based on acoustic and perceptual considerations", coordenado pelo professor Anders Eriksson da Universidade de Estocolmo, Suécia.[3]

[3] O autores deste trabalho não têm relação com o projeto.

O *corpus* compreende dados de sete línguas: alemão, estoniano, francês, inglês britânico, italiano, português brasileiro e sueco. As amostras das línguas individuais foram coletadas por pesquisadores integrantes do projeto em países em que cada uma das línguas é falada. Em virtude da uniformidade dos procedimentos de coleta, o *corpus* possibilita a comparação interlinguística do fenômeno de interesse em línguas com características diversas. São contempladas seis línguas da família indo-europeia (três do ramo românico e três do ramo germânico) e uma da família urálica (estoniano). Além da variedade de línguas, outra razão para a escolha desse *corpus* para uso no projeto é o fato de as amostras de fala variarem em termos do estilo de elocução. A literatura mostra que a variação no estilo de elocução é um dos fatores que causam variabilidade em medidas de longo termo de F_0, como a média e o desvio-padrão (ESKÉNAZI, 1993; HOLLIEN; HOLLIEN; JONG, 1997; LLISTERRI, 1992). Essa variabilidade é importante no contexto do presente trabalho porque possibilita a aplicação da regressão linear como método para estimar como o valor médio de F_0 varia em função do desvio-padrão, um dos fundamentos da metodologia de Traunmüller e Eriksson ([S.d.]), descrita na seção 2. Três estilos são coletados: entrevista, leitura de frases e leitura de palavras. No estilo entrevista, um entrevistador (em geral um membro da equipe do projeto) fez perguntas ao participante sobre assuntos como trabalho, estudos e outros interesses do entrevistado, visando obter respostas não planejadas e de extensão variável. Para o estilo leitura de frases, um membro da equipe do projeto selecionou frases ditas pelo participante na entrevista, transcreveu-as ortograficamente e pediu que o participante as lesse em voz alta em uma sessão de gravação realizada alguns dias após a entrevista. No estilo leitura de palavras, o procedimento consistiu na escolha de uma palavra de cada frase presente na etapa anterior e na sua apresentação ao participante na forma de uma lista a ser lida. Foram analisadas amostras de fala de dez falantes de cada língua, cinco do sexo masculino e cinco do feminino, uma amostra de cada estilo, totalizando 210 amostras de fala (= 7 línguas ×10 falantes × 3 estilos).

3.2 Extração dos dados

A primeira parte da análise consistiu na extração dos valores de F_0 de cada uma das 210 amostras de fala do *corpus*. A extração se deu em duas etapas: na primeira, o contorno de F_0 foi extraído por meio do uso de um *script* do programa Praat (BOERSMA, 2001) escrito pelo primeiro

autor, que otimiza a escolha dos parâmetros *floor* e *ceiling* do algoritmo de extração de F_0 do Praat. Essa heurística de otimização, proposta por Hirst (2011), tem o objetivo de diminuir os erros de estimação de F_0 mais comuns produzidos pela função *To Pitch*, baseada na técnica da autocorrelação; na segunda etapa, os arquivos *Pitch* gerados na fase anterior foram corrigidos manualmente. Nessa fase, um segundo *script* foi usado para auxiliar a identificação dos erros não eliminados na etapa anterior. O *script* identifica duas amostras de F_0 sucessivas, separadas por 80 milissegundos ou menos, em que o primeiro valor é 1,5 vezes maior ou menor do que o segundo. Nos pontos do contorno de F_0 indicados pelo *script* como suspeitos de conter erro de extração, o trecho do oscilograma correspondente foi examinado visualmente para que fosse possível decidir se os valores de F_0 estimados pelo Praat naquele trecho correspondiam à periodicidade identificada visualmente na forma de onda.

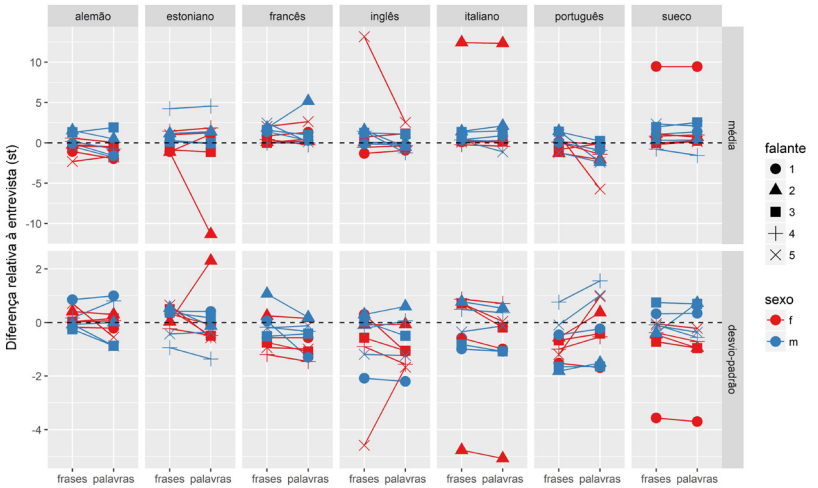
Os valores de F_0 foram mantidos na escala física Hertz (Hz) nas análises posteriores. Por conta de características típicas de amostras de F_0 mencionadas na Introdução, como o fato de serem frequentemente assimétricas e não se conformarem a uma distribuição normal, é comum que dados de F_0 sejam convertidos para uma escala não linear, como a escala de semitons. A decisão de manter os dados na escala Hertz, neste trabalho, foi tomada por uma questão de replicabilidade, uma vez que essa foi a escala usada por Eriksson nos trabalhos realizados no curso "Paralinguistic aspects of speech production and perception", realizado no Instituto de Estudos da Linguagem da Universidade Estadual de Campinas entre os dias 8 e 10 de abril de 2014. Nesse curso, os autores do presente trabalho foram treinados nos aspectos práticos da aplicação da metodologia que Eriksson e Traunmüller apresentam em seu trabalho seminal. Eriksson nos informou em comunicação pessoal que, nos materiais de fala usados no referido curso, a adoção da escala Hertz ou de semitons produz diferenças negligenciáveis nos valores estimados da constante *k*.

3.3 Verificação da congruência entre variação na média e no desvio-padrão

Na etapa de análise seguinte, um outro *script* do Praat processou os contornos corrigidos de F_0 das 210 amostras, para extrair os valores de média e desvio-padrão de cada um. Esses valores foram usados para verificar um pressuposto da metodologia dos autores suecos. Para que a técnica de regressão linear possa ter sucesso na estimativa do valor da constante *k*, é necessário que haja variação nas médias e nos desvios-

padrão dos três estilos de fala e que a variação no desvio-padrão seja diretamente proporcional à variação na média, isto é, o estilo com maior valor de média deve apresentar também o maior valor de desvio-padrão e vice-versa. Caso isso não ocorra, a reta estimada pela regressão pode ter um coeficiente de inclinação nulo ou negativo, o que resulta em um valor negativo para a constante *k*. Valores negativos para *k* não fazem sentido, pois resultariam em valores de base localizados acima da média de F_0, contrariando a intuição que fundamenta a proposição do valor de base.

FIGURA 3 – Diferenças (em semitons) entre a média e o desvio-padrão dos estilos entrevista e leitura de frases e entrevista e leitura de palavras



Fonte: Elaborado pelos autores.

A figura 3 mostra um panorama da relação entre a variação nos parâmetros média e desvio-padrão nos três estilos de elocução nas setes línguas do *corpus*. A figura mostra as diferenças entre os valores da média e do desvio-padrão das amostras de fala tanto do estilo leitura de frases quanto leitura de palavras em relação à média do estilo narrativa para os dez falantes de cada língua. A diferença entre os estilos foi calculada entre os valores de média e desvio-padrão expressos na escala de semitons. Esse procedimento foi adotado para que não houvesse grandes discrepâncias entre os dados dos falantes do sexo feminino e masculino. As falantes do sexo feminino são identificadas pela cor vermelha, e os masculinos, pela cor azul. Os cinco falantes de cada sexo são identificados por símbolos

diferentes, conforme a legenda. Os pontos abaixo da linha horizontal pontilhada correspondem aos casos em que os valores para os estilos leitura de frases ou palavras (indicados por marcas no eixo horizontal) são menores do que os da narrativa para o falante em questão. Os pontos acima da linha correspondem a casos em que o valor do estilo narrativa é menor do que aquele ao qual ele é comparado.

