# Capítulo 5

# Medidas melódicas e de qualidade de voz

Após uma apresentação, na primeira seção, de sistemas de notação para a entoação *stricto sensu*, delinearemos as formas de se obterem medidas fundamentadas tanto em aspectos qualitativos quanto em aspectos quantitativos do contorno melódico em torno das funções de acento de *pitch* e marcação de fronteira, bem como da realização de diferentes estilos de elocução. O capítulo também apontará o interesse dessas medidas para a pesquisa experimental.

## 5.1 Sistemas de notação melódica

Ao longo dos anos, muitas formas de notação da curva melódica, curva primariamente relacionada à veiculação da entoação da fala, foram propostas. Segundo Crystal (1997), as primeiras formas de notação foram propostas no séc. XVIII por Joshua Steele, a partir de um sistema semelhante a notas musicais, algo próximo ao que, já no séc. XX, foi usado por Fónagy (FÓNAGY; MAGDICS, 1963) para a descrição melódica da emoção na fala e da música. Nos anos 1940, Pike (1945) propôs um sistema que considerava quatro níveis de *pitch* para o inglês, enquanto as representações icônicas de Bolinger (1986, 1989) assinalavam, a partir da década de 1960, que o sistema pikeano tinha muitas limitações, por conta da riqueza melódica em diversos contextos comunicativos.

A chamada *British School* de Crystal (1969) trabalha com a noção de "configuração" que propõe, como parte obrigatória de um

sintagma entoacional, a configuração do núcleo[1], que pode ser um movimento de descida, descida-subida ou subida em nível baixo no quadro de uma gramática entoacional que pressupõe outras configurações opcionais que ocorrem na sequência de elementos: pré-cabeça, cabeça, núcleo e cauda (TAYLOR, 1992).

Tanto Bolinger (1951) quanto Ladd (1983a) apresentaram críticas aos sistemas propostos pelas escolas americana (Pike) e britânica. O sistema americano pikeano de níveis, por ser muito restritivo, não dá conta de curvas melódicas globais ou mesmo da declinação de Fo, muito frequente nos enunciados assertivos em inglês (e em português brasileiro, PB). Por outro lado, o sistema britânico não considera questões como a recorrência de alinhamento dos acentos de *pitch* com as sílabas acentuadas, por exemplo. O sistema de Bolinger, por ser uma espécie de cópia estilizada da curva de Fo, não tem as vantagens de uma concepção enxuta e analítica de eventos melódicos que pudessem servir de norte à construção de uma economia da entoação, embora seja muito interessante para a apreciação da expressividade da fala.

Os sistemas que examinaremos aqui procuram dar conta tanto do caráter combinatório dos acentos de *pitch* quanto da importância em se respeitar o alinhamento da curva melódica com pontos singulares como as sílabas acentuadas.

### 5.1.1 O sistema ToBI de notação

O sistema ToBI (de *Tone and Break Indices*) de notação entoacional deriva dos trabalhos de Pierrehumbert e colaboradores (PIERREHUMBERT, 1980; PIERREHUMBERT; HIRSCHBERG, 1990; SILVERMAN et al., 1992) e se propõe a capturar dois aspectos prosódicos: (1) o ritmo, pelo emprego de números assinalando quatro "forças"

1 O chamado acento nuclear, o último acento de *pitch* do sintagma entoacional.

de fronteira prosódica, daí a expressão *Break Indices* da sigla (BI) e (2) os eventos melódicos de tons de fronteira e acentos de *pitch* (BECKMAN; ELAM, 1993) que explicam o termo Tones da sigla (To). Os tons de fronteira são marcados pelos símbolos L- e H- para fronteiras de sintagmas intermediários (*intermediate phrase*) e pelos símbolos L% e H% para fronteiras de sintagma entoacional *(intonational phrase*). Quanto ao acento de *pitch*, há cinco tipos, conforme a descrição que segue. Essa descrição é reproduzida do apêndice A das instruções de transcrição do ToBI (BECKMAN; ELAM, 1993) para fins de comparação com o sistema DaTo que será apresentado na próxima seção.

- H* alvo tonal que está na parte superior ou média da gama de variação de Fo do locutor no respectivo sintagma;
- L* alvo tonal que está na parte inferior da gama de variação de Fo do locutor no respectivo sintagma;
- L*+H alvo tonal na parte inferior da gama de variação de Fo do locutor em sílaba acentuada, seguido de subida pronunciada para um pico na parte superior da mesma gama de variação de Fo;
- L+H* alvo tonal alto em sílaba acentuada imediatamente precedido de subida íngreme a partir de um vale de Fo na parte inferior da gama de variação do locutor;
- H+!H* descida de tom a partir de valor elevado em sílaba não acentuada precedente.

O número e tipos de acentos de *pitch*, reiteram em mais de uma publicação os proponentes do ToBI, foram concebidos para o inglês americano[2]. O exemplo da Figura 5.1, que pode ser ouvido em

2 Por isso os sistemas de notação baseados no ToBI para outras línguas tiveram que fazer adaptações, como nos casos do G-ToBI para o alemão e o Sp-ToBI para o espanhol europeu.

**To-BImadeH**, ilustra o uso do tom H* no enunciado *Marianna made the marmalade* nas palavras proeminentes “Marianna” e “marmalade”. Observe que ambos os tons estão altos e em nível semelhante de F0.

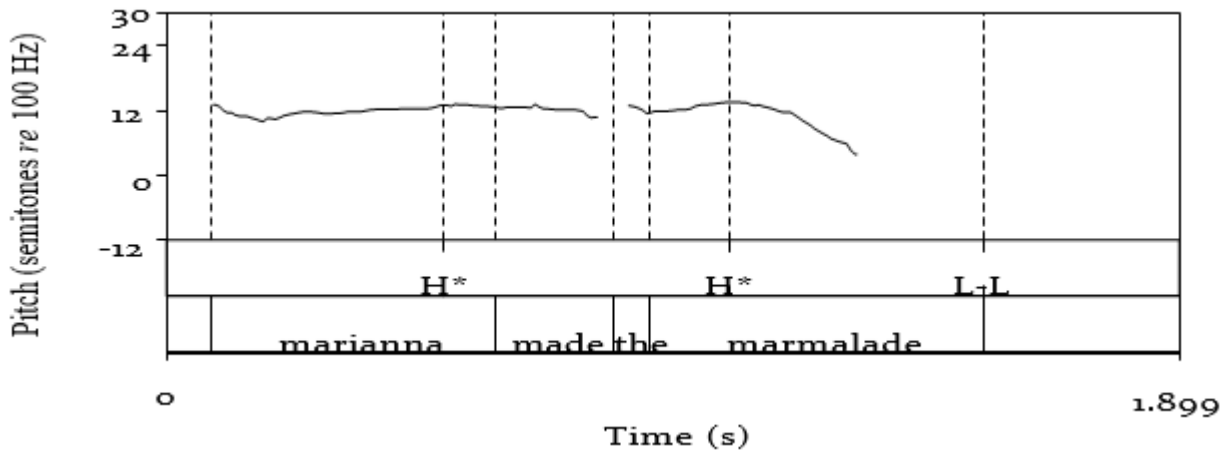


Figura 5.1 – Curva de Fo do enunciado *Marianna made the marmalade* com dois tons altos H, exemplo da oficina de aprendizado do sistema ToBI.

Ao compararmos com o evento bitonal L+H* da Figura 5.2, que pode ser ouvido em **ToBImadeLH**, no mesmo tipo de sentença, mas pronunciada de forma a assinalar foco contrativo em “Marianna”, a curva melódica da palavra em foco começa com uma subida a partir de nível baixo de F0 para atingir o pico que se vê na figura, como na instrução acima para esse tipo de acento de *pitch*.

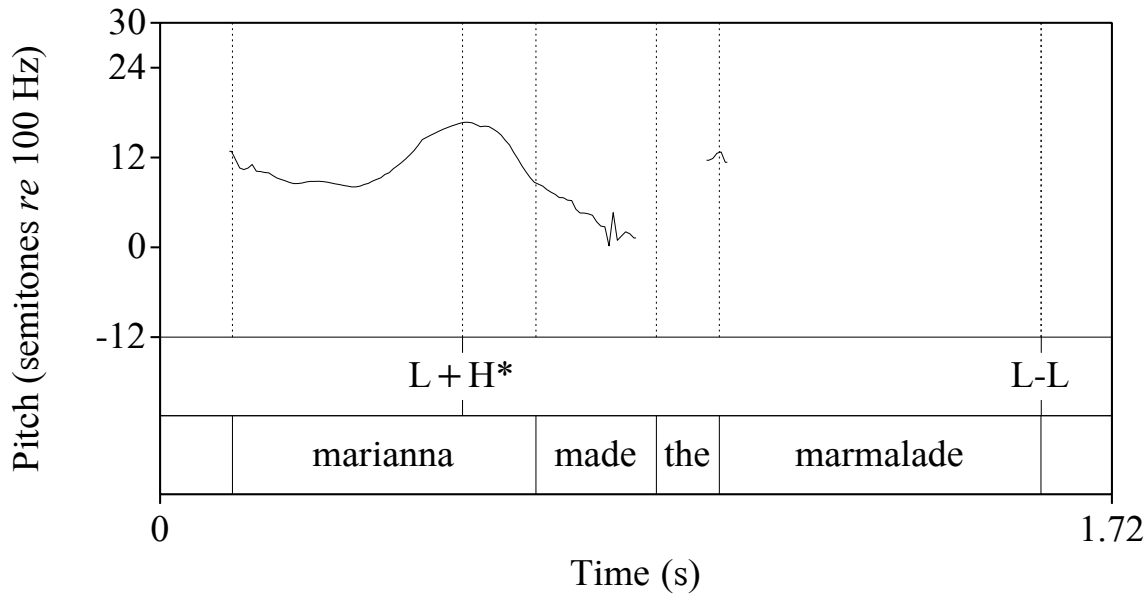


Figura 5.2 – Curva de F0 do enunciado *Marianna made the marmalade* com o evento bitonal L+H*, exemplo da oficina de aprendizado do sistema ToBI.

O último exemplo, na Figura 5.3, que pode ser ouvido em **ToBI-ma- deHH**, ilustra o uso do fenômeno de *downstep* no enunciado *Sublime mnemonic rhyme and free meter* nas palavras proeminentes “sublime”, “mnemonic”, “rhyme” and “meter”. Observe que houve uso do marcador ’!’ indicando queda do valor de Fo em relação ao nível precedente. A diferença entre os dois últimos acentos de *pitch* reside no fato de que em H+!H*, o tom alto que precede se encontra em sílaba não proeminente (“free”), o que não é o caso do tom !H* de “rhyme”, que é precedido da sílaba acentuada na palavra “mnemonic”.

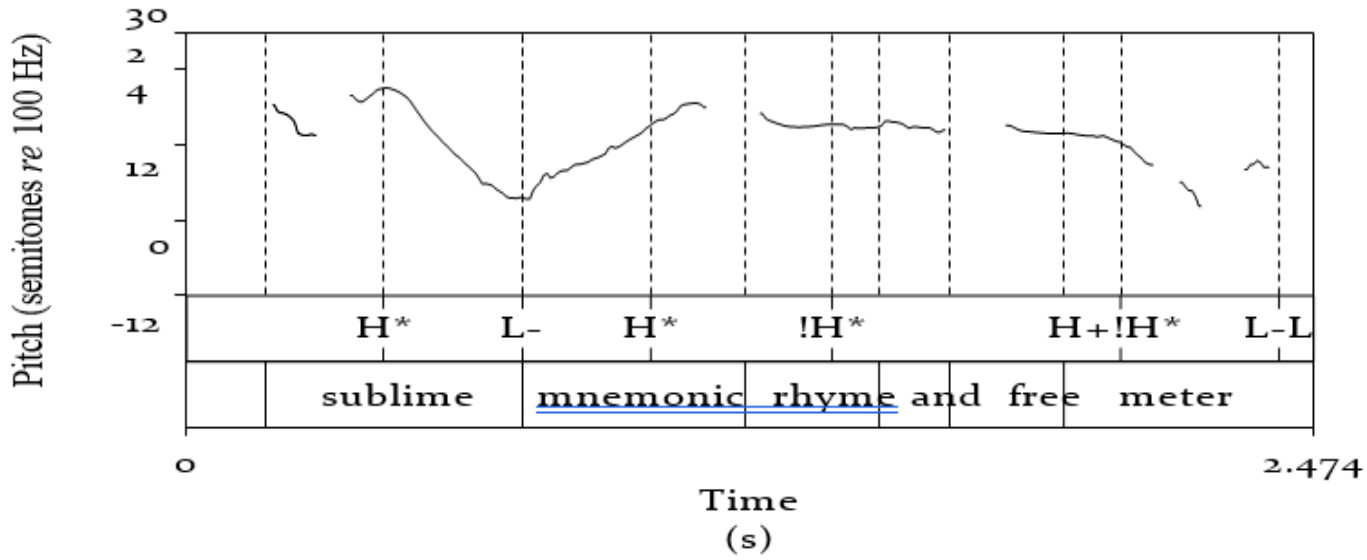


Figura 5.3 – Curva de Fo do enunciado *Ssublime mnemonic rhyme and free meter* para ilustrar o uso de *downstep* (!), exemplo da oficina de aprendizado do sistema ToBI.

Uma das principais críticas a sistemas como o ToBI veio do próprio grupo, dez anos depois de seu aparecimento (WIGHTMAN, 2002), por conta de um problema grave: o baixo acordo entre transcritores quanto ao tipo de tom a serem usados, por falta de instruções claras dos procedimentos de anotação. Esse resultado levou o autor a insistir para que se transcreva tão somente o que se "ouve" (sic), sem se guiar também pela informação dada pelos parâmetros acústicos, procurando se concentrar na função. Mas isso não vem sem trazer outros problemas, como colocou Hirst (2005) ao dizer que o que certamente essa nova proposta quer dizer se refere à interpretação do que se ouve, que diz respeito a sua função, algo que parecia não claramente separado da forma da curva melódica nos dez anos de aplicação do ToBI.

Essa mesma crítica é apresentada por Xu (2011), que assinala ainda que a forma de uma curva melódica na superfície pode estar associada a diferentes funções atuando em paralelo, como uma função de modalidade interrogativa na última palavra de um enunciado e, simultaneamente, uma função de foco na mesma palavra, que

afetam a forma do acento de *pitch* na superfície. No mesmo artigo e em outras publicações, Xu ainda aponta a necessidade de se conjugar um sistema notacional, se seu uso for realmente necessário, com abordagens de aprendizado como a dos modelos de geração da curva de Fo que vimos nas seções 2.2.3 e 2.2.4, que permitem inferir parâmetros que representam cada curva, sendo também passíveis de generalização. De fato, mostramos que é possível inferir características entoacionais a partir de abordagens fundamentadas no modelamento entoacional, como fizemos em PB (BARBOSA; MIXDORFF; MADUREIRA, 2011; BARBOSA,2016), com o modelo PENTA, e em alemão padrão (BARBOSA; MIXDORFF; MADUREIRA, 2011), com os modelos PENTA e de Fujisaki.

