

## Aspectos de produção e percepção de estilos de elocução profissionais e não profissionais em quatro línguas

Plínio A. Barbosa  
Unicamp

### Introdução

Ao fazer uma revisão dos trabalhos apresentados para o *Workshop on Integrating Speech and Natural Language*, Llisterri (1992) apontava para a diversidade de termos associados a estilo de elocução. Embora os estilos majoritariamente mencionados fossem (e ainda são) os da fala lida e da fala espontânea, o estilo de elocução profissional fora citado para se referir à fala usada por jornalistas e professores em seu trabalho (DELGADO; FREITAS, 1991).

Essa variação de terminologia levou Eskénazi (1993) a propor uma definição de estilo de elocução (*speaking style*, em inglês) fundamentada na observação de três dimensões de variação: grau de clareza, familiaridade e estrato social. A primeira dimensão vai do esforço mínimo à clareza máxima (de pouco a muito inteligível, como da “leitura para criança” à “locução de futebol”), a segunda depende do grau de familiaridade entre os interlocutores (de familiar a não familiar, como de “falando a um amigo próximo” a “falando a um estranho”), enquanto que a terceira depende da situação e da diferença de nível social entre os interlocutores (de baixo estrato a alto estrato, como de “conversa diária” a “conferência formal”).

Um estudo adequado de estilos de elocução pressupõe, assim, que se identifique com clareza o tipo de tarefa que realiza o locutor ao se dirigir a um único ou mais interlocutores (presumíveis ou presentes). E, portanto, a denominação restrita de “fala espontânea” é insuficiente quando não é descrita a partir de dimensões como as apontadas por Eskénazi ou por Barbosa (2012), esse último a partir de dois eixos de variação: de gênero textual e grau de controle do experimentador. O estilo de fala se dá, portanto, numa dimensão eminentemente de variação que

requer um estudo formal e controlado dos parâmetros acústicos que são modificados (BARRY, 1995; BARBOSA; MADUREIRA, 2016). A necessidade desse estudo é historicamente apontada por Eskénazi (1993), após uma revisão de 150 artigos sobre o tema que mostram que os estilos de elocução estão relacionados a mudanças em qualidade de voz, como sopro, tensa, laringalizada; taxa de elocução e situações específicas de comunicação, como fala profissional, entrevista e transmissão de notícias e esportes via rádio e televisão.

Tanto parâmetros segmentais quanto prosódicos são modificados entre os diferentes estilos. O estilo de elocução é, assim, entendido como um aspecto da diferenciação do modo de falar (IRVINE, 2001), que perpassa a interface entre fonética e sociolinguística, já apontada por Argente (1992) e cunhada pelo estudo da sociofonética, conforme propuseram Foulkes e Docherty (2006) ao fazer uma retrospectiva de seu uso:

The term has a reasonably long history (tracing back at least to Deshaies-Lafontaine, 1974), but it has been neither widely nor consistently used. [...] It has had its widest currency among phoneticians, who have used it to refer especially to descriptive accounts of speech production across different dialects, speaker groups or speech styles (e.g., Esling, 1991; Henton & Bladon, 1988).

Essa diferença no modo de falar, entre estilos de elocução, pode ser estudada pelo exame das diferenças prosódicas dos enunciados produzidos em cada estilo. Pressupõe-se que os ouvintes conseguem discriminar e identificar essas diferenças a partir da prosódia, conforme demonstra a literatura mais recente.

Ao usar enunciados deslexicalizados<sup>1</sup> de quatro diferentes estilos profissionais em francês padrão, a saber, o comentário esportivo, o sermão religioso bem como os discursos político e jornalístico, Obin et al. (2011) mostraram que é possível identificar acima do aleatório esses quatro estilos em teste de classificação de escolha forçada. Os resultados apontaram que o comentário esportivo é o melhor identificado, seguido do discurso jornalístico. O discurso político e o sermão religioso foram pior identificados. Esses dois últimos são muito confundidos entre si, enquanto o jornalístico é em parte confundido com o discurso político. O comentário esportivo é o menos confundido dos quatro. Os autores apontam que parâmetros como taxa de elocução, intensidade, uso de pausa e estruturação

---

1 A deslexicalização é uma técnica que permite mascarar ou eliminar a informação segmental impossibilitando a compreensão do conteúdo e preservando a informação prosódica do enunciado.

de proeminências prosódicas são os principais parâmetros que guiam os ouvintes na classificação.

São também características prosódicas que distinguem os estilos estudados por Higuchi et al. (1997) em japonês, a saber, enunciados naturais atuados com raiva, gentileza e de forma apressada. Os autores apontaram que o valor mínimo da frequência fundamental ( $F_0$ ) se manteve alto para a raiva com contornos melódicos monótonos. A variação melódica na fala com gentileza foi a maior entre os três estilos, enquanto que o estilo apressado exibiu contornos melódicos com variação intermediária.