A observação da figura 3 mostra que a influência dos estilos de elocução sobre a variabilidade da média e do desvio-padrão de F_0 não é uniforme entre os falantes e entre as línguas. Os falantes m4 do português, m3 do sueco e m1 do inglês são exemplos em que diferenças no desvio-padrão entre os estilos vão na mesma direção das diferenças na média – no caso dos dois primeiros, frases > entrevista e palavras > entrevista e, no caso do último, frases < entrevista e palavras < entrevista. Essa configuração favorece a aplicação da metodologia de Traunmüller e Eriksson ([s.d.]). Há casos em que a diferença entre os estilos observada na média se dá em sentido oposto no desvio-padrão, como ilustram os falantes f2 do italiano e m4 do estoniano: frases > entrevista e palavras > entrevista nas médias, mas frases < entrevista e palavras < entrevista nos desvios-padrão.

No que diz respeito à influência da língua sobre os valores de média e desvio-padrão de F_0, a figura 3 mostra que, no português, o estilo entrevista tem médias e desvios-padrão maiores do que os outros dois estilos. No francês e no italiano, por outro lado, predominam casos em que o estilo entrevista tem as médias mais baixas, embora esse padrão não se reflita no desvio-padrão. A tabela 1 apresenta a porcentagem de falantes em cada língua cuja variação de desvio-padrão é diretamente proporcional à da média. Para esse cálculo, as comparações entrevista-frases e entrevista-palavras foram agrupadas. A inspeção da tabela confirma que o português foi a língua na qual a estratégia de manipulação do estilo de elocução foi mais bem-sucedida no sentido de produzir dados adequados à aplicação da metodologia a ser testada. Entre os estilos, a porcentagem é de 57% tanto nas comparações entrevista-frases quanto nas comparações entrevista-palavras. Entre os sexos, a porcentagem é de 51% para as mulheres e 63% para os homens. Os dados de todas as línguas foram agrupados para a realização do cálculo nas comparações entre estilos e sexos.

TABELA 1 – Porcentagem de falantes cuja variação entre a média e o DP se dá no mesmo sentido

Língua	%
Alemão	60
Estoniano	55
Francês	55
Inglês	55
Italiano	55
Português	75
Sueco	60

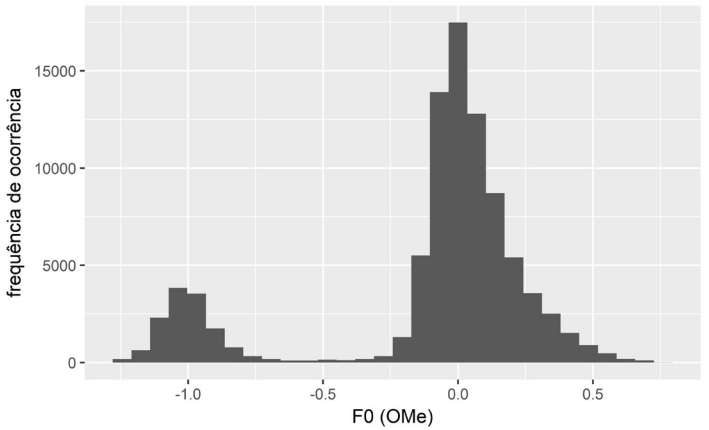
Fonte: Elaborada pelos autores.

Considerando todas as 140 diferenças pareadas (sexo e língua agrupados), os 60 casos de incongruência entre a variação na média e no desvio-padrão dividem-se igualmente entre as comparações frases-entrevista e palavras-entrevista. Em 82% dos casos em que não houve congruência, isso se deveu ao fato de o estilo entrevista apresentar média menor do que o outro estilo do par, embora seu desvio-padrão fosse o maior. O português apresenta o maior índice de congruência. Em uma publicação que analisa o mesmo *corpus* (ARANTES; LINHARES, 2017) e procura mostrar o efeito da língua, estilo de elocução e sexo dos falantes sobre descritores estatísticos de longo termo de F_0, observa-se que o português é a única língua na amostra para a qual o estilo entrevista teve valores de média estatisticamente maiores do que os outros estilos. Estoniano, francês e italiano mostram a tendência inversa, significativa do ponto de vista estatístico. Em termos do desvio-padrão, por outro lado, o estilo entrevista apresenta valores mais altos do que os demais estilos, e essa diferença é estatisticamente significativa em todas as línguas.

Uma das explicações para a incongruência entre o comportamento da média e do desvio-padrão, especialmente o caso em que a média da entrevista não é a maior entre os estilos, mas o desvio-padrão é, pode ser a presença de registro vocal não modal nas amostras de fala. A figura 4 mostra o histograma dos valores de F0 da amostra do estilo entrevista da falante f2 do italiano. Os valores de F_0 estão expressos na escala OMe (O*ctave Median*), proposta por De Looze e Hirst (2014). Os valores de F_0 em Hz (f_{Hz}) são transformados para a escala OMe (f_{OMe}) por meio da fórmula $f_{OMe} = \log_2(f_{Hz}/f_{med})$, onde f_{med} é o valor da mediana de F_0 do falante, estimada com base em todos os valores presentes no contorno

a ser convertido.[4] O histograma indica que a amostra de F0 é bimodal.[5] A parte da distribuição centrada no valor -1 está uma oitava abaixo da mediana, que para essa falante é 207 Hz. A inspeção do histograma ajuda a entender que a bimodalidade tem como efeito baixar a média da amostra (183 Hz para o contorno todo, 228 Hz excluindo da amostra de F0 os valores abaixo de -0.35 OMe), mas aumentar seu desvio-padrão (48 Hz se toda a amostra for considerada, 25 Hz se apenas os valores acima de -0.35 OMe forem considerados). No caso dessa falante, quase 38% de todos os valores de F0 da amostra estão bastante abaixo do valor mediano, concentrados em um uma região quase uma oitava abaixo da mediana da amostra completa.

FIGURA 4 – Histograma da distribuição de F0 (na escala OMe) da falante italiana f2, estilo entrevista



Fonte: Elaborado pelos autores.

[4] A utilidade dessa escala está no fato de que ela usa um valor considerado típico para o falante – a mediana – como fator de normalização para todos os valores de um determinado contorno e expressa a variabilidade de F_0 em torno do valor de referência em termos de oitavas. Essa operação possibilita identificar facilmente os valores que estão muito acima ou abaixo do valor da mediana nos histogramas.

[5] Bimodal no sentido estatístico e não no sentido de voz bitonal, que apresenta simultaneamente vibrações de duas frequências diferentes.

3.4 Extração da constante *k*

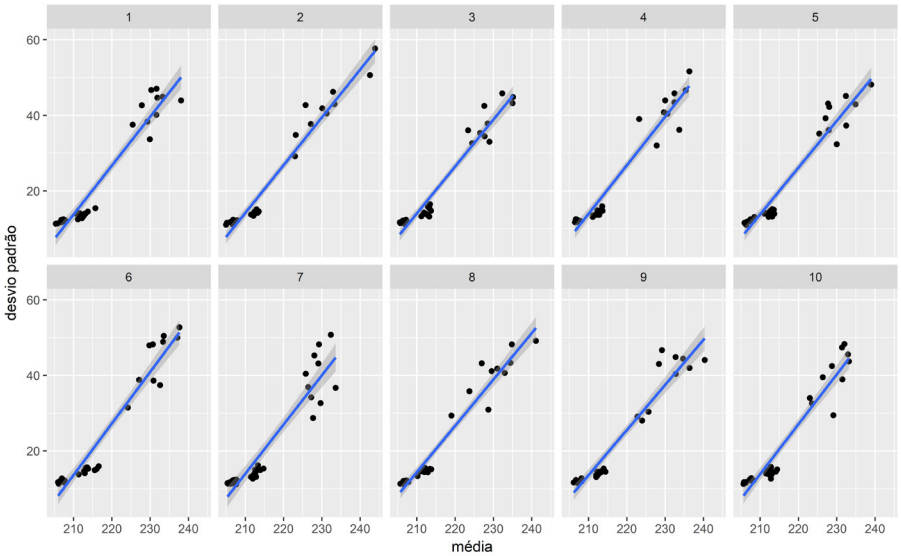
Para estimar o valor da constante *k* usando a metodologia de Traunmüller e Eriksson é preciso dispor de uma distribuição de valores de média e desvio-padrão de F_0. Esses valores foram gerados a partir dos contornos de F_0, cuja extração é descrita na seção 3.2, segundo o procedimento descrito a seguir. Os arquivos de som foram segmentados manualmente para identificar os trechos de fala, e as marcações foram armazenadas em arquivos *TextGrid* do programa Praat. Nas amostras do estilo entrevista, foram marcados os trechos de fala entre pausas maiores do que 300 ms. Nos estilos leitura de frases e leitura de palavras, foram marcadas as frases e palavras individuais. Um *script* do Praat foi desenvolvido para selecionar aleatoriamente trechos marcados no arquivo *TextGrid* até que a duração acumulada desses trechos atinja pelo menos 60 segundos. O contorno de F0 dos trechos individuais selecionados é concatenado, e a média e o desvio-padrão do contorno resultante são calculados. A operação é repetida dez vezes para cada estilo de fala, de modo que são obtidos para cada falante trinta pares <média, desvio-padrão>. O procedimento de regressão linear é aplicado aos trinta pontos da amostra, e o valor da constante *k* é determinado pelos parâmetros relevantes, conforme explicado na seção 2.