Antes de mostrar como combinar uma abordagem qualitativa fundamentada num sistema notacional com uma abordagem quantitativa a partir de descritores estatísticos da curva melódica, apresentamos o sistema DaTo, que se fundamenta numa abordagem dinamicista que leva em conta o ancoramento da curva melódica com pontos singulares da sílaba.

### 5.1.2 O sistema DaTo de notação melódica

O sistema DaTo de notação melódica, cuja sigla significa *Dynamic Tones* (LUCENTE; BARBOSA, 2009; LUCENTE, 2012), pressupõe a existência de um sincronismo entre a curva melódica e os movimentos articulatórios que geram os padrões espectrais, apesar de serem movimentos controlados por mecanismos distintos. Essa pressuposição o distancia do sistema ToBI, por conceber o trecho de curva melódica associado a uma determinada função como um perfil integral e não como uma composição de tons (e.g., como na notação L+H*).

O sistema DaTo não marca graus distintos de fronteiras prosódicas, deixando isso a cargo de um algoritmo semi-automático

de marcação de fronteira pela via da duração da unidade VV normalizada que foi apresentada na seção 4.2. Essa decisão também permite que o transcritor se concentre na função melódica e na descrição da forma da configuração melódica e seu alinhamento com o material linguístico. Do ponto de vista melódico, deve-se marcar apenas se a curva melódica precedendo imediatamente uma fronteira terminal ou não terminal é alta (H%) ou baixa (L%).

Antes da marcação de qualquer contorno melódico, o sistema requer que se reconheçam as palavras proeminentes e as fronteiras prosódicas do trecho sendo anotado, o que ressalta seu aspecto funcional. Tendo feito isso, então se passa a identificar o tipo de contorno ou tom. Para selecionar criteriosamente a palavra proeminente, recomenda-se que essa função de proeminência seja feita a partir de um conjunto de ouvintes leigos que assinalariam as palavras que se destacam do "fundo". As palavras assinaladas em destaque pela maioria dos ouvintes são então consideradas como proeminentes. O mesmo se faz com as fronteiras, solicitando aos ouvintes que indiquem como o locutor agrupou as palavras no trecho falado.

Quanto aos tipos de acentos de *pitch* no sistema DaTo, eles pertencem a duas classes, contornos dinâmicos, por se referirem a um movimento de subida (LH, >LH e HLH) ou de descida (HL, >HL e LHL), e tons estáticos alto (H) e baixo (L). Para todos esses contornos e tons, o aspecto crucial, que diz respeito ao sincronismo articulatório mencionado acima, é o alinhamento do pico de F0 (nos contornos ascendentes e no tom alto) ou do vale de F0 (nos contornos descendentes e no tom baixo) com a sílaba lexicalmente acentuada, bem como o alinhamento do movimento que prepara a subida ou a descida dos contornos dinâmicos com a sílaba átona precedente, como veremos adiante.

Estudos conduzidos em corpora de fala espontânea do PB (LUCENTE, 2012, 2017), sugerem que o contorno ascendente LH seja o contorno *default* do enunciado assertivo, marcando um foco

estreito. Em posição final de enunciados interrogativos esse contorno também pode aparecer, mesmo sendo mais comum a ocorrência de >LH (LUCENTE; BARBOSA, 2009).

O alinhamento do pico de F0 ao final da subida presente nos contornos LH e HLH se dá em algum ponto da vogal tônica da palavra proeminente, enquanto esse alinhamento se dá ao final ou mesmo depois da vogal tônica no contorno >LH, fazendo com que a subida desse contorno fique contida inteiramente no intervalo da vogal tônica. Essa diferença de alinhamento do pico de Fo em relação à tônica foi estudada por Kohler (2006a) em línguas como o alemão e o inglês, com o pico no início da vogal tônica sendo interpretado como marca de finalidade, no meio da tônica como abertura para novo argumento e, ao final da tônica, como no caso de >LH (*late peak* para Kohler), suscita uma interpretação de surpresa ou de algum tipo de expectativa. O número de possíveis significações desse atrasodo pico é bastante ampliado no trabalho de Ward (2019, p. 75-95), com aspectos como incredulidade, sugestão, pedido, oferta, convite, especulação, entre muitas outras. Esses estudos mostram como o alinhamento do pico ou vale (KOHLER, 2006b) com a vogal tônica é crucial para a interpretabilidade de um enunciado.

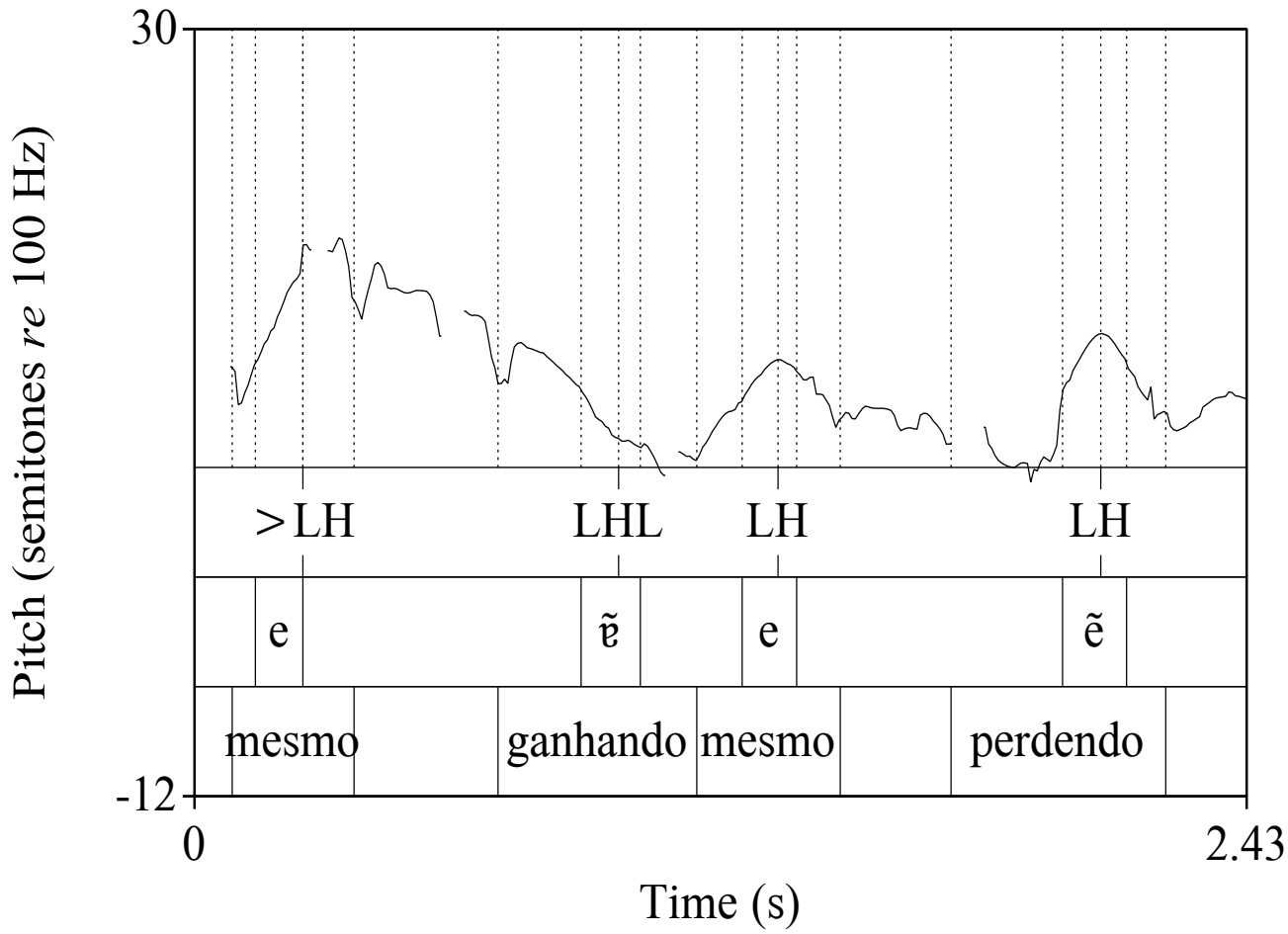


Figura 5.4 – Curva de Fo e camadas de anotação do trecho de enunciado"Mesmo o Brasil ganhando, mesmo o Brasil perdendo", ilustrando os contornos ascendentes LH e >LH. Somente as palavras proeminentes estão transcritas para facilitar a visualização. Indicam-se também os intervalos das vogais tônicas. Trata-se de uma locutora paulista durante um programa da rádio Você de Campinas.

Uma comparação entre os contornos LH e >LH pode ser vista na Figura 5.4, na qual se observa que o contorno >LH tem seu pico no extremo direito da vogal tônica de "mesmo", enquanto o contorno LH tem seu pico mais próximo ao meio da vogal tônica, especialmente a segunda ocorrência na palavra "perdendo". O contorno LHL é um contorno descendente, com descida lenta que se estabiliza na vogal tônica de "ganhando". Pode-se escutar o enunciado inteiro no arquivo **MesmoMesmo**.

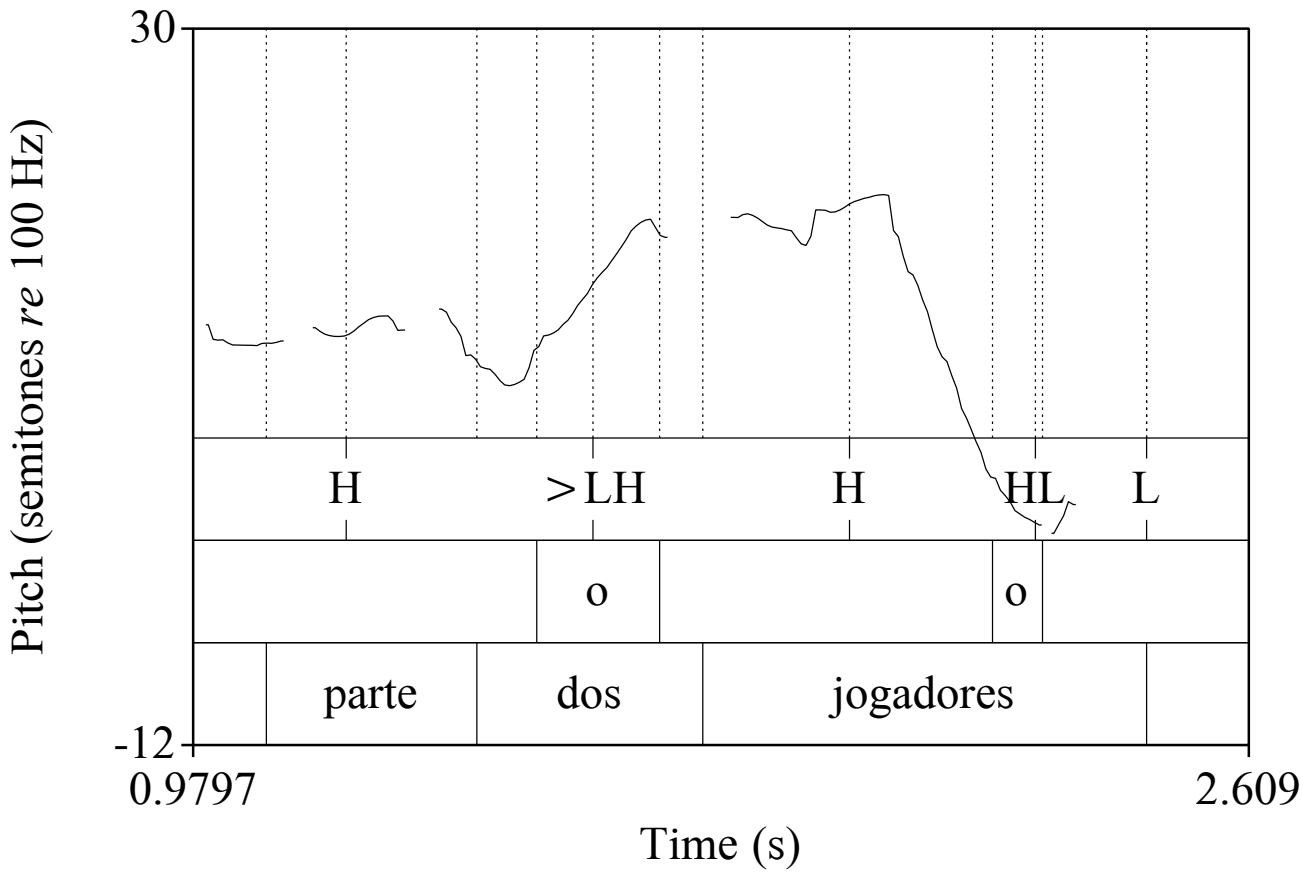


Figura 5.5 – Curva de F0 e camadas de anotação do trecho de enunciado “da parte dos jogadores”, ilustrando o contorno descendente HL, que pode ser comparado ao contorno também descendente LHL, porém com descida mais lenta da Figura 5.4. Indicam-se também os intervalos das vogais tônicas (ou proeminente, no caso de “dos”). Trata-se de uma locutora paulista durante um programa da rádio Você de Campinas.

O contorno HL, que habitualmente precede a fronteira de um enunciado assertivo, é composto por dois estágios: (1) subida da curva melódica em sílaba pré-tônica, podendo ser o clítico precedente, que culmina em pico de F0 alinhado normalmente à parte medial de vogal pré-tônica e (2) descida de F0 para alinhamento da curva em tom baixo durante a tônica. A subida precedendo a descida, de modo especular ao contorno LH, é necessária para a definição desse contorno e o diferencia do tom de nível L.

O contorno >HL tem a mesma forma que HL, mas se encontra atrasado, tendo seu vale mais à frente, fazendo com que seu pico se situe habitualmente no início da vogal tônica. O contorno LHL, por sua vez, assinala uma descida lenta de F0 própria de finais de enunciados assertivos e se alterna com HL, sendo o último mais enfático. Para

ser marcado como LHL, o contorno melódico também deve ter uma subida de F0 antecedendo sua descida.

O contraste entre as descidas dos contornos LHL e HL pode ser visto comparando-os nas Figuras 5.4 e 5.5, na qual se pode ver a rapidez com que a curva de F0 cai para atingir um mínimo ao final da vogal tônica de "jogadores". Observe no trecho que a contração "dos" foi realizada com ênfase e com contorno >LH, cujo pico está alinhado com o final da vogal.