Dado o componente aleatório no procedimento descrito no parágrafo anterior, decidimos investigar se as estimativas de *k* para cada falante produzidas por sua aplicação é estável. Para tanto, o procedimento descrito no parágrafo anterior foi repetido dez vezes para cada falante, de modo que para cada um deles obtivemos dez estimativas para o valor de *k*.

4 Resultados e discussão

A figura 5 mostra os gráficos de dispersão e a curva de regressão linear ajustada aos dados dos 10 conjuntos coletados para o falante f1 do português. É possível ver que, apesar de haver alguma variabilidade, a distribuição dos dados em cada gráfico de dispersão é bastante similar, o que indica que a estimativa de *k* é estável para esse falante em particular. Os valores de *k* variam entre 0,73 e 0,84, com coeficiente de variação de 0,4%. Os valores altos do coeficiente de determinação (r^2) da regressão linear – entre 0,88 e 0,95 – indicam um bom ajuste da reta em relação aos pontos.

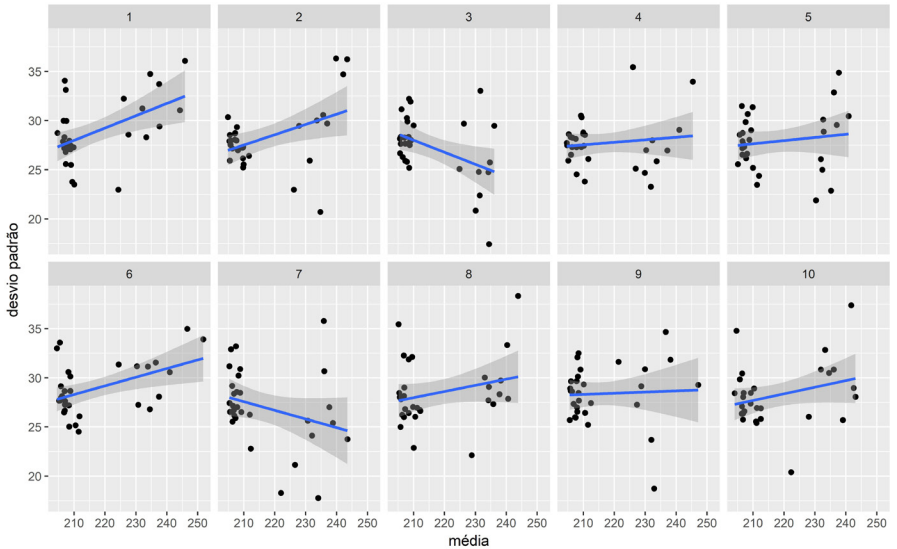
FIGURA 5 – Gráficos de dispersão (média e desvio-padrão da F0 em Hz) com reta de regressão linear superposta de 10 amostras da falante brasileira f1



Fonte: Elaborados pelos autores.

Em contraste, a figura 6 ilustra o caso de uma falante, f2 do português, cujo padrão de variação da média e do desvio-padrão não é adequado à aplicação da metodologia de estimação da constante *k*. É possível observar que a reta de regressão ora tem inclinação positiva (repetições 1 e 6, por exemplo), ora, inclinação negativa (repetições 3 e 7, p.e.) e, em alguns casos, aparenta ter inclinação nula (repetição 9). Conforme é possível observar na figura 3, a falante apresenta diferenças na média entre os estilos (entrevista maior do que leitura de frases e palavras), embora o desvio-padrão seja basicamente o mesmo para os três estilos. Essa característica não faz dessa falante uma boa candidata à aplicação da metodologia de estimativa de *k* por meio da análise de regressão. Podemos ver isso na imensa variabilidade dos valores de *k* que a técnica estima para esse falante: mínimo de -11,36 e máximo de 38,89, com coeficiente de variação de 140%. Os valores de r^2 são bastante baixos, variando entre 0,002 e 0,26.

FIGURA 6 – Gráficos de dispersão (média e desvio-padrão da F0 em Hz) com reta de regressão linear superposta de 10 amostras da falante brasileira f2

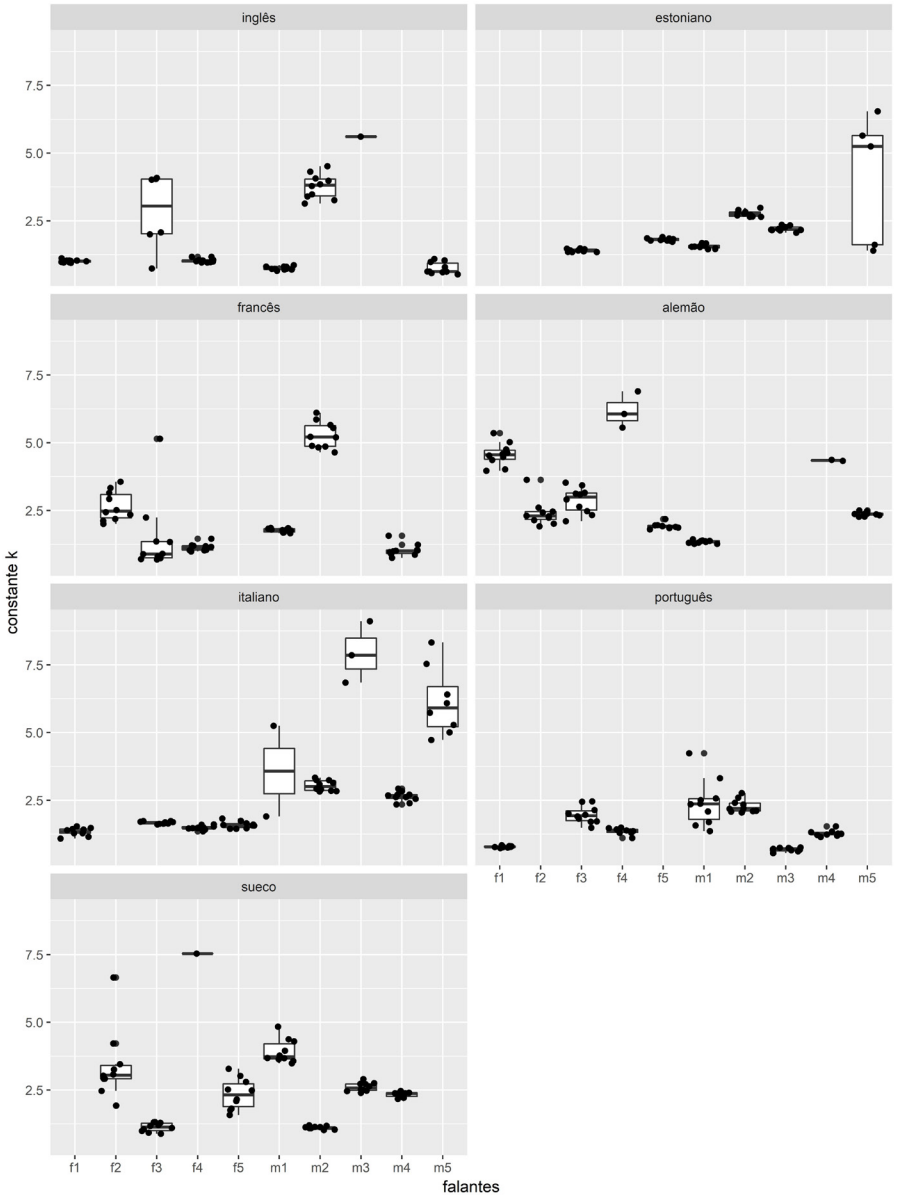


Fonte: Elaborados pelos autores.

A figura 7 mostra a distribuição dos valores da constante *k* estimados para as sete línguas do *corpus*. Os falantes estão dispostos no eixo horizontal, e os valores estimados para a constante *k* aparecem no eixo vertical. Cada ponto corresponde a uma estimativa do valor de *k*. Em todas as línguas há falantes, como f2 do português, para os quais a aplicação da metodologia resulta em valores de *k* negativos, que não fazem sentido e são omitidos. O número de falantes que se enquadram nesses casos variou entre um no italiano e quatro no francês e no estoniano.[6] O valor médio de *k* para a amostra total é 2,24 com intervalo de confiança de 95% em torno da média de ± 0,13. A tabela 2 lista a média, o intervalo de confiança em torno da média e o coeficiente de variação das estimativas do coeficiente *k* para cada língua.

[6] A título de comparação, podemos observar que, nos dados mostrados na figura 2 de Traunmüller e Eriksson ([s.d.], p. 8), três dos dez falantes não apresentam variação congruente entre média e desvio-padrão de F_0, uma proporção semelhante à que observamos nas amostras analisadas aqui.

FIGURA 7 – Valores da constante ***k*** em função dos falantes e da língua



Fonte: Elaborado pelos autores.

TABELA 2 – Média, intervalo de confiança de 95% e coeficiente de variação da constante ***k*** para cada língua

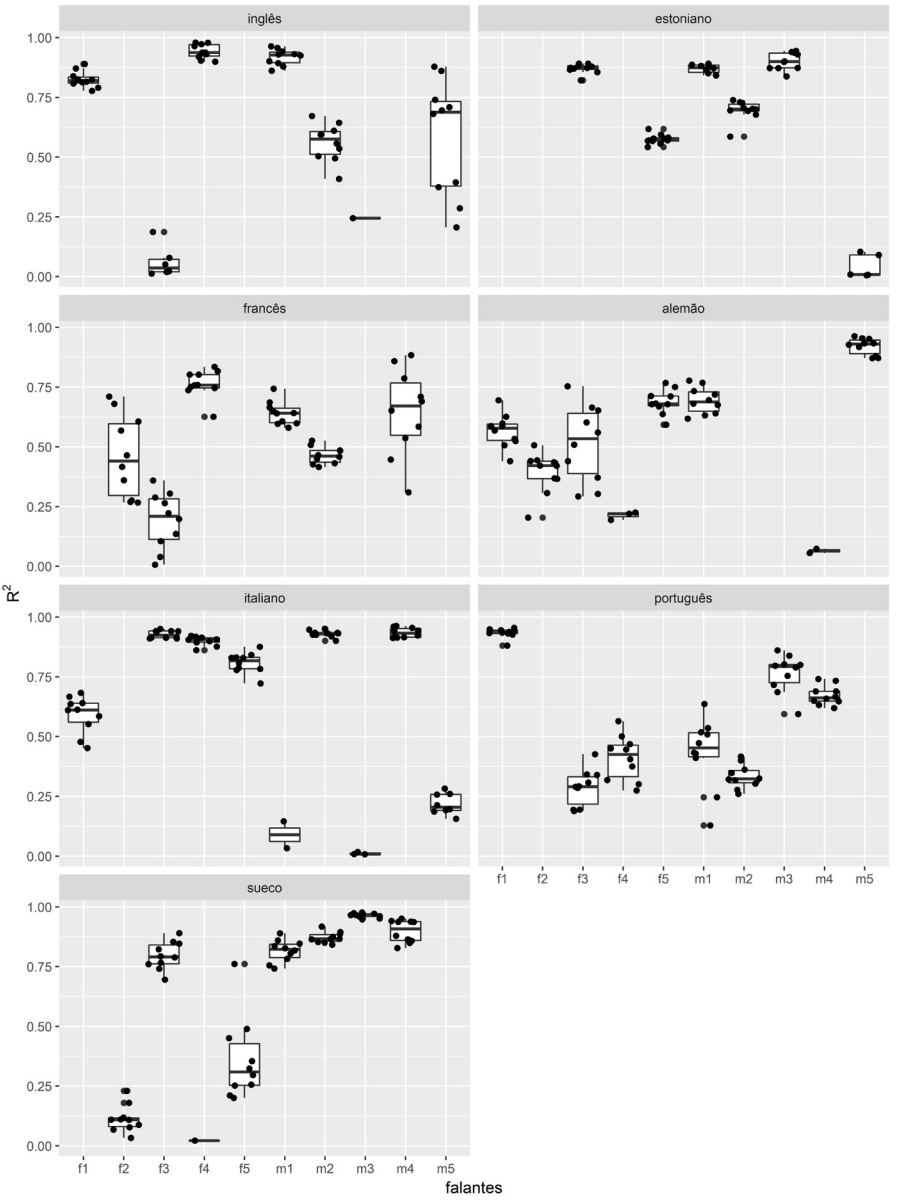
Língua	Média	Intervalo de confiança	CV (%)
Alemão	2,8	± 0,32	47
Estoniano	2,14	± 0,27	48
Francês	2,23	± 0,4	72
Inglês	1,68	± 0,36	82
Italiano	2,71	± 0,43	69
Português	1,54	± 0,17	48
Sueco	2,48	± 0,3	51

Fonte: Elaborada pelos autores.

Para estabelecer a significância da variável independente língua sobre o valor de *k*, recorremos à aplicação de um teste estatístico de hipótese. A amostra não cumpre o pressuposto da homogeneidade de variância, necessário para o uso de um teste paramétrico, conforme testado pelo teste Fligner-Killeen: [X^2 (6) = 17,3 $p < 0,01$]. O teste não paramétrico Kruskal-Wallis foi usado no lugar da análise de variância e indica um efeito estatisticamente significativo do fator língua sobre o valor médio de *k* [X^2 (6) = 77 $p < 0,001$]. Análise das comparações pareadas indica que o português e o inglês, as línguas com os menores valores médios de *k*, formam um grupo homogêneo. As demais línguas não se agrupam de nenhuma maneira particular. Os valores do português e do inglês são os que mais se aproximam dos valores pontuais 1,5, usado em Traunmüller e Eriksson ([s.d.]), e 1,47, sugerido por Lindh e Eriksson (2007). A maioria das médias concentra-se em uma faixa muito próxima à indicada por Traunmüller e Eriksson ([s.d.]), que vai de 1,1 a 2. Considerando o limite inferior dos intervalos de confiança em torno da média, alemão, italiano e sueco ficam acima do limiar de 2.

Na figura 8, apresentamos os valores de r^2, estimados nas dez amostras de cada falante, agrupados por língua. Os falantes estão dispostos no eixo horizontal e os valores de r^2 no eixo vertical. Quanto mais próximo de 1 é o valor de r^2, melhor é o ajuste da reta estimada por meio da técnica de regressão linear aos dados da amostra.

FIGURA 8 – Valores do coeficiente de determinação (r^2)
das análises de regressão linear em função dos falantes e da língua



Fonte: Elaborado pelos autores.

Como é possível observar, o sueco e o italiano são as línguas que apresentam as maiores proporções de falantes com valores de r^2 acima de 0,75, que indica um bom ajuste da reta estimada através da técnica de regressão linear. A inspeção conjunta das figuras 7 e 8 sugere que os falantes que apresentam valores baixos de r^2 tendem a apresentar maior variabilidade nos valores de *k*. A correlação entre o valor médio de r^2 por falante e o desvio-padrão de *k* calculado por falante é de -0,74, o que indica uma relação forte entre as duas variáveis. Uma análise de regressão simples foi usada para predizer os valores médios de r^2 com base nos valores médios de desvio-padrão de *k*. Uma equação de regressão significativa foi encontrada [$F(1, 47) = 58,7$ $p < 0,001$], com coeficiente de determinação (r^2) de 0,54. O aumento de uma unidade de desvio-padrão no valor de *k* implica redução de aproximadamente 40% no r^2 da regressão linear.