Para apontar os aspectos abstratos da notação, a despeito de diferenças de implementação fonética, a Figura 5.6 mostra a mesma sequência de contornos empregados no mesmo enunciado da locutora paulista:

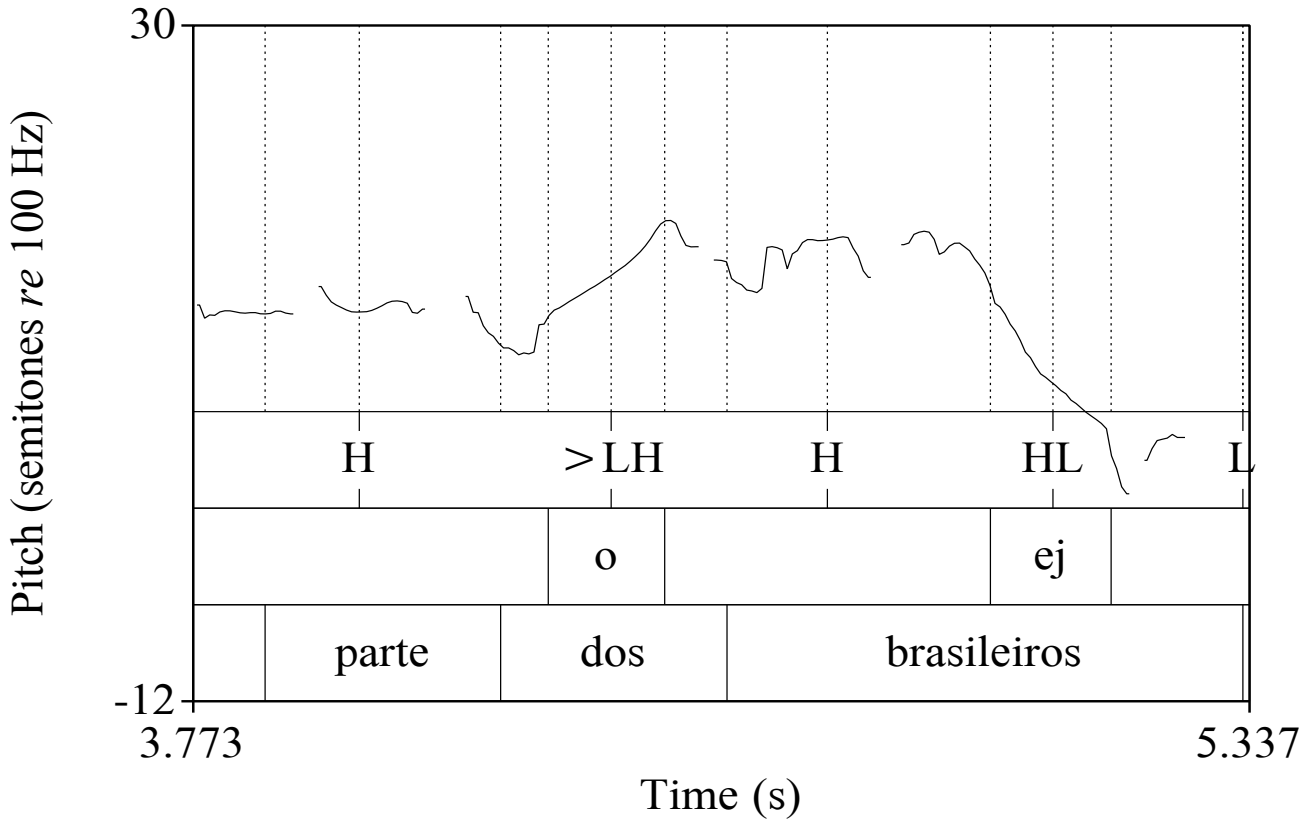


Figura 5.6 – Curva de F0 e camadas de anotação do trecho de enunciado "da parte dos brasileiros", ilustrando as regularidades do uso dos contornos, observando seu paralelismo com o uso na Figura 5.5. Indicam-se também os intervalos das vogais tônicas (ou proeminente, no caso de "dos"). Trata-se de uma locutora paulista durante um programa da rádio Você de Campinas.

"Não tenho percebido isto da parte DOS jogadores, eu te-

nho percebido isto da parte DOS brasileiros.” (contração “dos” em maiúsculas por ter sido pronunciada enfaticamente nas duas ocorrências), como se pode escutar do arquivo **Jogadores**. Nas duas sequências o perfil melódico não é exatamente o mesmo, mas a função é a mesma.

O contorno ascendente HLH, que integra uma proeminência secun dária, é ilustrado na Figura 5.7, realizado por jornalista masculino da CBN de São Paulo na palavra “JULgar”. E pode ser comparado na Figura 5.8 ao uso que ele faz desse mesmo contorno em “o governo”, com pico inicial no artigo, bem como o uso de >LH para dar ampla ênfase na palavra “toda” (o enunciado pode ser escutado no arquivo **Dinheirama**).

Uma representação esquemática dos contornos dinâmicos e sua relação com a vogal tônica podem ser vistas nas Figuras 5.9 e 5.11 para as subidas e descidas, na Figura 5.10 para o contorno HLH e na Figura 5.12 para o contorno LHL. Essas representações foram extraídas do trabalho de Lucente (2017)[3].

3 Os contornos comprimidos vHL e vLH são propostas ulteriores de Lucente (2017) para representar uma compressão da curva melódica durante a vogal (com tanto o pico quanto a descida de F0 na vogal, em vHL, e adiantamento da subida de F0 numa pré-tônica em vLH), nos casos em que, logo após a realização de uma proeminência, segue uma outra com apenas uma sílaba de intervalo. A compressão está ligada ao sincronismo entre produção segmental e realização da curva de F0 mencionado acima.

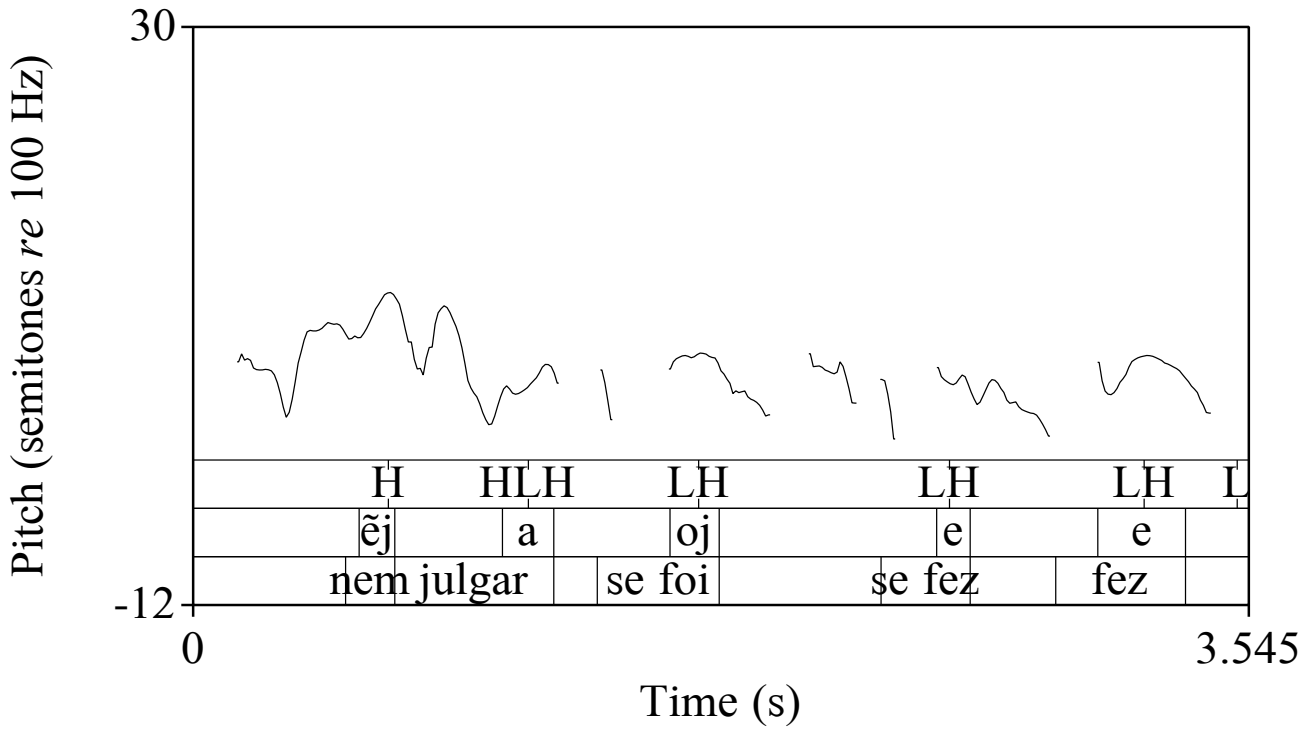


Figura 5.7 – Curva de Fo e camadas de anotação do trecho de enunciado "Não vamos nem JULgar se foi ou não foi, se fez ou não fez." para ilustrar o contorno HLH e o comparar com três instâncias do contorno LH, com a característica descida que precede a subida na partícula "se". Indicam-se os intervalos das vogais tônicas. Trata-se de um locutor paulista que comanda um programa na rádio CBN de São Paulo.

Os contornos de nível, H e L, representam alvos estáticos. Esses contornos são associados a uma proeminência que não tenha a subida obrigatória precedente dos contornos descendentes ou a descida obrigatória precedente dos contornos ascendentes. Podem aparecer acompanhados dos diacríticos ‘ !’ e ‘¡’, indicando *downstep* e *upstep*, respectivamente.

Uma forma de avaliar diferenças no emprego dos contornos anotados com o sistema DaTo é calcular a frequência relativa de cada um deles em trechos de fala. Em seu trabalho de mestrado, Freire (2020) anotou com o sistema DaTo a fala de imigrantes holandeses moradores da Holambra (SP), brasileiros e holandeses moradores da Holanda lendo uma história infantil curta. Os brasileiros não imigrantes leram a tradução da mesma para o PB, enquanto imigrantes e holandeses leram a história em holandês. É preciso ter em mente que, se os brasileiros não tiveram contato com o holandês, nem os holandeses

com o por- tuguês, os imigrantes de Holambra se consideravam holandeses e eram bilíngues ou trilíngues, falando, mesmo que de forma não simétrica, o PB, o holandês e seu dialeto original da Holanda.

O trabalho apontou uma diferença significativa entre as proporções dos contornos >LH e do tom L entre as mulheres imigrantes e as holandesas (nenhuma diferença entre homens holandeses nativos e imigrantes para essa comparação). Quanto à comparação entre imigrantes e brasileiros, o tom H mostrou diferença significativa para os homens. Os resultados desse trabalho apontaram que as mulheres imigrantes estão se aproximando da entoação das brasileiras, pelo uso mais frequente do contorno >LH, característico de marcação de proeminências no PB. Os homens, por sua vez, estão em situação intermediária entre a entoação dos holandeses e a dos brasileiros, embora tendam a se aproximar dos holandeses (FREIRE, 2020).

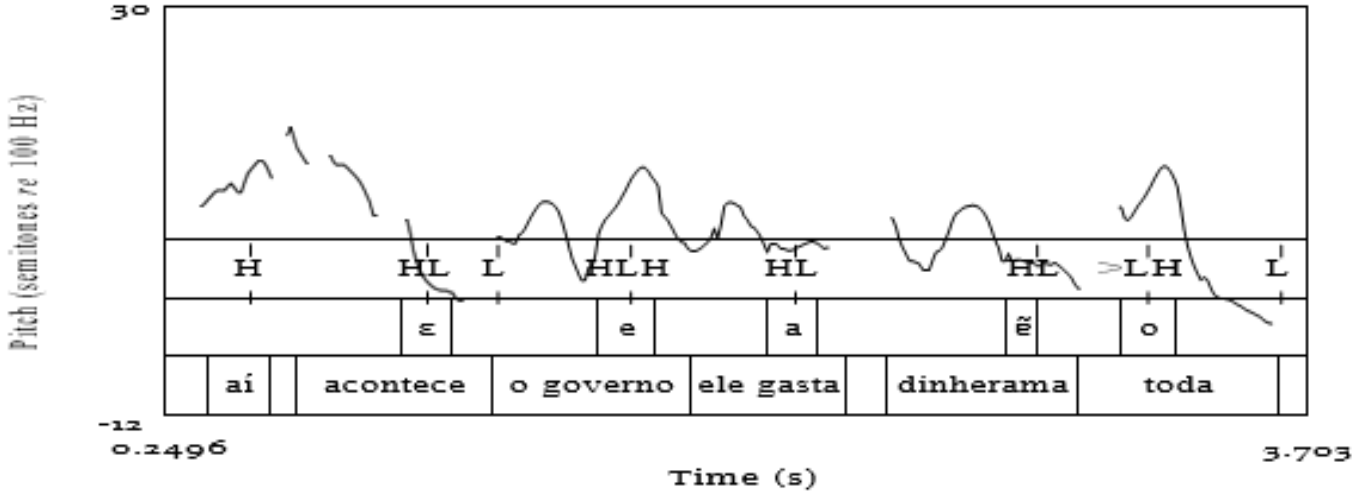


Figura 5.8 – Curva de Fo e camadas de anotação do trecho de enunciado “E aí o que acontece.... O governo, ele gasta aquela dinheirama toda...” para ilustrar a consequência da ênfase em “toda” para o perfil melódico, bem como o emprego do tom de fronteira baixo e a proeminência secundária no sintagma “o governo”. Indicam-se os intervalos das vogais tônicas. Trata-se de um locutor paulista que comanda um programa na rádio CBN de São Paulo.

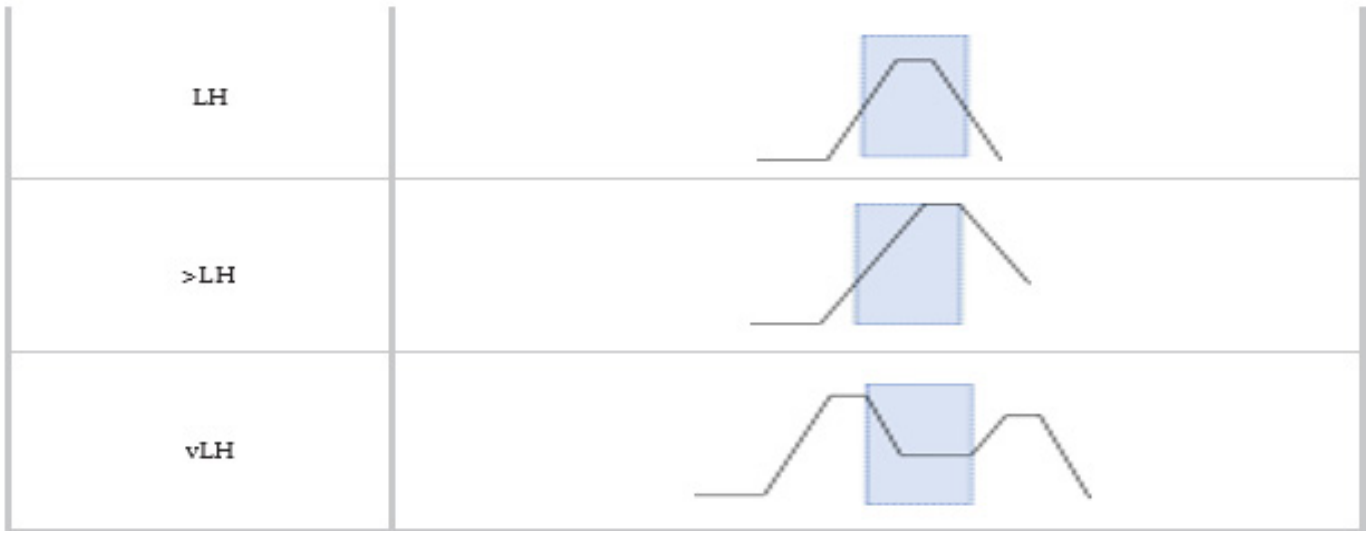


Figura 5.9 – Representação esquemática de contornos ascendentes mostrando forma e alinhamento com a vogal tônica (retângulo sombreado).