A amostra de valores de *k* foi analisada por meio da razão *F* (*F*-ratio, em inglês) de maneira semelhante à empregada por Nolan(NOLAN, 1993, 2002). O propósito é analisar a variabilidade da estimativa de *k* considerando dois pontos de vista, os falantes e as línguas, e estabelecer a relação entre a variabilidade intrafalante e interfalante, de um lado, e a variabilidade intralinguística e interlinguística, de outro. A estatística *F* expressa numericamente a razão entre a variância das médias dos falantes/línguas e a média das variâncias dos falantes/línguas. Seguimos aqui as indicações apresentadas em Nolan (2002) para o cálculo da razão *F*. A chave de interpretação do valor da razão *F* é que valores menores do que 1 indicam que a variabilidade intrafalante ou intralinguística é maior do que a variabilidade interfalante ou interlinguística.

Do ponto de vista dos falantes, a razão *F* calculada separadamente para cada língua apresenta os seguintes valores: inglês (9,364), estoniano (1,037), francês (6,169), alemão (19,111), italiano (5,631), português (3,683), sueco (13,178) e média geral 8,31. Esses resultados sugerem que a variabilidade interfalante é maior do que a variabilidade intrafalante, isto é, os diferentes falantes em cada língua do *corpus*, com a possível exceção dos falantes estonianos, variam mais entre si do que variam relativamente a si mesmos. A influência dos falantes sobre as estimativas de *k* não é surpresa, uma vez que já assinalamos que nem todos os falantes produzem dados que permitem a aplicação da técnica de regressão linear.

Do ponto de vista das línguas, a razão *F* tem o valor de 0,127, que indica que a variabilidade intralinguística das estimativas de *k* é

maior do que a variabilidade interlinguística. Esse resultado sugere que a variabilidade das estimativas de *k* é relativamente uniforme entre as línguas analisadas. Interpretamos isso como evidência de robustez, uma vez que a metodologia produz resultados similares em termos de variabilidade a despeito das diferenças existentes entre as línguas presentes no *corpus* estudado.

5 Esforço vocal e frequência fundamental

O modelo que embasa a proposição da metodologia de estimação do valor da constante *k* testado neste trabalho supõe que o nível de esforço vocal se mantenha estável e que a variação na F0 seja motivada por outros fatores. Em parte dos experimentos descritos em Traunmüller e Eriksson ([s.d.]), a simulação de diferentes graus de envolvimento ou atitude por parte de atores foi a estratégia usada para tentar obter mudanças na F_0 e controlar o nível de esforço vocal. No presente trabalho, elegemos uma estratégia para induzir variação na F_0, a mudança no estilo de elocução, que possibilita um grau de controle menor do que o uso de atores em uma situação de atuação. A variação em F_0 devida aos diferentes estilos de elocução pode interagir de forma complexa com outros fatores, entre os quais, o aumento no esforço vocal. É possível, portanto, que, em nossos dados, parte da variação observada na média e no desvio-padrão de F_0 dos diferentes estilos não seja causada por um ajuste ativo, mas seja uma consequência indireta de variações no esforço vocal. Com a finalidade de saber se os níveis de esforço vocal dos três estilos de fala presentes no *corpus* afetam a F_0, fizemos uma análise em que correlacionamos os valores do esforço vocal com os valores de média e desvio-padrão dos contornos de F_0. Para uma revisão da literatura a respeito da influência do esforço vocal sobre a F0 consultar Jessen, Koster, Gfroerer (2005).

Adotamos como medida para detectar aumento no esforço vocal a diminuição na inclinação espectral calculada com base no espectro médio de longo termo (*long-term average spectrum*, em inglês, LTAS, em forma abreviada). Para a obtenção do LTAS com base na análise dos arquivos de áudio do *corpus*, usamos o algoritmo de extração proposto em Boerma; Kovacic (2006) e implementado no Praat na função *To Ltas (pitch-corrected)*. A inclinação do espectro LTAS foi calculada relativamente a duas bandas. A inferior compreendeu valores de frequência entre 0 e 1,5 vezes o valor da F_0 média no contorno correspondente ao arquivo de

áudio, e a banda superior incluiu frequências entre aquele valor e 5 kHz. Em seguida, foram realizados separadamente testes de regressão linear simples para predizer os valores de média ou desvio-padrão da F0 com base nos valores do esforço vocal (inclinação do espectro de LTAS). Os parâmetros mais relevantes para a presente análise são a inclinação da reta de regressão e o coeficiente de determinação (r^2), isto é, a porcentagem de variância dos dados de média ou desvio-padrão explicada pela variância no esforço vocal. Os dados de falantes do sexo feminino e masculino foram analisados separadamente. Os dados das diferentes línguas foram analisados em conjunto e também separadamente.

Não foram encontradas evidências fortes nos dados do corpus entre mudanças no esforço vocal e mudanças na média de F0. A inclinação do modelo de regressão não é significativamente diferente de zero para nenhum dos dois sexos. Os valores de r^2 são 0,012 e 0,007 para o modelo dos dados dos falantes femininos e masculinos, respectivamente. A análise separada das línguas mostra que a inclinação da reta de regressão só é significativamente diferente de zero no caso dos falantes do francês – inclinação positiva de 2,5 (r^2 = 0,5) para mulheres e 1,6 (r^2 = 0,26) para homens – e do estoniano – inclinação é negativa para os falantes femininos (-4,6, r^2 = 0,21) e positiva para os masculinos (1,69, r^2 = 0,24). Esses resultados indicam que o impacto do esforço vocal sobre as mudanças na F_0 é bastante limitado e, no caso do estoniano, a influência se dá em direções opostas para falantes femininos e masculinos.

O esforço vocal influencia em alguma medida a variabilidade de F_0. O aumento no esforço vocal parece provocar aumento no desvio-padrão, mas apenas nos dados das falantes do sexo feminino. A inclinação do modelo de regressão tem o valor de 0,89 e é significativamente diferente de zero, embora o r^2 seja baixo (0,12). A análise individual das línguas mostra que, para os falantes masculinos do inglês, a inclinação da reta de regressão é significantemente diferente de zero (1,05, com r^2 = 0,59). A inclinação da reta de regressão é significativamente diferente de zero nos modelos estimados com base nas amostras das falantes do sexo feminino do estoniano (1,87, r^2 = 0,36) e do francês (1,65, r^2 = 0,6).

O francês é a língua em que o aumento no esforço vocal parece influenciar de maneira mais consistente o aumento no valor típico e a variabilidade da F_0. No estoniano, a influência atua de maneira heterogênea nas falantes do sexo feminino: níveis maiores de esforço vocal têm o efeito de abaixar a média e aumentar o desvio-padrão. Em

termos da magnitude do efeito, as inclinações da reta de regressão não nulas do ponto de vista estatístico tendem a não ser muito elevadas, concentrando-se entre 0,9 e 2,5.

Uma análise de variância de dois fatores tendo como variáveis independentes o sexo dos falantes e os estilos de fala e como variável dependente a inclinação do espectro de LTAS mostra um efeito significativo (considerando um nível de rejeição da hipótese nula de 5%) do sexo [$F(1, 204) = 14{,}1$ $p < 0{,}001$] mas não do estilo de fala [$F(2, 204) = 2{,}3$ *ns*] ou da interação entre os dois [$F(2, 204) = 2{,}3$ *ns*]. A inclinação média do espectro LTAS das falantes do sexo feminino é -10,06 dB e a dos falantes masculinos é -8,16 dB. A inclinação média dos diferentes estilos apresenta-se da seguinte forma: entrevista (-9,68 dB), leitura de frases (-9,75 dB) e leitura de palavras (-10,75 dB) para os falantes do sexo feminino e entrevista (-7,21 dB), leitura de frases (-8,44 dB), leitura de palavras (-8,82 dB) para os falantes do sexo masculino.

Os resultados, em seu conjunto, sugerem que há uma correlação entre o esforço vocal e a F_0, embora limitada a duas línguas entre as sete presentes no *corpus*. Os resultados da análise de variância, no entanto, indicam que a variação no esforço vocal é estável entre os estilos de fala. Do ponto de vista do desenho do presente experimento, esse resultado é importante, já que indica que a variação observada em F_0 entre os três estilos de fala é, em boa medida, independente da variação no esforço vocal observada nos dados. Jessen e colegas (2005) notam que os falantes podem diferir em sua resposta quando apresentados a condições que induzem o aumento no esforço vocal, e que, mesmo em casos em que há um aumento mensurável acusticamente no esforço, o impacto disso na F_0 pode ser variável entre eles. Não é possível elaborar uma explicação para a diferença significativa no nível de esforço vocal observada entre os sexos, detectada nos dados de nosso *corpus* com base nas resenhas e dados apresentados em Jessen; Koster; Gfroerer (2005).

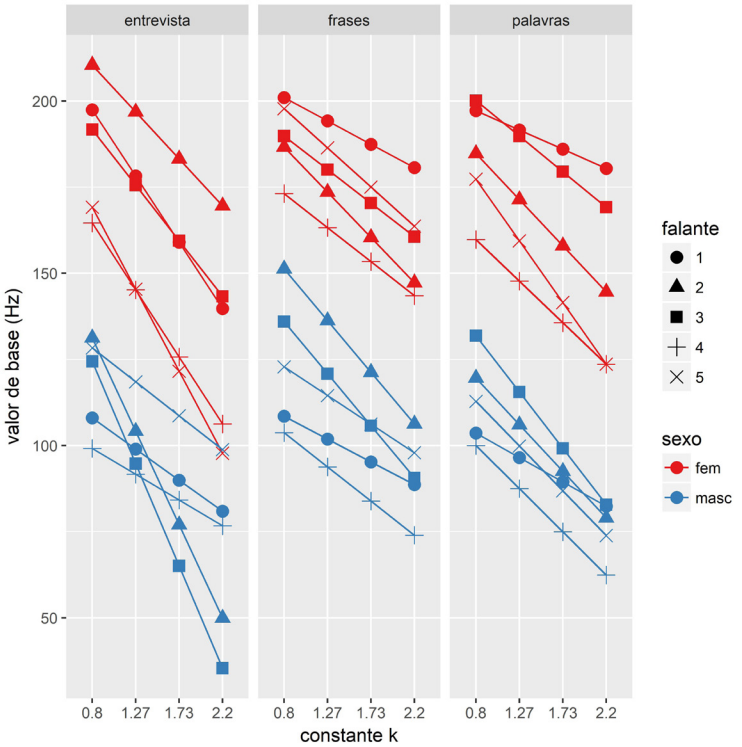
6 Sensibilidade do valor de base em relação à constante *k*

O valor estimado para a constante *k* varia entre os falantes de uma mesma língua e entre línguas diferentes. Por isso, é importante ter uma ideia da variabilidade causada na estimativa do valor de base pelo uso de diferentes valores possíveis de *k*. Para esse fim, fizemos uma simulação em que o valor de *k* foi sistematicamente variado, e o valor

de base correspondente foi calculado. Utilizamos nessa simulação os contornos de F_0 de todos os falantes e todos os estilos da amostra de dados do português brasileiro. Para cada contorno, o valor de base da F_0 foi calculado por meio da fórmula $F_b = F_{média} - k\sigma$, variando o valor de k entre 0,8 e 2,2, com passos intermediários em 1,27 e 1,73. Os valores mínimo e máximo estão próximos aos limites da faixa de variabilidade encontrada nas análises reportadas na seção 1.

A figura 9 mostra os resultados dessa variação, separados pelos estilos de fala. O sexo dos falantes é codificado pela cor, e os diferentes falantes, por símbolos diferentes. No eixo horizontal, estão os quatro valores de *k* testados e, no eixo vertical, o valor de base para cada falante, em Hertz.

FIGURA 9 – Variação do valor de base (Hz) em função do valor da constante ***k*** (formulação original)



Fonte: Elaborado pelos autores.

Replicamos o teste com a formulação alternativa do valor de base sugerida por Lindh; Eriksson (2007), que estima aquele valor como um determinado quantil da amostra de F_0. A constante *k* pode ser entendida como a indicação de quantos desvios-padrão abaixo do valor da média está localizado o valor de base. Se assumirmos que os valores de F_0 seguem uma distribuição normal centrada em zero e com desvio-padrão unitário, a função *pnorm(-k)* da linguagem de programação R retorna o valor cumulativo de probabilidade da distribuição normal compreendido no intervalo [-∞, -*k*]. Esse valor, que chamaremos de *q*, pode ser interpretado como o quantil que corresponde ao valor de base. Seguindo esse método, o valor de base foi estimado como sendo os quantis 0,01, 0,04, 0,1 e 0,21. Na formulação alternativa do valor de base, Lindh e Eriksson (2007) sugerem o uso do quantil 0,074 para a determinação do valor de base. A tabela 3 a seguir mostra os valores de *k* selecionados para a simulação e o correspondente valor de *q*.

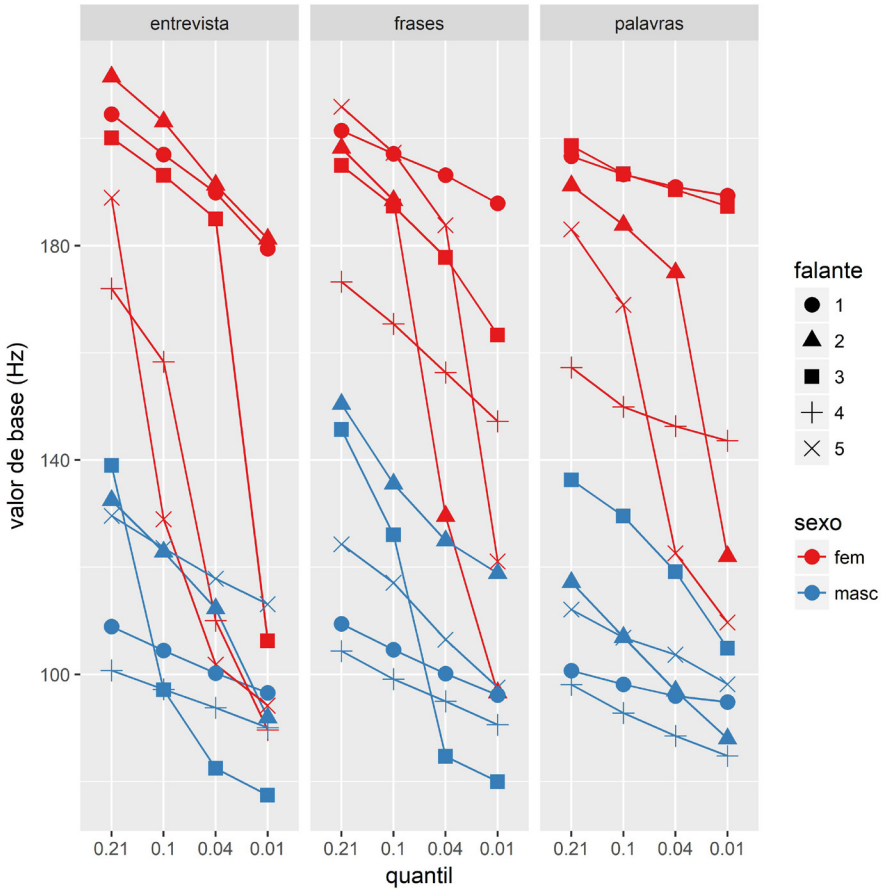
TABELA 3 – Quantis correspondentes ao valor de base e sua relação com os valores de ***k***

Valor de *k*	**Quantil (*q*) correspondente ao F_b**
0,8	0,21
1,27	0,1
1,73	0,04
2,2	0,01

Fonte: Elaborada pelos autores.

A figura 10 mostra os resultados da variação do valor de base segundo a formulação alternativa, separados pelos estilos de fala. O sexo dos falantes é codificado pela cor, e os diferentes falantes são codificados por símbolos diferentes. No eixo vertical está o valor de base, em Hertz, para cada falante, e, no eixo horizontal, os valores do quantis que correspondem à localização do valor de base (conforme mostrado na tabela 3).