A combinação de descrições qualitativas, como essas reveladas pelos sistema de notação melódica, com medidas quantitativas relacionada à F0 fornece elementos muito ricos para a compreensão das diversas funções e usos comunicativos da entoação da fala. Passamos, assim, a apresentar descritores estatísticos de medidas melódicas para apontar diferenças entre situações comunicativas distintas.

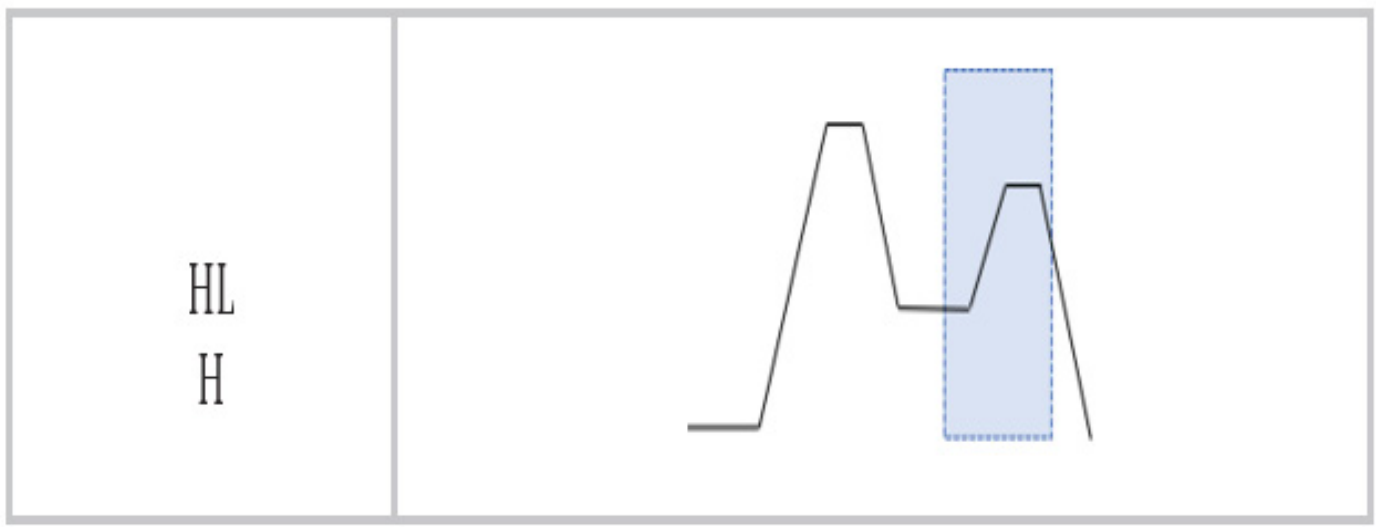


Figura 5.10 – Representação esquemática do contorno HLH mostrando forma e alinhamento com a vogal tônica (retângulo sombreado).

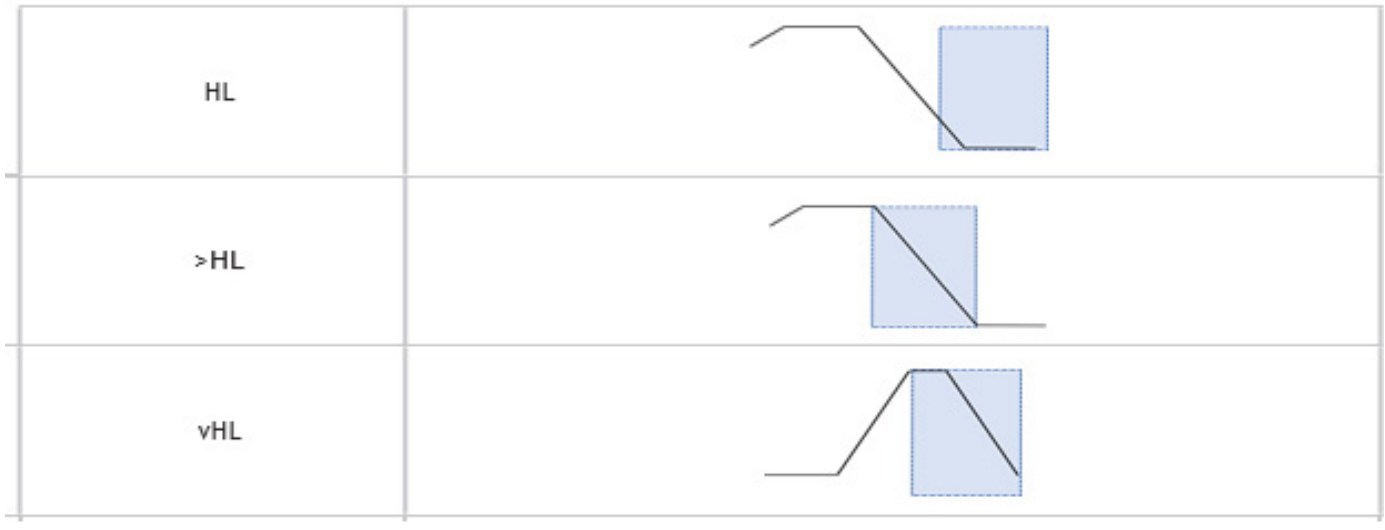


Figura 5.11 – Representação esquemática de contornos descendentes mostrando forma e alinhamento com a vogal tônica (retângulo sombreado).

## 5.2 Descritores melódicos

Os descritores estatísticos de uma variável assinalam diversos aspectos das amostras de valores da variável. Com o intuito de revelar semelhanças e diferenças entoacionais entre locutores e estilos de elocução, elencamos e explicamos abaixo o interesse de alguns descritores da F0, da primeira derivada da F0 (taxa de mudança da F0), além de outros que podem se revelar interessantes para a pesquisa experimental. Todos esses descritores permitem a realização de testes de estatística inferencial (vide seção 6.1).

### 5.2.1 Descritores de centralidade

Os descritores de centralidade de uma amostra de valores são medidas que refletem a região dos dados mais frequentes e que estão no centro de uma distribuição assumida como normal (gaussiana), por isso a referência à centralidade. O mais conhecido de todos é a média, mas há ainda a mediana. A mediana é o valor ou ponto central que divide a quantidade de dados em 50% à esquerda e à direita desse

descritor. Embora ambas revelem algo sobre a maior frequência dos valores, a mediana é mais robusta do que a média em pelo menos dois sentidos. Quando há erros de medida, a mediana continua refletindo os valores mais frequentes, enquanto a média é afetada pelo erro de medida, que pode ocorrer para determinada curva de F0. Além disso, se a amostra tem, por exemplo, uma cauda à direita, isto é, alguns poucos valores válidos de F0 bem mais altos do que os demais, a média refletirá esses valores, enquanto a mediana não, desde que os valores mais altos de F0 sejam em pequeno número.

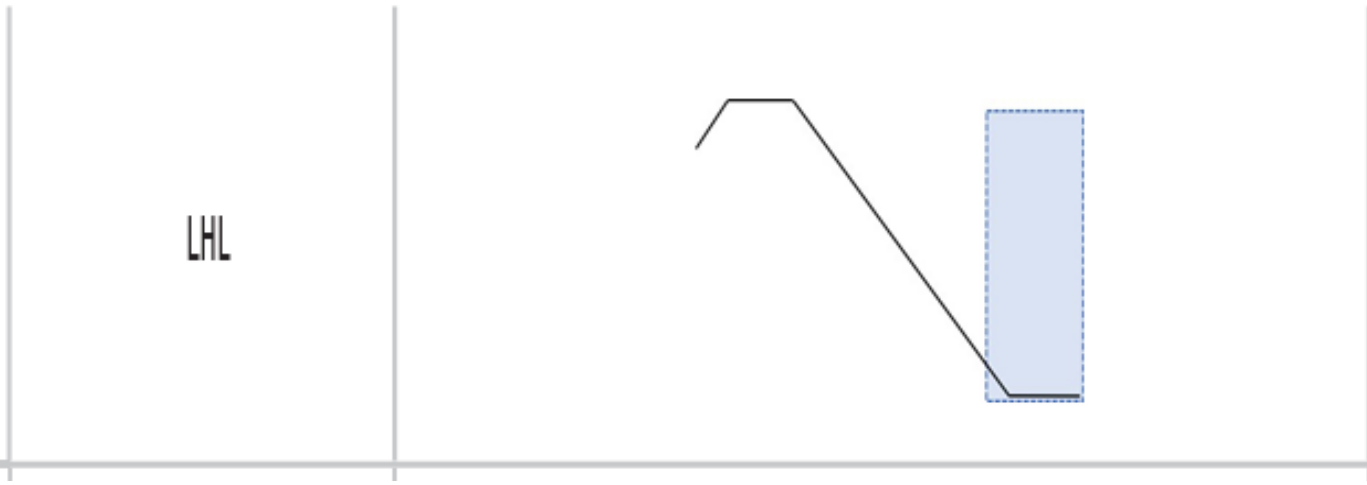


Figura 5.12 – Representação esquemática do contorno LHL mostrando forma e alinhamento com a vogal tônica (retângulo sombreado).

Para exemplificar, suponhamos que o algoritmo que calcula os valores de Fo tivesse dado a seguinte sequência de valores em Hertz (120, 127, 132, 136, 138, 140, 280), com claro erro de salto de oitava (a frequência dobra) no valor de 280. Para essa sequência, a média é 153 Hz e a mediana é de 136 Hz, pois divide exatamente à metade o número de valores à sua esquerda e à sua direita. Observe que 136 Hz é mais semelhante aos demais valores do que 153 Hz, por conta do efeito do erro de salto de oitava que entrou no cálculo da média. É sempre mais seguro, em dados sujeitos a erro, usar a mediana para estimar a centralidade da amostra.

Para ilustrar a utilidade das medidas de centralidade, observe a Figura 5.13, que mostra a curva de F0 de um homem e uma mulher lendo o trecho “Subiu a tribuna”. Os valores são expressos em Hertz (acima) e em semitons (abaixo), uma medida logarítmica de herança musical mais próxima da percepção da frequência (BARBOSA, 2019).

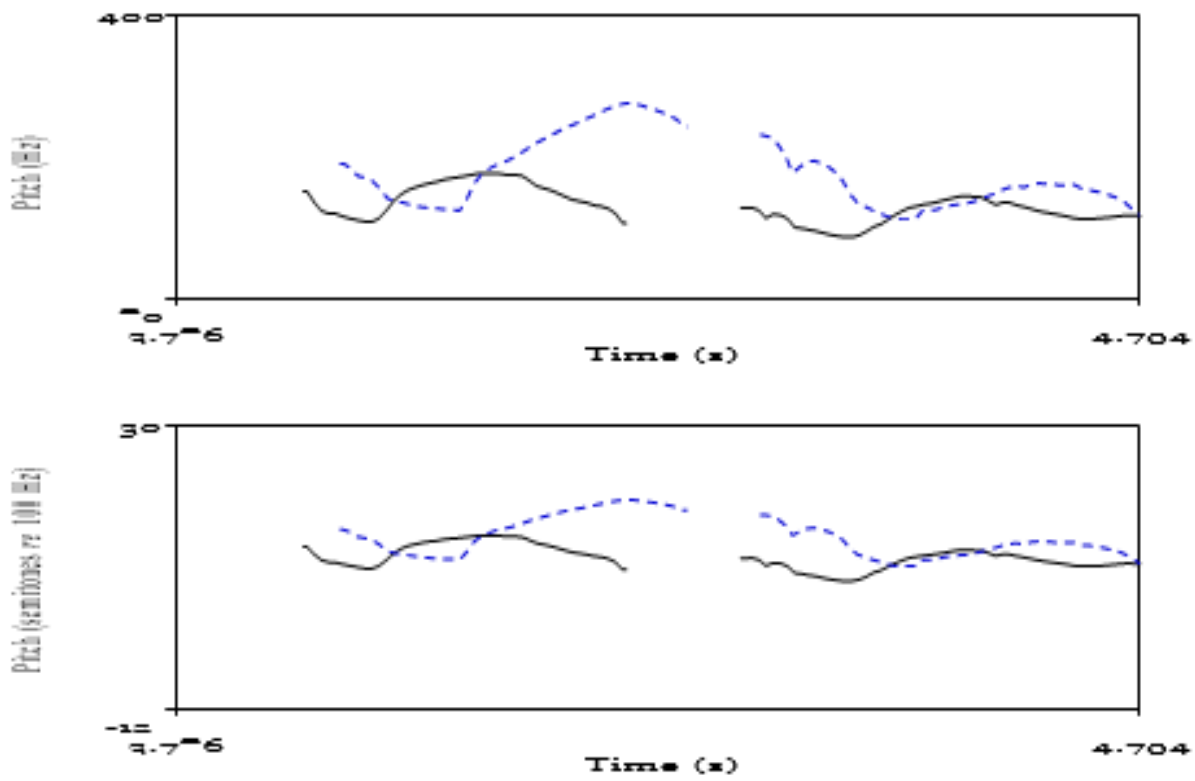


Figura 5.13 – Curva de F0 de um homem (linha cheia) e uma mulher (linha pontilhada) lendo o trecho ‘Subiu a tribuna”. Acima, em Hertz, e, abaixo, em semitons relativos a 100 Hz. Observe os valores da mulher mais altos nos dois casos.

Para o homem, a média de F0 no trecho é de 185 Hz (ou 11 semitons rel. a 100 Hz) e, na mulher, de 222 Hz (ou 14 semitons rel. a 100 Hz). Já o valor da mediana é, para o homem, de 183 Hz (ou 10 semitons relativos a 100 Hz) e para a mulher, de 210 Hz (ou 13 semitons relativos a 100 Hz). Observe que ambas as medidas de centralidade, em ambas as unidades físicas, expressam o mesmo: que o valor médio feminino é maior do que o masculino, refletindo a sensação de *pitch* mais agudo na mulher.

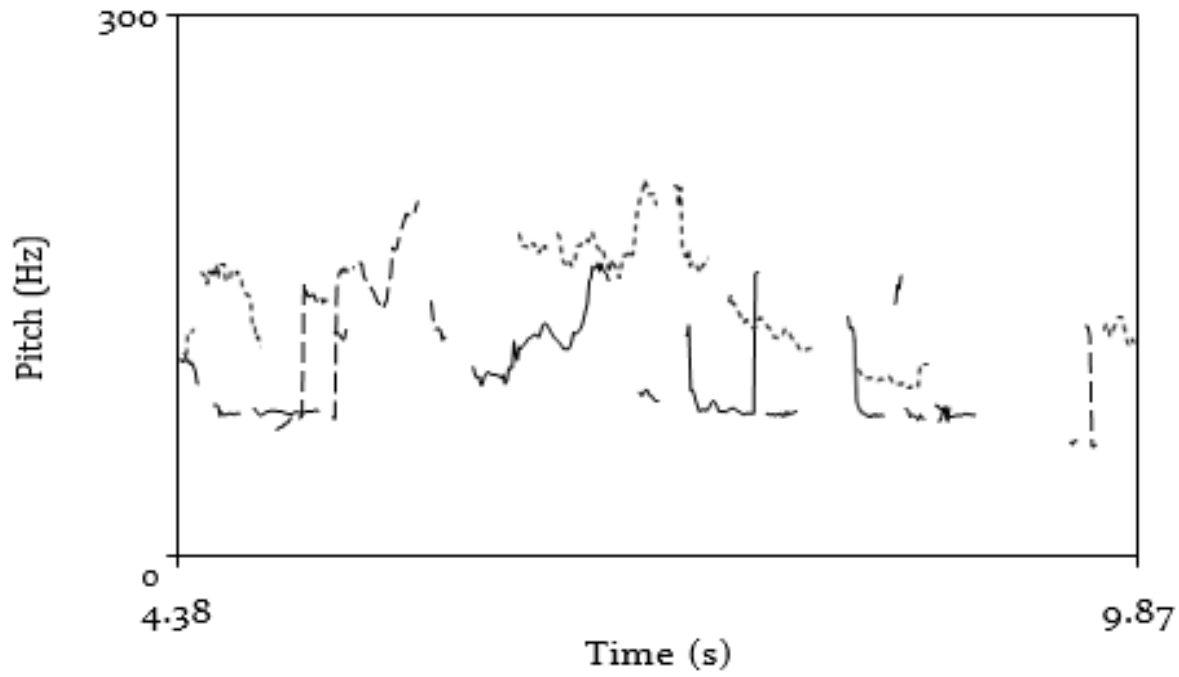


Figura 5.14 – Curva de F0 de um homem de cerca de 50 anos lendo o trecho do Primo Basílio 'Era a primeira vez que lhe escreviam aquelas sentimentalidades" numa leitura para informar (linha cheia) e outra interpretada de forma bem enfática (linha tracejada).