FIGURA 10 – Variação do valor de base (Hz) em função do quantil (formulação alternativa)



Fonte: Elaborado pelos autores.

Para os propósitos em que o uso do valor de base pode ser mais útil, robustez não significa que os valores retornados pela fórmula sejam estritamente invariantes para um mesmo falante, mas sim a preservação das diferenças entre os valores calculados pela fórmula para os diferentes falantes.

A tendência geral, dedutível por meio da fórmula, é que quanto maior k, menor será o valor de F_b. Observe-se o painel central da figura 9, que corresponde à leitura de frases. Ali, os valores de F_b obtidos quando

k é igual a 0,8 estabelecem uma ordenação entre os falantes: f1 > f5 > f3 > f2 > f4 > m2 > m3 > m5 > m1 > m4. Apesar de haver diferenças nos valores absolutos do F_b para cada falante, a ordenação observada anteriormente permanece inalterada quando o valor de *k* sobe para 1,27; uma única alteração aparece quando *k* é igual a 1,73 (m3 = m5). Finalmente, quando *k* é igual a 2,2 há uma inversão (m5 > m3). Pelo menos no estilo leitura de frases, verificamos que o cálculo do valor de base segundo a formulação original é relativamente robusto em relação às possíveis variações de *k* no sentido que definimos anteriormente: a ordenação dos dez falantes em termos de seu valor de base fica quase inalterada, não importando qual seja o valor definido para *k*.

Considere-se, agora, para o painel central da figura 10: o mesmo tipo de análise nos leva a observar que, para parte dos falantes, a ordenação tende a permanecer estável a despeito das mudanças no quantil que corresponde ao F_b, com exceção dos falantes f2, f5 e m3, que, em algum momento, apresentam mudança brusca na passagem de um valor de quantil a outro.

Para poder quantificar o grau de robustez das duas formulações do cálculo de F_b, a original e a alternativa, além dos diferentes estilos de fala, determinamos, para cada estilo e para cada valor de *k* ou *q*, a distância euclidiana entre os valores de F_b de todos os falantes, tomados em pares. O desvio-padrão das distâncias será então tomado como um indicador de robustez, considerados os diversos agrupamentos de variáveis independentes (formulação do valor de base, estilo de fala e sexo dos falantes e os valores de *k* e *q*). Menores valores de desvio-padrão indicarão maior robustez.

TABELA 4 – Desvio-padrão (Hz) das distâncias entre o valor de base dos falantes, agrupado pelos estilos de fala

Estilo de fala	Original	Alternativa
Entrevista	35,3	36,8
Frases	30,6	32,8
Palavras	33,5	33,5

Fonte: Elaborada pelos autores.

TABELA 5 – Desvio-padrão (Hz) das distâncias entre o valor de base dos falantes, agrupado pelo sexo dos falantes

Sexo do falante	Original	Alternativa
Feminino	21,2	30,6
Masculino	18,5	14,3

Fonte: Elaborada pelos autores.

TABELA 6 – Desvio-padrão (Hz) da distância entre o valor de base dos falantes, agrupado pelos valores de ***k*** e ***q***

Valores de *k*	**DP**	**Valores de *q***	**DP**
0,8	30,4	**0,21**	32,2
1,27	30,7	**0,1**	32,5
1,73	31,3	**0,04**	33,8
2,2	33	**0,01**	33,2

Fonte: Elaborada pelos autores.

TABELA 7 – Desvio-padrão (Hz) da distância entre o valor de base dos falantes, agrupado pela interação entre estilos de fala e sexo dos falantes

Estilo de fala	**Sexo do falante**	**Original**	**Alternativa**
Entrevista	Feminino	24,8	37,2
	Masculino	20,9	13,6
Frases	Feminino	13,1	27
	Masculino	16	16
Palavras	Feminino	19	24,9
	Masculino	14	11,8

Fonte: Elaborada pelos autores.

As duas formulações parecem ter o mesmo grau de robustez quando se compara o fator estilo de fala, uma vez que o desvio-padrão das distâncias entre os falantes não varia muito em razão dessa variável. O sexo dos falantes apresentou uma relação de interação complexa: a formulação original parece ser mais robusta para as mulheres, e a alternativa, para os

homens; além disso, de forma geral as duas formulações parecem mais robustas quando aplicadas aos dados dos falantes masculinos. A Tabela 6, que mostra a interação entre estilo de fala e sexo do falante mostra que a diferença de robustez mais pronunciada entre os sexos se dá na formulação alternativa, em especial no estilo entrevista.

Observando-se a figura 10, percebe-se que alguns falantes apresentam comportamento mais discrepante em relação aos demais em termos da mudança no valor de base em razão da variação no valor do quantil associado a ele. No estilo entrevista, os falantes f3, f4, f5 e m3 têm mudanças mais abruptas. No estilo leitura de frases, os falantes f2, f5 e m3 devem ser destacados e, no estilo leitura de palavras, os falantes f2 e f5. A observação dos histogramas dos contornos produzidos por esses falantes em cada estilo indica o uso sistemático do registro não modal de vozeamento, semelhante ao padrão mostrado na figura 4. Por conta disso, quando o valor do quantil que corresponde ao valor de base assume valores mais baixos, como 0,04 ou 0,01, o F_b estimado começa a estar localizado possivelmente na região de registo não-modal, bem mais baixo do que os valores típicos do registro modal.

6 Conclusão

O principal objetivo do presente trabalho é testar a robustez da metodologia desenvolvida e apresentada por Traunmüller e Eriksson ([S.d.]) para a determinação do valor de base da F_0. O valor de base seria característico de cada falante, em tese invariante ou pelo menos bastante robusto a diversos fatores que afetam a F_0 se determinado com base em uma amostra suficientemente extensa. Na proposta dos autores, a fórmula para a determinação do valor de base depende do valor da média e desvio-padrão do falante, além de uma constante *k*, cujo valor é determinado empiricamente. No trabalho mencionado anteriormente, os autores apresentam uma metodologia para a estimação da constante, baseada na aplicação de regressão linear a dados de média e desvio-padrão de F_0. Nos experimentos descritos pelos autores, lança-se mão de fala produzida por atores, que simulam o efeito de fatores paralinguísticos, como, por exemplo, diferentes graus de envolvimento em relação aos enunciados produzidos. Esse recurso é usado para produzir enunciados idênticos do ponto de vista segmental, mas variáveis do ponto de vista da média e do desvio-padrão da F_0. Uma característica fundamental que

as amostras de F_0 precisam exibir para que a metodologia seja aplicada é proporcionalidade direta entre a variabilidade nas médias e nos desvios-padrão, isto é, que as amostras com maior média sejam também as que apresentem os maiores desvios-padrão.

No presente trabalho, testamos se o uso de diferentes estilos de elocução de fala não atuada é capaz de produzir o tipo de variabilidade na média e no desvio-padrão dos contornos de F_0 necessário para a aplicação da metodologia para estimar o valor da constante *k*. Além desse fator, testamos ainda o papel de falantes e línguas como fonte de variabilidade na estimação de *k*. Para tanto, nossa investigação analisa dados produzidos por 70 falantes de sete línguas diferentes.

Os resultados reportados aqui indicam que a estratégia de usar diferentes estilos de elocução para conseguir variabilidade na média e no desvio-padrão dos contornos de F_0 produz padrões que possibilitam a aplicação bem-sucedida da metodologia. O uso de registro não modal, bastante expressivo em termos quantitativos no caso de alguns dos falantes do *corpus*, no entanto, é um fator que parece em parte explicar os casos em que mudanças na média e no desvio-padrão não estão correlacionados. Em estudos posteriores pode ser interessante propor um critério objetivo para eliminar dos contornos os trechos de vozeamento não modal e verificar o impacto dessa eliminação nos resultados. Casos discutidos na seção 5, em que o nível de esforço vocal é uma fonte de variabilidade nos níveis médios e/ou desvio-padrão de F_0, também podem ser a razão para as incongruências que dificultam a aplicação da metodologia testada aqui.