As medidas de centralidade podem ainda ser bem úteis para revelar a mudança global da F0 com a mudança de estilo de elocução, como se vê na Figura 5.14, considerando duas leituras, uma para informar e outra, interpretando o trecho de modo bem enfático. Essa mudança de interpretação faz com que a média passe de 97 Hz, no primeiro caso, para 137 Hz, no segundo. Há também alterações na variabilidade melódica, mas antes vejamos o exame da taxa de variação da F0, calculada sua derivada primeira. A derivada da F0 é relevante para o estudo me- lódico porque essa taxa de variação revela muito sobre a forma como o locutor realiza um acento de *pitch* ou marca uma fronteira em diferentes situações. Essa taxa tem uma relação direta com a articulação dos sons, pois eventos como acentos de *pitch* devem ter seus picos ou vales realizados na proximidade da sílaba tônica, para se fazerem mais audíveis. E para que isso se dê num determinado intervalo silábico, é preciso mudar a taxa de subida ou de descida da F0.

A Figura 5.15 retoma o traçado da F0 do homem que leu o trecho “subiu a tribuna”, acrescentando-se duas curvas: (1) uma curva de F0 suavizada com frequência de corte de 5 Hz para eliminar da primeira derivada valores bruscos sem significado fonético-linguístico, e (2) a primeira derivada de F0, obtida pelo script *fo_extrema* (ARANTES, 2008). Ao longo dessa curva, os valores acima de zero correspondem aos trechos de subida da curva de F0 e, os valores abaixo de zero, aos trechos em que a curva de F0 desce. Além disso, as taxas máximas das subidas e das descidas da curva de F0 são dadas respectivamente pelo picos e vales da derivada.

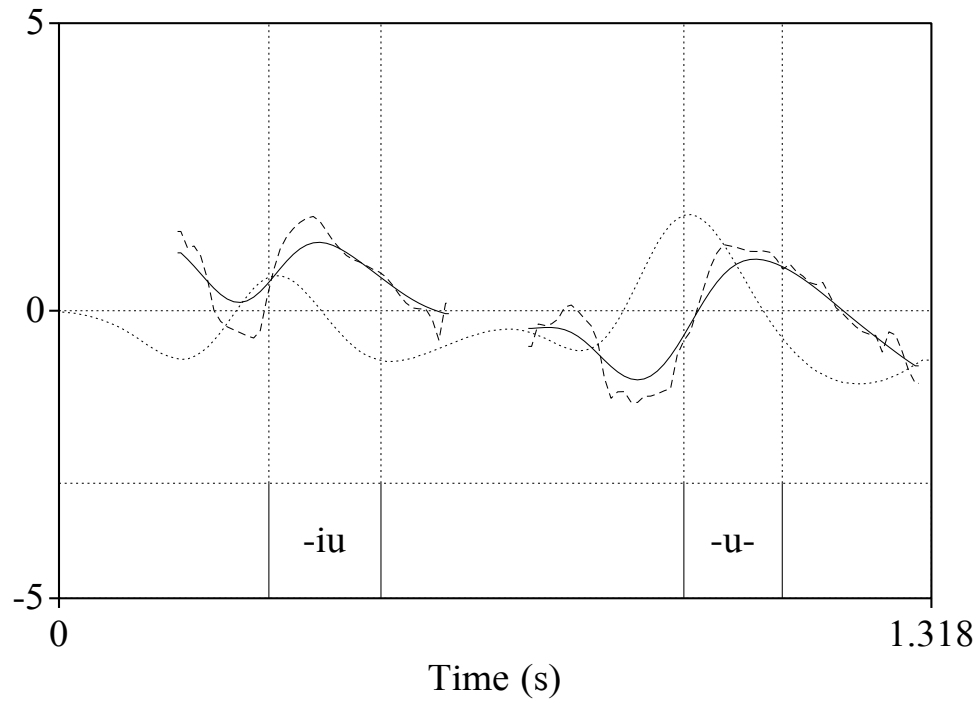


Figura 5.15 – Curva de F0 suavizada com frequência de corte de 5 Hz (linha cheia) e primeira derivada da mesma curva (linha pontilhada) de um homem que leu o trecho ‘Subiu a tribuna”. A curva tracejada é a curva original de Fo, para comparar com a suavizada.

Observe algo muito interessante: os picos das taxas de subida da curva de F0 coincidem com o início das tônicas de “subiu” e “tribuna”, cujos intervalos estão assinalados na figura. O valor máximo da taxa de subida se dá na segunda palavra lexical, “tribuna”, muito embora o pico da curva de F0 seja mais elevado na primeira palavra. A

subida rápida contribui para uma percepção de maior ênfase e para a percepção de um ritmo mais rápido, por contribuir com uma sucessão mais alta de acentos de *pitch*. Por isso, para bem estudar diferenças entre estilos de elocução e entre locutores, é importante o cálculo da média das taxas de subida, bem como das taxas de descida da curva de F0 nos trechos de fala. O script *ProsodyDescriptor*, que desenvolvemos para cálculo de parâmetros prosódicos em trechos previamente segmentados pelo pesquisador, faz esses cálculos automaticamente. Seu funcionamento é descrito no repositório em https://github.com/pabarbosa/prosody-scripts. A variabilidade tanto dessas taxas quanto dos valores da curva de F0 e os valores de seus pontos extremos, que dão a tessitura do locutor, são medidas úteis para o estudo prosódico, e são também calculadas pelo script, juntamente com os demais descritores desse capítulo, e explicados na próxima seção.

### 5.2.2 Descritores de dispersão e valores extremos

Os descritores de dispersão e os valores-limite de uma amostra de um conjunto de medidas refletem a variabilidade da medida. No caso da F0, quanto mais dispersa, menos monótono soa o trecho de fala. A mais conhecida dessas medidas é o desvio-padrão, que tem uma definição precisa, sendo a média quadrática das distâncias dos valores em relação à média.

Da mesma forma que a média, o desvio-padrão é sensível aos erros de medida e, como alternativa, se pode calcular a semi-amplitude entre quartis (SAQ), definida pela equação 5.1, em que Q1 é o primeiro quartil, o valor que divide o número de dados em 25% à esquerda e 75% à direita, e Q3 é o terceiro quartil, o valor que divide o número de dados em 75% à esquerda e 25% à direita.

$$SAQ_{F0} = \frac{Q3_{F0} - Q1_{F0}}{2} \quad (5.1)$$

Os valores mínimo e máximo de F0 de um trecho de fala definem a amplitude de variação da F0 nesse trecho, enquanto se calculada para um enunciado inteiro, esses limites definem a tessitura do locutor, seus limites superior e inferior da F0. Embora o máximo valor seja condicionado a estilo de elocução e emoção na fala, o mínimo varia bem menos com esses fatores.

Retomando o exemplo da Figura 5.13, o desvio-padrão da F0 no trecho é, para o homem, de 19 Hz (ou 2 semitons relativos a 100 Hz) e, na mulher, de 39 Hz (ou 3 semitons relativos a 100 Hz). Já o valor da SAQ, é, para o homem, de 12 Hz (ou 1,5 semitom relativos a 100 Hz) e, na mulher, de 31 Hz (ou 2,5 semitons relativos a 100 Hz). Quanto aos valores mínimo e máximo, entre 150 e 222 Hz no homem (72 Hz de amplitude de variação no trecho) e entre 171 e 301 Hz na mulher (130 Hz de amplitude de variação no trecho). Em semitons, os extremos estão entre 7 e 14 semitons no homem e entre 9 e 19 semitons na mulher.

O desvio-padrão é um descritor que pode ser usado para calcular a variabilidade das taxas de subida e de descida da F0 num trecho de fala, pois permite revelar aspectos relevantes do modo de falar de alguém numa certa circunstância de comunicação. Além dessa descrição de variabilidade da fala, outros descritores melódicos, como os que seguem, permitem revelar aspectos da vivacidade da fala.

### 5.2.3 Outros descritores melódicos

A taxa de picos (máximos locais) da F0 por segundo, desde que a curva melódica seja suavizada de forma a ressaltar os picos salientes para a percepção, está ligada ao ritmo da fala também, uma vez que

assinala a maior ou menor produção de acentos de *pitch* por unidade de tempo.

Tanto os valores desses picos locais da Fo quanto os momentos no tempo em que ocorrem podem variar, assinalando uma maior vivacidade ou criatividade do modo de falar, quanto maior for essa variabilidade. Assim, o cálculo dos desvios-padrão dos valores e dos intervalos de tempo de ocorrência de picos locais da Fo podem revelar semelhanças e diferenças melódicas.

Retomando o exemplo da Figura 5.13, há maior variabilidade na fala feminina, pois o desvio-padrão entre os dois picos da Fo é de 64 Hz na mulher e de 18 Hz no homem, significando que ela fez os dois acentos de *pitch* com valores máximos bem mais distintos que o homem.

Um outro exemplo permite observar os valores de taxas em diferentes trechos de fala. Trata-se de trecho inicial da leitura de uma fábula de Esopo em PB, "O vento sul e o sol discutiam qual dos dois era o mais forte", e em francês como língua estrangeira (FLE), *La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort* na Figura 5.16.

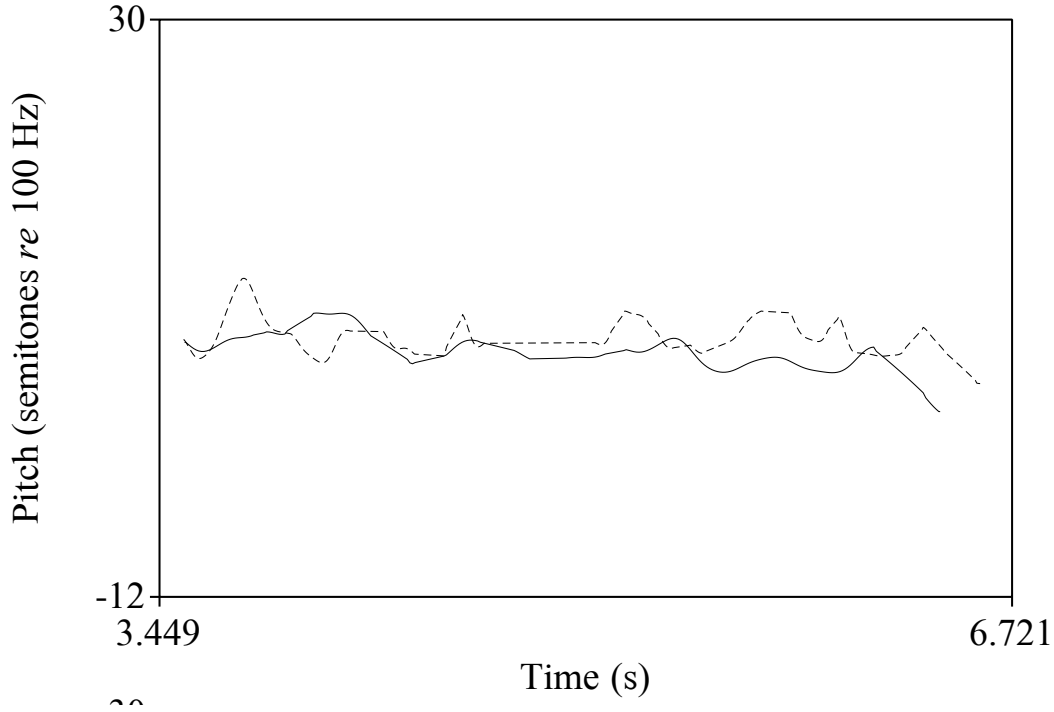


Figura 5.16 – Curva de F0 suavizada (e interpolada) com frequência de corte de 5 Hz de locutor de cerca de 20 anos de trecho de leitura de fábula em PB, "O vento sul e o sol discutiam qual dos dois era o mais forte" (linha cheia), e em francês como língua estrangeira, nível de proficiência básico, *La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort* (linha tracejada). A abscissa se refere ao tempo de leitura em PB; em francês, ela durou 2,3 segundos a mais.

O locutor tinha cerca de 20 anos no momento da leitura e tinha nível de proficiência básico na língua estrangeira. O tempo da leitura em francês está encolhido na figura para caber no mesmo gráfico, tendo sido gastos 2,3 segundos a mais para ler em francês, por conta da inserção de pausas silenciosas e uma articulação mais lenta. No trecho em PB, a taxa de picos de Fo é de 3,5 picos por segundo e, em FLE, de 1,8 picos por segundo. Além disso, o desvio-padrão temporal de sua ocorrência é de 20 ms em PB e 50 ms em FLE, atestando maior lentidão e variabilidade de incidência no tempo em FLE, o que tem a ver mais com a dificuldade de produção na língua.

Outro indicador dessa dificuldade pode ser examinado pelos contornos melódicos que usou nas duas línguas. Utilizando o sistema DaTo que vimos neste capítulo, em francês o tom de fronteira H% foi usado 72% das vezes, indicando não terminalidade de vários trechos de fala (foram 28 tons desse tipo em 39 do total de tipos de con-

torno/tom). As demais proporções importantes foram 13% de tom H para marcar proeminência e 13% de contorno >LH. Já em PB, 21% de todos os tipos de contorno/tom foram do tipo L%, marcador de terminalidade, e as proeminências foram assinaladas a 21% pelo tom H, 21% pelo contorno >LH e 29% pelo contorno HL, muito frequente ao final de trecho assertivo.

Outro descritor interessante para a descrição melódica é o grau de abertura média dos picos da F0, que guardam uma relação com carisma, como mostraram Niebuhr, Thumm e Michalsky (2018), assinalando que aberturas maiores dos picos da F0 tendem a ser associados a uma fala mais carismática, como se pode ver na Figura 5.17. Os picos, possivelmente por serem menos abruptos e talvez, considerados menos contundentes, são interpretados como uma espécie de convite.

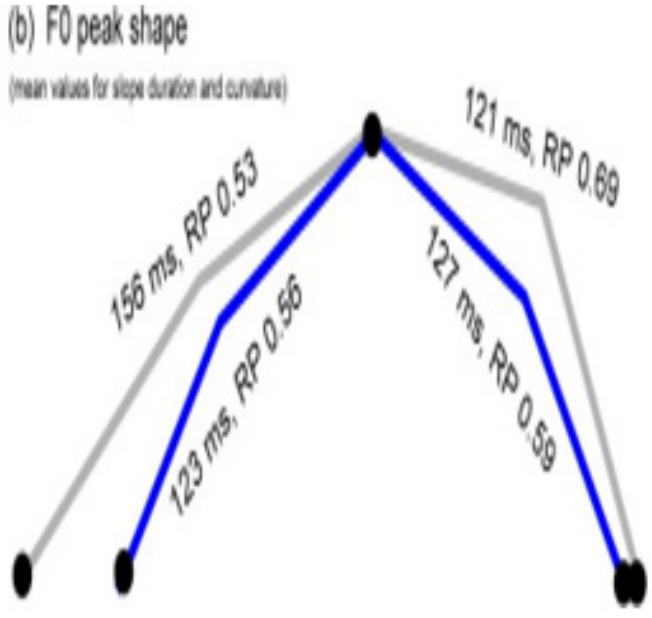


Figura 5.17 – Representação esquemática da abertura de picos da F0, sendo o mais aberto encontrado mais frequentemente em discurso de Steve Jobs, segundo Niebuhr, Thumm e Michalsky (2018). Adaptado da figura 3 do mesmo artigo. As durações são das subidas e descidas, sendo maiores em Jobs. RP é uma medida do grau de convexidade, sendo tanto mais convexa quanto maior a partir de 0,5.

### 5.2.4 Servindo-se dos descritores melódicos

No que segue, examinaremos descritores melódicos em diferentes tipos de contrastes, para ver o que revelam, em seu conjunto, sobre a entoação da fala. A Figura 5.18 mostra os valores de desvio-padrão da Fo e média das taxas de variação positivas de F0 (média dos trechos positivos da primeira derivada de F0) de uma interpretação da lenda do uirapuru por Camila Pitanga (LIS-PECTOR, 2000). O trecho lido foi dividido em trechos discursivos hierarquizados segundo a proposta de Grosz e Sidner (1986), por isso, embora os trechos apareçam na sequência como foram ditos, da esquerda para a direita, sua numeração reflete seu lugar na hierarquia temático-discursiva. O que importa aqui é a observação das maiores mudanças, que são reflexos de mudanças na interpretação em função do conteúdo.

Selecionamos para ilustrar nas figuras os parâmetros melódicos com mudanças mais bruscas para determinada passagem entre os trechos. De fato, do trecho DS10, em que se relata o lançamento de uma flecha para matar o uirapuru, para o trecho DS5, em que se introduz algo inesperado que será relatado na sequência, enquanto o desvio-padrão da F0 passa de 3 a 3,4 semitons, o valor médio das taxas de subida da curva da F0 passa de 5 a 8 Hertz/quadro, apontando que, embora os trechos tenham extensões temporais muito distintas, como se vê na Figura 5.19, o trecho 5 tem uma subida central da F0 bem mais rápida. Pode-se ver aí também que, por conta dessa maior subida, a variabilidade da F0 aumenta no trecho mais curto. Os dois trechos podem ser ouvidos no repositório do livro como **DS10Pitanga** e **DS-5Pitanga**

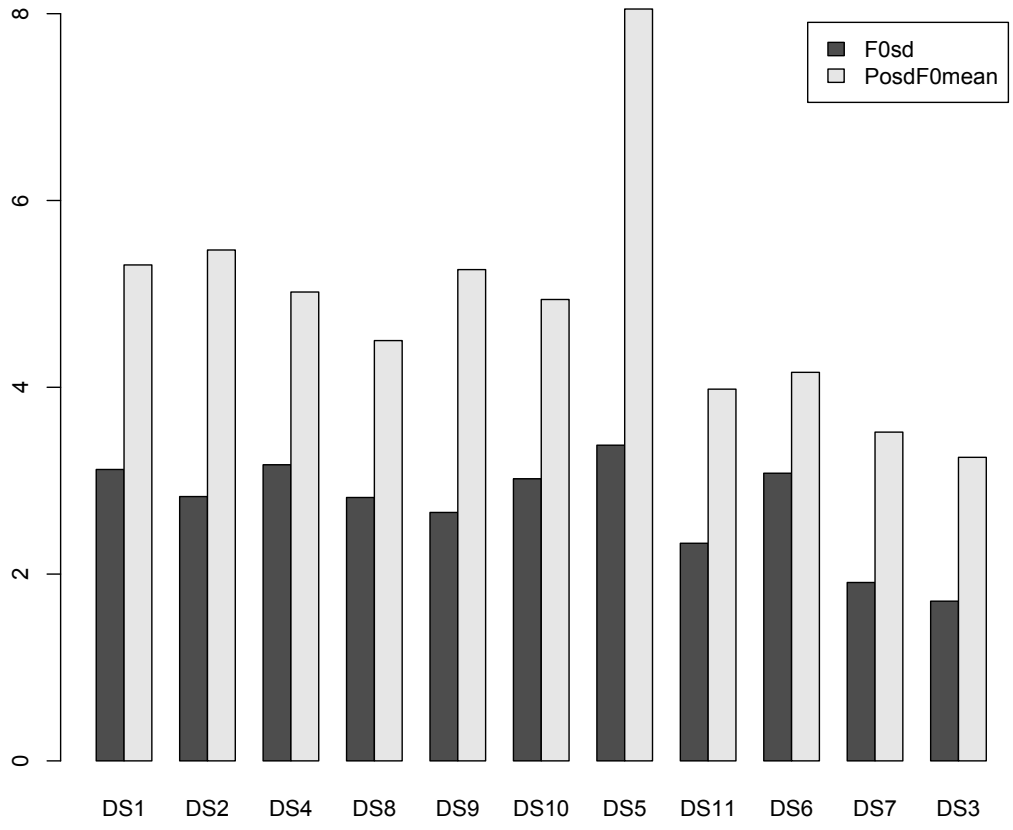


Figura 5.18 – Valores de desvio-padrão de F0 (semitons), barras escuras, e média das taxas de variação positivas de Fo (Hertz/quadro), barras claras, dos trechos discursivos da interpretação da lenda do uirapuru por Camila Pitanga.

Outros descritores melódicos que podem ser comparados estão assinalados na Figura 5.20: enquanto a taxa média de produção dos picos da F0 se mantém praticamente a mesma ao longo de toda a interpretação, a variabilidade dos valores dos picos da Fo vai particularmente aumentando do trecho DS8 até DS5, para indicar justamente, pela vivacidade que essa variação provoca na percepção, o elemento de surpresa que será relatado a partir do trecho DS11, que segue DS5 na linha temporal.

Com o fim de ilustrar a utilidade em conjugar descritores melódicos e duracionais, vistos no capítulo anterior, vamos ver agora o que acontece com a melodia e a pausa numa fala telejornalística. A locutora é uma jornalista de Campinas, cujos dados fazem parte do

trabalho de Mareüil e Barbosa (2018). Ela foi convidada a ler o texto da fábula de Esopo adaptada, "O Vento Sul e o Sol", de duas maneiras: leitura habitual (etiquetada 'normal') e, em sequência, leitura imitando uma locução telejornalística. A leitura foi dividida em 10 trechos de mesmo conteúdo que incluíram ao menos uma pausa silenciosa em um dos estilos. Em outra camada de anotação, as pausas silenciosas produzidas foram segmentadas, não havendo nenhuma pausa preenchida nas duas leituras. O script *Prosody Descriptor* permitiu o cálculo de 12 parâmetros melódicos e dois parâmetros relativos às pausas. Desses, onze parâmetros melódicos apresentaram diferença significativa entre os estilos e um parâmetro, entre os dois relativos às pausas, a duração média. Os diagramas de blocos podem ser vistos na Figura 5.21[4].

4 De cima para baixo e da esquerda para a direita as variáveis são: mediana de F0 em Hertz (f0med), máximo de F0 no trecho em Hertz (f0max), mínimo de F0 no trecho em Hertz (f0min), desvio-padrão de F0 no trecho em Hertz (f0sd), desvio-padrão de máximos de F0 no trecho em Hertz (sdf0peak), grau de abertura do pico de F0 em Hertz (f0peakwidth), desvio-padrão das posições dos picos de F0 no trecho em segundos (sdtf0peak), média da primeira derivada positiva de F0 em Hertz/quadro (df0posmean), desvio-padrão da primeira derivada positiva de F0 no trecho em Hertz/quadro (df0sdpos), média da primeira derivada negativa de F0 em Hertz/quadro (df0negmean), desvio-padrão da primeira derivada negativa de F0 no trecho em Hertz/quadro (df0sdneg) e duração de pausa silenciosa em ms (durSIL).

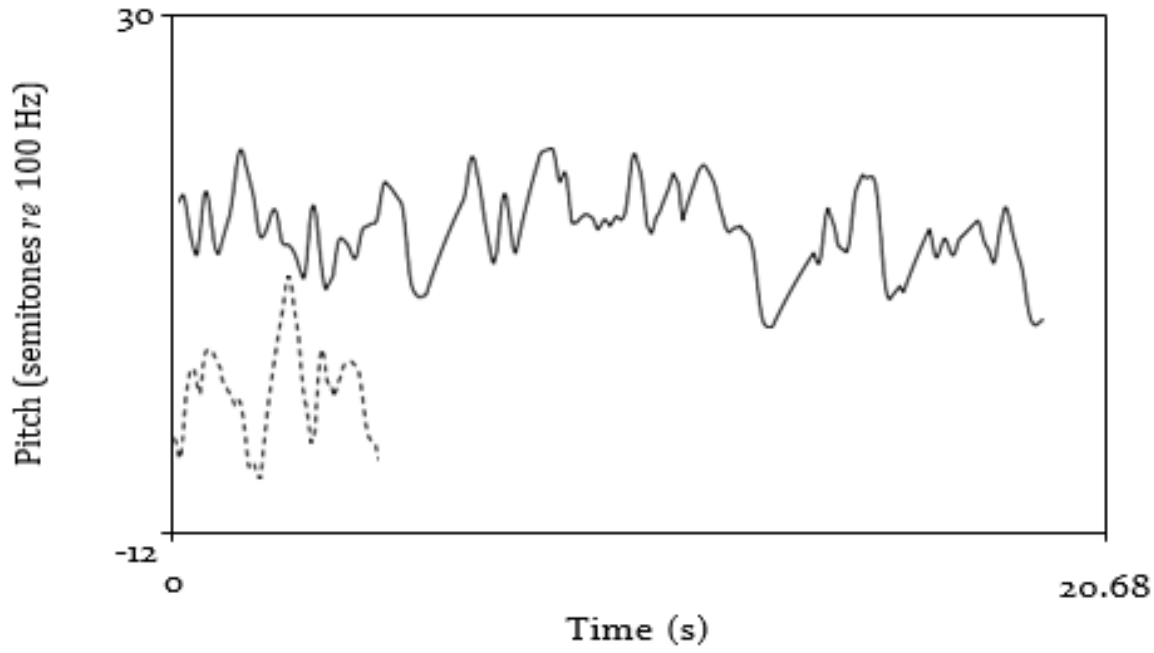


Figura 5.19 – Curvas de F0 suavizada (e interpolada) com frequência de corte de 5 Hz e traçados em semitons rel. 100 Hz dos trechos discursivos 10 (linha cheia) e 5 (linha tracejada) da interpretação da lenda do uirapuru por Camila Pitanga. Escalas verticais deslocadas para facilitar a visualização das duas curvas.

Com exceção das variáveis "desvio-padrão das posições dos picos de F0" e "mínimo da F0", as diferenças aqui encontradas são todas significativas para o nível de significância de 5% em testes de Wilcoxon e revelam o que se vê claramente na Figura 5.21: no estilo telejornalístico, os seguintes parâmetros têm valores maiores do que na leitura habitual, a saber, mediana da F0, máximo da Fo, desvio-padrão da F0, desvio-padrão dos máximos da F0, média e desvio-padrão da taxa de subida da F0, média (em módulo) e desvio-padrão da taxa de descida da F0. São menores no estilo telejornalístico o grau de abertura dos picos da F0 e a duração das pausas silenciosas. Quanto às variáveis não significativas, enquanto o mínimo de Fo deva ser investigado para ver se seria menor na fala telejornalística com mais sujeitos, a variabilidade entre os trechos do próprio desvio-padrão das posições dos picos da F0 deve também merecer um exame num maior volume de dados, pois pode estar relacionada a escolhas variadas sobre em que palavras dar ênfase.

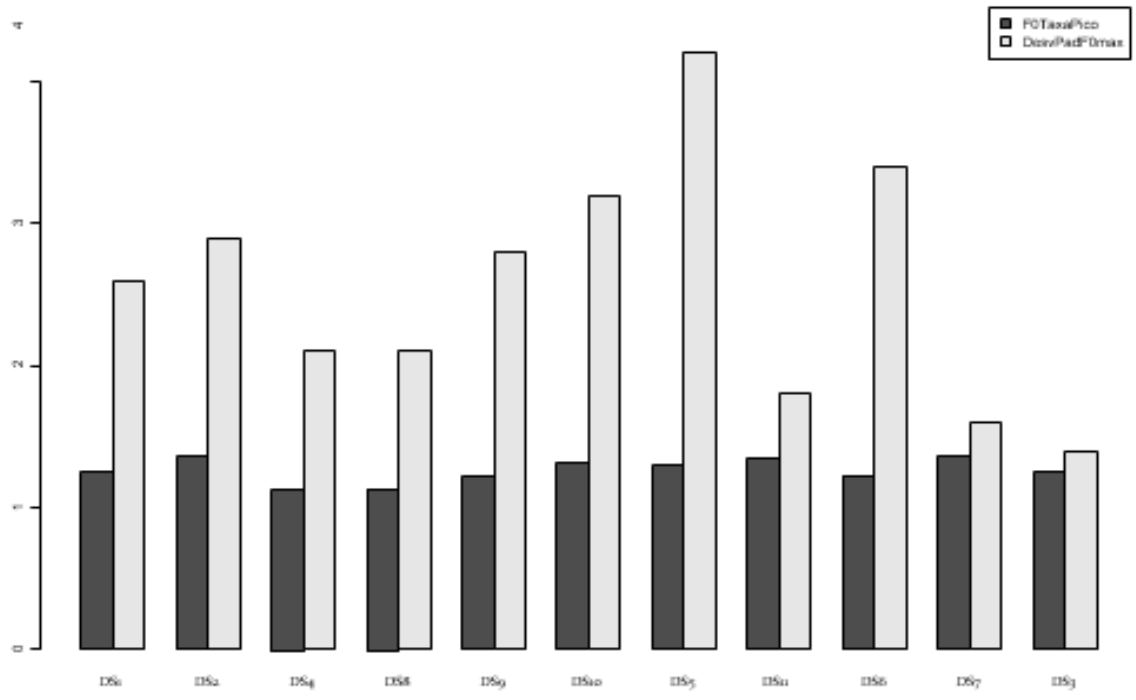


Figura 5.20 – Valores de taxa de picos locais de Fo (picos/segundo), barras escuras, e desvio-padrão de valores máximos de Fo (semitons), barras claras, dos trechos discursivos da interpretação da lenda do uirapuru por Camila Pitanga.