De modo geral, os valores de *k* estimados com base nas amostras de fala não atuada são bastante próximos àqueles que os autores suecos reportam em seu trabalho e que foram derivados de amostras de fala atuada. Portanto, pode-se dizer que a técnica é robusta ao uso de fala não atuada. Os resultados apresentados na seção 4 mostram que os valores da constante *k* estimados usando a metodologia de Traunmüller e Eriksson são, em alguma medida, dependentes dos falantes. Não consideramos que essa dependência em relação aos falantes seja uma limitação severa da metodologia. Como sua aplicação depende da existência de uma dependência linear entre variação da média e do desvio-padrão de F_0, esse pressuposto precisa ser atendido. As diferenças observadas entre falantes podem ser associadas em grande parte aos casos em que a regressão linear tem um valor de r^2 baixo e ocorrem nos dados dos falantes

com maior prevalência de uso do registro não modal. A variabilidade no comportamento dos falantes pode estar relacionada com o fato de a estratégia de usar estilos de elocução diferentes para induzir mudanças na média e no desvio-padrão da F_0 não possibilitar, por seu caráter mais naturalístico, um controle tão grande da produção vocal como o que é possível conseguir por meio do uso da fala atuada.

Em termos da robustez interlinguística, os resultados indicam a existência de diferenças na média de *k* entre as línguas, que, embora significativas do ponto de vista estatístico, não são de grande extensão. O valor médio de *k* de quatro das sete línguas está dentro do intervalo [1,1 2] relatado por Traunmüller e Eriksson ([s.d.]). Além disso, os resultados da análise da razão *F* reportados na seção 4 mostram que a variabilidade interlinguística das estimativas de *k* não é maior do que a variabilidade intralinguística.

Finalmente, os resultados da simulação apresentados na seção 6 mostram que o próprio valor de base é uma medida que é bastante robusta às variações no valor da constante *k*. Dado um grupo de falantes, sua ordenação baseada no valor de base é pouco alterada pelo valor de *k* que se escolha usar. Uma vez que um dos usos mais interessantes para o valor de base é como um estimador do valor típico ou característico de um falante, essa quase invariância nas distâncias entre o valor de base dos falantes é uma propriedade interessante.

Agradecimentos

Os autores agradecem ao professor Anders Eriksson, da Universidade de Estocolmo, pela cessão do *corpus* analisado no trabalho e por discussões a respeito dos resultados. A segunda autora agradece à FAPESP pela Bolsa de Iniciação Científica (processo 2014/21161-5).

Referências

ARANTES, Pablo; LINHARES, Maria E. N. Efeito da língua, estilo de elocução e sexo do falante sobre medidas globais da frequência fundamental. *Letras de Hoje*, PUCRS, v. 52, n. 1, p. 26-39, 2017. Doi: http://dx.doi.org/10.15448/1984-7726.2017.1.25419

BOERSMA, Paul. Praat, a system for doing phonetics by computer. *Glot International*, Elsevier, v. 5, n. 9/10, p. 341-345, 2001.

BOERSMA, Paul; KOVACIC, Gordana. Spectral characteristics of three styles of Croatian folk singing. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 119, p. 1805-1816, 2006. Doi: https://doi.org/10.1121/1.2168549.

DE LOOZE, Céline; HIRST, Daniel J. The OMe (Octave-Median) scale: A natural scale for speech melody. 2014, Dublin: [s.n.], 2014. p. 910-914.

ERIKSSON, Anders. Aural/Acoustic vs. Automatic Methods in Forensic Phonetic Case Work. In: NEUSTEIN, A.; PATIL, H. A. (Org.). *Forensic Speaker Recognition: Law Enforcement and Counter-terrorism*. [S.l.]: Springer, 2011. p. 41-70.

ESKÉNAZI, Maxine. Trends in Speaking Styles Research. 1993, Berlin: ISCA, 1993. p. 501-509. Disponível em: <http://www.isca-speech.org/archive/eurospeech_1993>.

FUJISAKI; HIROSE, K. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustic Society of Japan*, Acoustical Society of Japan, v. 5, n. 4, p. 233-242, 1984. Doi: 10.1250/ast.5.233

GÅRDING, Eva. A Generative Model of Intonation. In: CUTLER, A.; LADD, D. R. (Org.). *Prosody: Models and Measurements*. Berlin: Springer-Verlag, 1983. p. 11-25.

HIRST, Daniel J. Prosodic aspects of speech and language. In: BROWN, K. (Org.). *Encyclopedia of Language and Linguistics*. [S.l.]: Elsevier Science, 2005. v. X. p. 167-178.

HIRST, Daniel J. The Analysis by Synthesis of Speech Melody: from Data to Models. *Journal of Speech Sciences*, Unicamp, v. 1, n. 1, p. 55-83, 2011.

HOLLIEN, Harry; HOLLIEN, Patricia; JONG, Gea De. Effects of three parameters on speaking fundamental frequency. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 102, n. 5, p. 2984-2992, 1997. Doi: https://doi.org/10.1121/1.420353

JASSEM, Wiktor. Normalisation of F0 curves. In: FANT, GUNNAR; TATHAM, M. A. A. (Org.). *Auditory Analysis and Perception of Speech*. London: Academic Press, 1975. p. 523-530.

JESSEN, Michael. Forensic Phonetics. *Language and Linguistics Compass*, Wiley Online Library, v. 2, n. 4, p. 671-711, 2008. Doi: 10.1111/j.1749-818X.2008.00066.x

JESSEN, Michael; KÖSTER, Olaf; GFROERER, Stefan. Influence of vocal effort on average and variability of fundamental frequency. *Speech, Language and the Law*, Equinox Publishing, v. 12, n. 2, p. 174-213, 2005. Doi: 10.1558/sll.2005.12.2.174

KENNEY, J. F.; KEEPING, E. S. Mathematics of Statistics. [s.l.]: Van Nostrand, 1962. p. 50-54.

LINDH, Jonas; ERIKSSON, Anders. Robustness of Long Time Measures of Fundamental Frequency. 2007, Antwerp, Belgium: [s.n.], 2007. p. 2025-2028.

LLISTERRI, Joaquim. Speaking styles in speech research. 1992, Dublin, Ireland: [s.n.], 1992.

MAIDMENT, J. A.; LECUMBERRI, M. L. Pitch analysis methods for cross-speaker comparison. 1996, Delaware: [s.n.], 1996.

MIXDORFF, Hansjörg. Extraction, Analysis and Synthesis of Fujisaki Model Parameters. In: HIROSE, KEIKICHI; TAO, JIANHUA (Org.). *Speech Prosody in Speech Synthesis*: Modeling and generation of prosody for high quality and flexible speech synthesis. Berlin: Springer, 2015. p. 35-47.

NOLAN, Francis. Intonation in speaker identification: an experiment on pitch alignment features. *Forensic Linguistics*, International Association for Forensic Phonetics and Acoustics, v. 9, n. 1, p. 3-21, 2002. Doi: 10.1558/sll.2002.9.1.1

NOLAN, Francis. *The Phonetic Bases of Speaker Recognition*. Cambridge, UK: Cambridge University Press, 1993.

ROSE, Philip. How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? *Speech Communication*, Elsevier, v. 10, n. 3, p. 229-247, 1991. Doi: https://doi.org/10.1016/0167-6393(91)90014-K

SCHULTZ, Tanja. Speaker Characteristics. In: MÜLLER, CHRISTIAN (Org.). *Speaker Classification I: Fundamentals, Features, and Methods*. [S.l.]: Springer, 2007. p. 47-74.

STEVENS, S. S. On the theory of scales of measurement. *Science*, American Association for the Advancement of Science, v. 103, Issue 2684, p. 677-680, Jun. 7, 1946. Doi: 10.1126/science.103.2684.677

TITZE, Ingo. *Principles of voice production*. Englewood Cliffs: Prentice Hall, 1994.

TRAUNMÜLLER, Hartmut. Conventional, biological and environmental factors in speech communication: a modulation theory. *Phonetica*, Karger Publishers v. 51, p. 170-183, 1994. Doi:10.1159/000261968

TRAUNMÜLLER, Hartmut; ERIKSSON, Anders. *The frequency range of the voice fundamental in the speech of male and female adults*. [s.d.]. Disponível em: <http://www2.ling.su.se/staff/hartmut/f0_m&f.pdf>.

VAISSIÈRE, J. Language-Independent Prosodic Features. In: CUTLER, A.; LADD, D. R. (Org.). *Prosody: Models and Measurements*. Berlin: Springer-Verlag, 1983. p. 53-66.