Do ponto de vista da percepção, esses resultados revelam que, comparada com a leitura não profissional, a fala telejornalística da jornalista é mais aguda, com entoação mais variável (evitando a monotonia e criando momentos de surpresa), subidas da curva melódica mais rápidas e variáveis, contribuindo para maior vivacidade. Ela também faz picos da F0 menos abertos, o que pode assinalar maior atratividade na fala, e as pausas são mais curtas, o que a torna mais ágil. Os áudios podem ser ouvidos no repositório nos arquivos **FalaJornal** e **FalaHabitual**.

Além de parâmetros melódicos *stricto sensu*, os parâmetros de qualidade de voz (QV) concernem à atividade vibratória das pregas vocais e, por isso, serão descritos nesse capítulo. Eles dizem respeito à alteração de longo termo do modo de fonação e a suas consequências acústicas. Para uma excelente discussão sobre o tema, ver os trabalhos de Fujimura (1988), Fujimura e Hirano (1995), Titze (2000), Esling e Harris (2005) e Kreiman e Sidtis (2011).

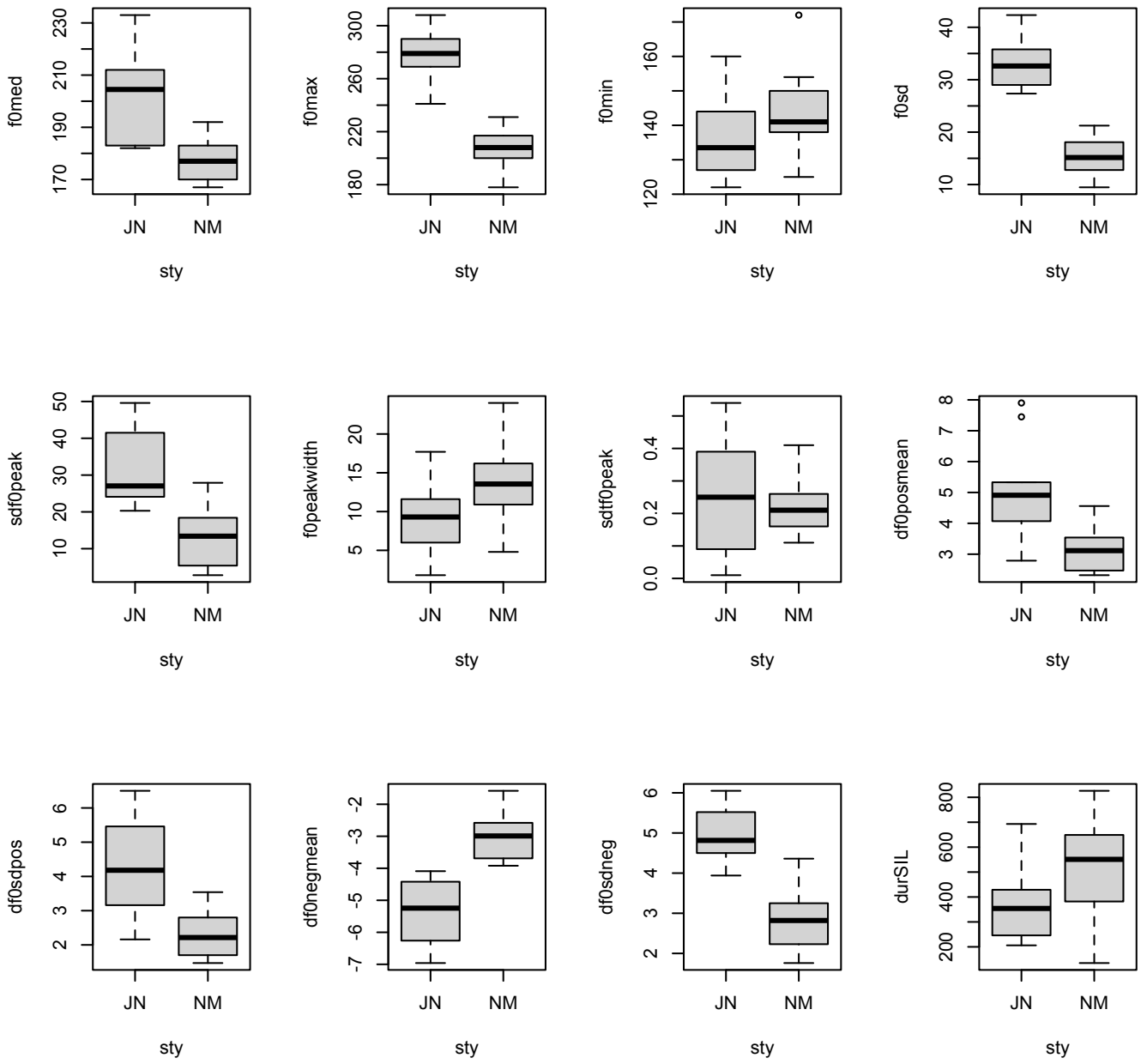


Figura 5.21 – Diagramas de blocos (*boxplots*) de onze variáveis melódicas e uma duracional da leitura em dois estilos e elocução, telejornalístico (JN) e habitual (NM).

## 5.3 Descritores acústicos de Qualidade de Voz (QV)

Os descritores acústicos da qualidade de voz (QV) são calculados a partir do sinal de fala, mas refletem direta ou indiretamente o que se passa no sinal glotal. Para uma leitura didática,

sem deixar de ser aprofundada, recomendamos o panorama dado por d'Alessandro (2006).

De interesse prosódico são as mudanças no modo de fonação, também chamada de qualidade de voz, como no caso da voz modal (*modal voice*), voz soprosa (*breathy voice*) e voz laringalizada (*creaky voice*). Em termos articulatórios, esses modos de fonação alteram o quociente de abertura do ciclo glotal (OQ, na sigla em inglês para *Open Quoti- ent)*. O OQ é a razão entre o intervalo de tempo em que as pregas vocais estão abertas em relação ao período glotal. Para a fonação modal (normal), esse quociente é cerca de 50%, enquanto para a fonação soprosa, por exemplo, o OQ é bem maior.

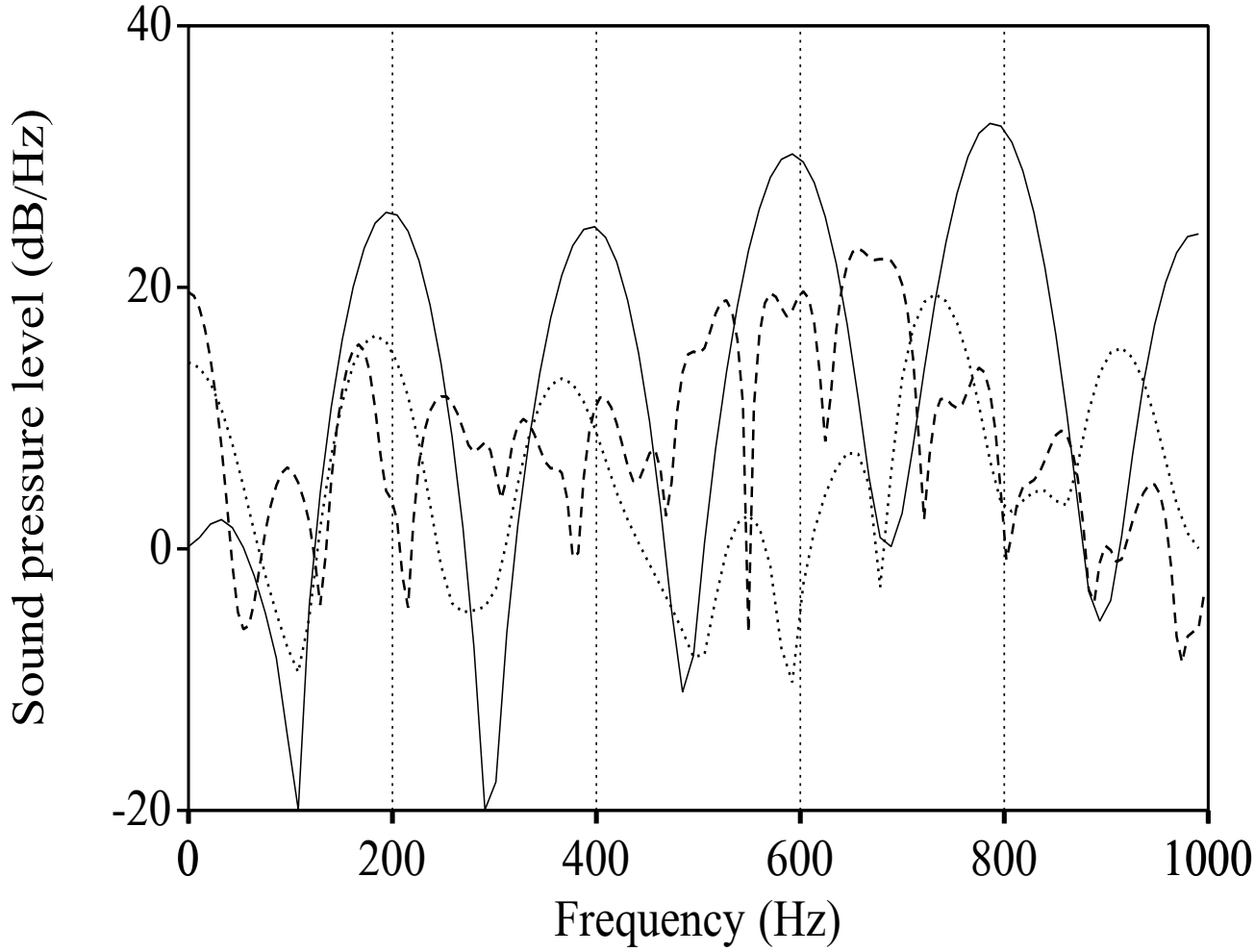


Figura 5.22 – Espectros de Fourier na região média da vogal [ɛ] da palavra "fonética" pronunciada com três modos de fonação na frase "O objeto de estudo da Fonética é essa complexa, variável e poderosa face sonora da linguagem, a fala." por uma fonoaudióloga paulista de cerca de 40 anos na época da gravação. A linha cheia se refere à fonação modal, a linha tracejada à fonação laringalizada e a linha pontilhada à fonação soprosa.

Autores como Shue, Chen e Alwan (2010) mostraram uma correlação de cerca de 65% entre as diferenças entre o primeiro (H1) e segundo harmônicos (H2) do espectro de uma vogal emitida em determinado modo de fonação com o quociente de abertura. Embora haja grande variação interindividual nesse tipo de correlação, como mostraram logo depois Kreiman et al. (2012), vale a pena o cálculo dessa medida, denominada de H1-H2, em vogais com F1 elevada (vogais abertas), para evitar o efeito do primeiro formante sobre a amplitude dos dois primeiros harmônicos.

A Figura 5.22 mostra os espectros de Fourier na região média da vogal [ɛ] da palavra "fonética" pronunciada numa frase-veículo por uma fonoaudióloga paulista. Observe que as amplitudes do primeiro harmônico nas fonações modal e soprosa são maiores do que a do segundo harmônico, mas a relação inversa se dá na fonação laringalizada. Os valores das diferenças de amplitude calculados numa janela de 50 ms centrada na vogal são de 2,0 dB (modal), 6,3 dB (soprosa) e -6,6 dB (laringalizada), o que vai na direção do que tem sido observado na literatura. Os áudios correspondentes podem ser ouvidos no repositório como **VQPBModal**, **VQPBLaringalizada** e **VQPBSoprosa**.

Uma relação no mesmo sentido, maior na fala soprosa e menor na fala laringalizada, foi encontrada no estudo de Shue, Chen e Alwan (2010), embora tenham usado a vogal [i] sustentada[5].

Uma outra medida acústica que reflete o modo de fonação é o pico de proeminência cepstral (CPP, na sigla em inglês para *Cepstral Prominence Peak* ). Como vimos em outro lugar (BARBOSA; MADUREIRA, 2015, p. 162-167), o cepstro é técnica de análise espectral que permite separar os sinais da fonte sonora do efeito de filtragem do trato vocal, permitindo observar isoladamente as características sono-

5 Embora haja técnicas para amenizar o efeito dos formantes, recalculando a amplitude dos harmônicos sem uma estimativa desses efeitos, quantidade que tem a sigla H1*-H2*, é recomendável evitar esse artifício, por isso a recomendação de escolher as vogais baixas.

ras do som laríngeo. Um ciclo glotal regular dá maiores picos de proeminência cepstral do que um ciclo irregular. Assim, vozes roucas e soprosas têm valores menores para CPP. Para as mesmas vogais [ɛ] da palavra "fonética" acima, os valores de CPP foram de 26,7 dB (modal), 13,1 dB (soprosa) e 18,1 dB (laringalizada), tendo valor máximo na voz modal e mínimo na voz soprosa, como previsto.

Medidas mais diretas da perturbação da vibração das pregas vocais são as medidas de *jitter* e *shimmer*. O *jitter* é a medida da irregularidade nos períodos glotais. Pode ser calculado ciclo a ciclo (*jitter* local) ou considerando janelas contendo 3 ou 5 ciclos glotais ou ainda a diferença média entre os ciclos numa determinada janela. Pode ser expresso em unidades de tempo ou de modo percentual. Quanto maior seu valor, mais irregular a vibração, sendo menor na voz modal e maior na voz rouca de vibração irregular. Para os trechos lidos pela fonoaudióloga que estamos usando para ilustrar as medidas, o valor do *jitter* local percentual nos três modos de fonação são 1,9% (modal), 2,1% (soprosa) e 6,2% (laringalizada), conforme esperado.

O *shimmer* é a medida da irregularidade nas amplitudes dos ciclos glotais. Pode ser calculado ciclo a ciclo (*shimmer* local) ou considerando janelas contendo 3, 5, 7 ou 11 ciclos. Pode ser expresso em dB ou de modo percentual. Quanto maior seu valor, mais irregular a amplitude vibratória, sendo menor na voz modal e maior na voz rouca de vibração irregular, especialmente com tensão laríngea. Para os trechos lidos pela fonoaudióloga, o valor do *shimmer* local percentual nos três modos de fonação foram: 9,1% (modal), 13,4% (soprosa) e 22,9% (laringalizada). Observe que tanto para o *jitter* quanto para o *shimmer* os maiores valores ocorrem na voz laringalizada, que foi feita de forma bem tensa em sua fala. A voz com vibração mais regular das pregas vocais é mesmo a modal, como esperado.

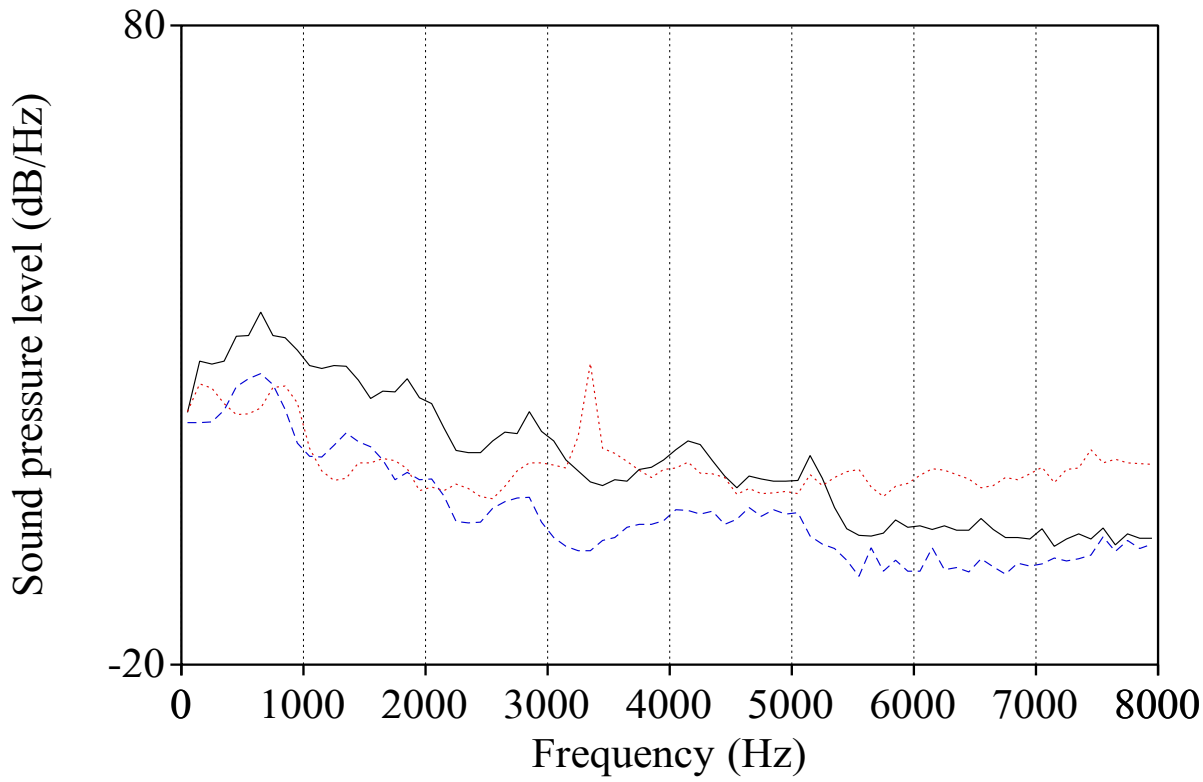


Figura 5.23 – Espectros médios de longo termo (LTAS) da frase "O objeto de estudo da Fonética é essa complexa, variável e poderosa face sonora da linguagem, a fala." pronunciada em três modos de fonação por uma fonoaudióloga paulista de cerca de 40 anos na época da gravação. A linha cheia se refere à fonação modal, a linha tracejada azul à fonação laringalizada e a linha pontilhada vermelha à fonação soprosa. Observe as diferenças de energias de 0 a 1000 Hz e depois nas faixas 1000 a 4000 Hz.

Os dois próximos descritores se referem ao descompasso em energia entre diferentes bandas espectrais. São a ênfase espectral e a inclinação do espectro de longo termo (LTAS, na sua sigla em inglês, *Long-Term Average Spectrum*). A ênfase espectral (*spectral emphasis*, em inglês) foi definida por Traunmüller e Eriksson (2000) como uma medida acústica indireta do esforço vocal. De forma simplificada, pode ser definida como a diferença entre a intensidade total de um som e a intensidade numa faixa de frequência baixa para englobar toda variação da frequência fundamental, segundo a equação 5.2.

$$\text{Ênfase espectral} = I - I_0 \qquad (5.2)$$

Em que I é a intensidade até a frequência máxima do sinal e $I_0$

é a intensidade do som de 0 a 400 Hz, ambas em dB, com o limiar da banda baixa de 400 Hz fixado para melhor operacionalizar os cálculos[6]. Na prática, uma vez que os sinais de fala são armazenados de forma digital, a frequência máxima disponível para análise é a frequência de Nyquist, que é a metade da frequência de amostragem. Os valores de ênfase espectral para os trechos ilustrados aqui são de: 9,5 dB (modal), 6,6 dB (soprosa) e 8,1 dB (laringalizada), com esforço maior na fala modal.

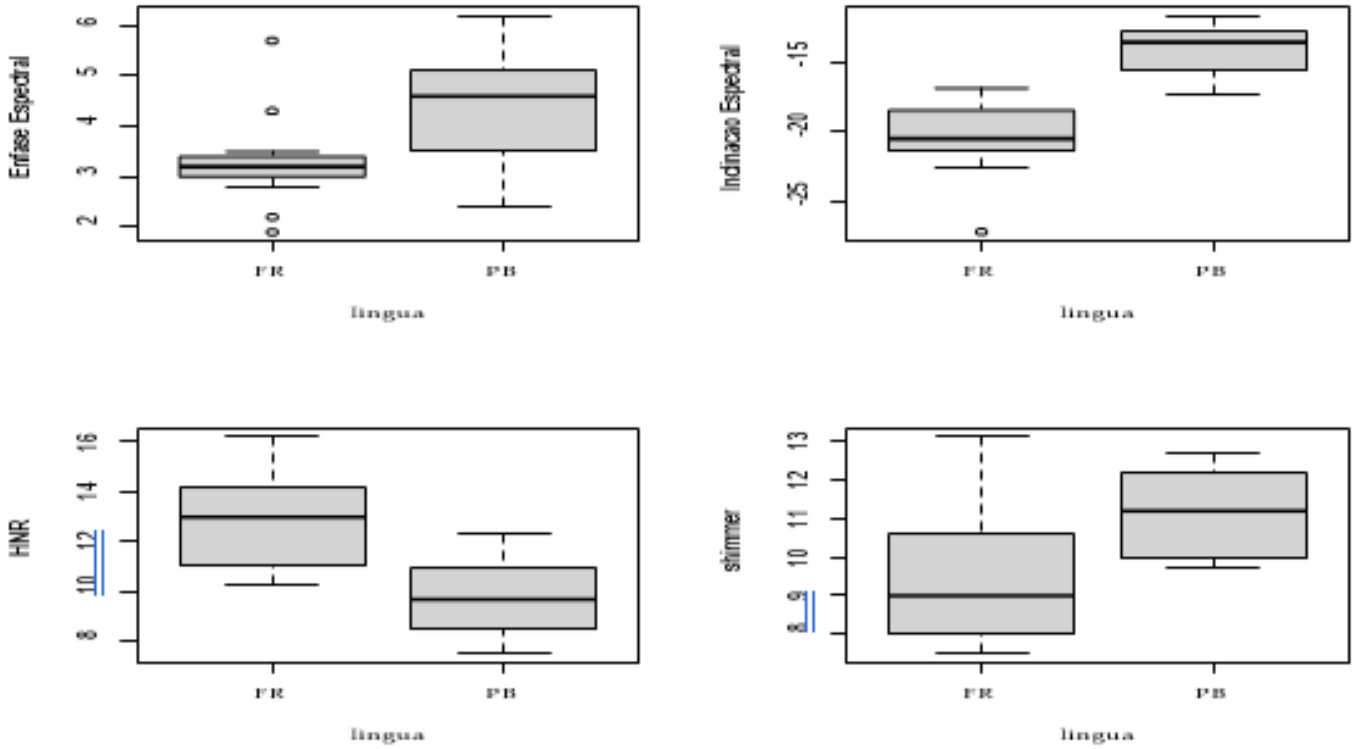


Figura 5.24 – Diagramas de blocos de medidas de QV para estudante de francês em nível básico em leituras em PB e em francês (FR) de fábula de Esopo. As medidas são, de cima para baixo e da esquerda para a direita: ênfase espectral, inclinação espectral, HNR, as três em dB, e *shimmer* em porcentagem.

A inclinação do espectro de longo termo se refere à diferença de energia média entre duas bandas de frequência desse mesmo espetro. Quanto menos inclinado for esse espectro, mais energia se encontra na banda de mais alta frequência, que é um reflexo da produção de

6 Os autores definiram esse limiar exatamente como 1,43 x F0 médio no trecho, isto é, 43% acima da frequência fundamental média no trecho, mas também testaram com valores de 50% e valores fixos como os 400 Hz sugeridos aqui. Os resultados de correlação com o esforço vocal foram praticamente os mesmos.

componentes de frequências altas com amplitude elevada, uma consequência seja de maior esforço vocal, seja de maior soprosidade. Uma medida muito usada da inclinação é a diferença de energia entre a banda de 1000 a 4000 Hz e a banda de 0 a 1000 Hz. Observe na Figura 5.23 que essa diferença é menor na fala soprosa. De fato os valores para o cálculo sobre toda a frase são: -9,6 dB (modal), -7,1 dB (soprosa) e -12,6 dB (laringalizada), revelando a maior energia produzida na banda de 1000 a 4000 Hz por conta da soprosidade.

Outra medida acústica muito útil de qualidade de voz é a razão harmônico-ruído (HNR, da sigla do nome em inglês, *Harmonic to Noise Ratio*), que mede a relação, em dB, da energia dos harmônicos num trecho de fala e e energia do ruído, em toda a faixa espectral. Diferentemente das medidas de ênfase e inclinação espectrais, que consideram bandas diferentes e não separam o que é devido apenas à energia de ruído do que é devido apenas à energia harmônica, a HNR faz isso. Os valores de HNR para os trechos ilustrados aqui são de: 10,2 dB (modal), 7,6 (soprosa) e 2,3 dB (laringalizada), revelando que há bastante ruído na fala laringalizada, afetando todo o espectro de fala. O efeito de ruído em baixa frequência nessa fala não é considerado de forma separada nas medidas de ênfase espectral e inclinação espectral, por isso a diferença com os dois outros modos de fonação não são tão grandes quanto a mostrada com essa medida.

Retomando o exame da fala de um estudante de francês cuja leitura em PB e em francês usamos na seção 5.2.3 para ilustrar a importância de se usar outros descritores melódicos, observa-se em sua fala em língua materna (PB) uma grande frequência de laringalização, como o leitor pode conferir nos áudios nas duas línguas, **MCFR** e **MCPB**. Essa laringalização é acompanhada de maior tensão laríngea e possíveis irregularidades de vibração das pregas vocais. De fato, os diagramas de bloco dos quatro descritores de QV mostrados na Figura 5.24 apontam para essa conclusão para sua leitura em PB, a qual exibe maior ênfase espectral, um espectro de longo termo menos inclinado,

um HNR menor que revela mais ruído na fala e um *shimmer* mais alto, apontando para maior irregularidade de amplitude do ciclo glotal.

Parâmetros melódicos e de qualidade de voz são muito usados para compor personagens na indústria de entretenimento. Os diagramas de bloco da Figura 5.25 mostram claramente os recursos empregados por Camila Pitanga para interpretar três personagens da história de Pedro Malazarte: o personagem principal tem uma fala mais aguda com melodia mais variável e ainda menos ruidosa (maior valor de HNR) e com ciclos glotais mais regulares do que a fala dos outros dois personagens, o marido e sua mulher. Esse dois outros personagens são diferenciados sobretudo pela menor variabilidade melódica na mulher (maior valor de desvio-padrão da F0) e a fala mais ruidosa do marido.

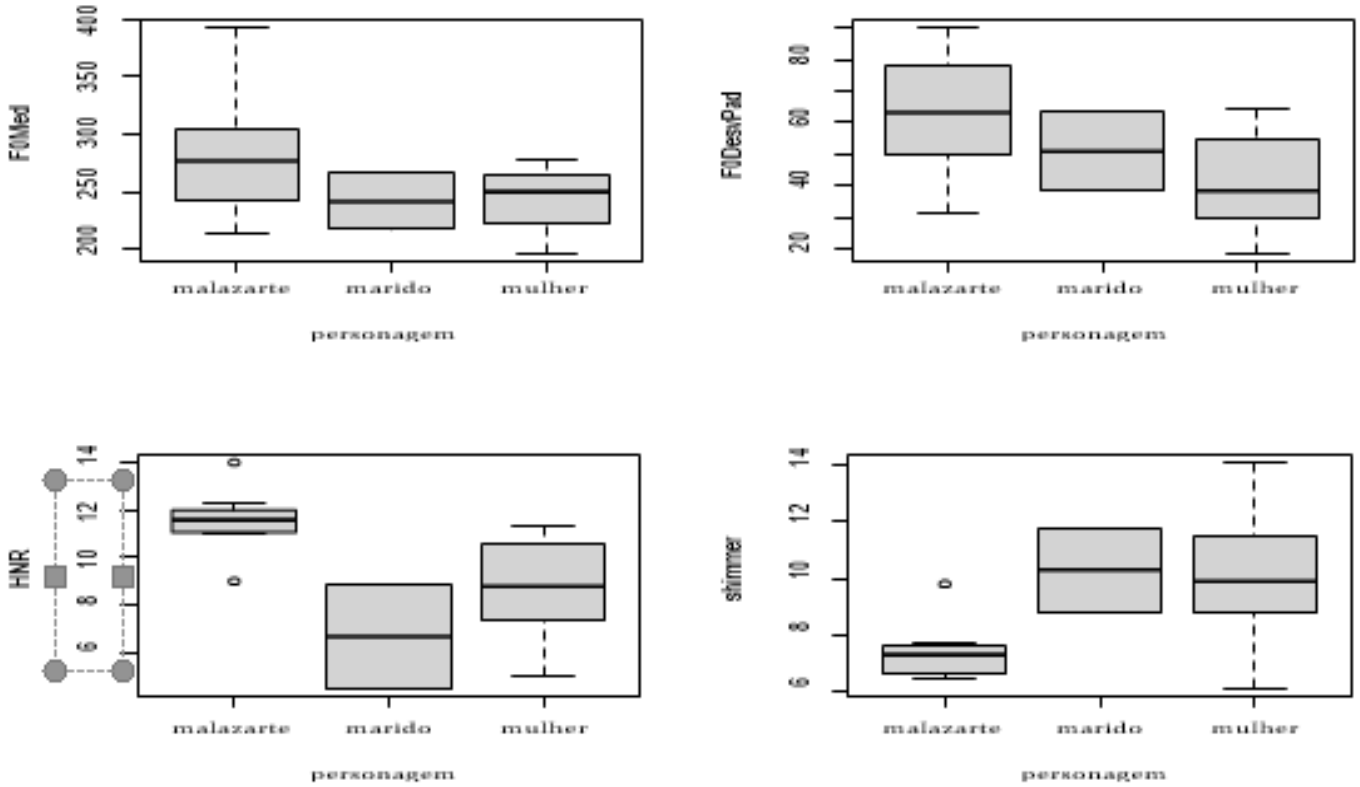


Figura 5.25 – Diagramas de blocos de medidas melódicas e de qualidade de voz na interpretação de três personagens por Camila Pitanga na história de Pedro Malazarte. As medidas são a mediana da Fo em Hertz, o desvio-padrão da Fo em Hertz, a razão harmônico-ruído em dB e o *shimmer* em porcentagem.

## 5.4 Prelúdio para o próximo capítulo

No próximo capítulo vamos aplicar, no contexto de exemplos de desenho experimental, o que aprendemos com todas as medidas prosódico-acústicas vistas até o momento. Começaremos com uma apresentação sucinta das principais técnicas de análise estatística inferencial usadas na área de prosódia experimental.