Barbosa e Silva (2012) mostraram que os ouvintes discriminam fala lida da narrada em português brasileiro também com a fala deslexicalizada e que as taxas de elocução, de sílabas acentuadas e o uso de pausas são os principais parâmetros empregados para discriminar esses dois estilos. Recentemente, mostramos que parâmetros temporais têm papel mais relevante, embora não exclusivo, que parâmetros melódicos para realizar a discriminação entre falas narrada e lida (BARBOSA, 2015).

Assim, como mostramos nesses trabalhos e em pesquisas anteriores (cf. BARBOSA, 2006), embora a organização da duração de unidades silábicas ao longo dos enunciados tenha um papel fundamental para a estruturação rítmica da fala e, portanto, para a diferenciação entre estilos como a narração e a leitura, é possível que, com a inclusão de outros estilos, especialmente os profissionais, a melodia tenha um papel que tenha sido até então negligenciado. É esse o intuito deste trabalho, ao comparar diferenças prosódicas acústicas e percebidas em quatro línguas e quatro estilos de elocução, sendo dois profissionais (discursos político e jornalístico) e dois não profissionais (leitura e narração).

## 2 Metodologia

### 2.1 Corpus

O corpus CROSS-RHYTHM compreende gravações em português brasileiro (PB), português europeu (PE), francês padrão (FR) e alemão padrão (AL). Ele é composto de material de fala lida e narrada de 10 sujeitos (5 homens e 5 mulheres) entre 25 e 40 anos em cada língua, a que chamamos de fala não profissional, uma vez que nenhum desses sujeitos usa a voz profissionalmente. Para os discursos político e jornalístico, colhemos gravações de até 5 minutos do *YouTube* de 6 sujeitos (3 homens e 3 mulheres), também em cada língua, que denominamos estilo profissional. As gravações da fala não profissional foram feitas no Brasil por Sandra Madureira (PUC-SP) e por mim, enquanto que as gravações nas demais

línguas foram realizadas por colaboradores em Portugal (Catarina Oliveira, Univ. de Aveiro), França (Philippe Boula de Mareüil, LIMSI) e Alemanha (Hansjörg Mixdorff, Beuth Univ. of Berlin), todas usando um microfone unidirecional. Os áudios dos vídeos da seleção de fala profissional foram extraídos em formato WAV. Todos os enunciados foram amostrados ou reamostrados a 16 kHz, com 16 bits de quantização. A partir das gravações, extraíram-se trechos para análise acústica e perceptiva.

Todas as seleções de trecho para análise nos quatro estilos foram feitas para as quatro línguas, totalizando 517 trechos. Os trechos se estendem de 10 a 20 segundos, extensão considerada ideal para a análise perceptiva de diferenças no modo de falar, como mostrou estudo anterior (BARBOSA; SILVA, 2012).

O texto escolhido para a leitura foi “O Monge Desastrado”, um texto originalmente escrito em PE sobre a origem dos pastéis de Belém. O texto foi adaptado para o PB por Juva Batella e traduzido para o francês por Philippe de Mareüil e por mim, e, para o alemão, por Hansjörg Mixdorff (a partir de uma tradução para o inglês, feita por mim). As narrações foram feitas logo após as leituras pelos mesmos sujeitos em cada língua com gravações conduzidas pelos mesmos colaboradores mencionados acima.

## 2.2 Parâmetros acústicos

O script *ProsodyDescriptor2*, que roda em Praat (BOERSMA; WEENINK, 2016), que desenvolvemos para este trabalho, calcula oito parâmetros prosódicos a partir de um arquivo de áudio. Os parâmetros são mediana, desvio padrão e assimetria de  $F_0$  em semitons (re 1 Hz), mediana, desvio padrão e assimetria da taxa de  $F_0$  em semitons por segundo, taxa de picos da curva suavizada a 1,5 Hz de  $F_0$  e um parâmetro ligado ao esforço vocal, a ênfase espectral (TRAUNMÜLLER; ERIKSSON, 2000), medida obtida através do cálculo da diferença entre a intensidade acústica do sinal original e a intensidade do sinal após ser submetido a um filtro passa-baixas com frequência de corte em 400 Hz. Os seis primeiros parâmetros são estritamente melódicos, enquanto que a taxa suavizada de picos de  $F_0$  é um parâmetro rítmico, por envolver uma taxa de produção de proeminências. A ênfase espectral se relaciona ao esforço vocal: quanto maior seu valor, maior esse esforço.

Os parâmetros prosódico-acústicos mencionados foram obtidos pelo script para uma seleção de 218 trechos de fala lida e narrada dos 10 sujeitos gravados por língua. Os demais 299 trechos foram obtidos das falas profissionais. A partir dos valores dos parâmetros, usamos a estatística inferencial para avaliar as diferenças entre suas médias.

### 3 Análises acústicas dos estilos de elocução

#### 3.1 Estilos não profissionais: falas lida e narrada

O teste não paramétrico Kruskal-Wallis para  $\alpha = 5 \%$ , seguido de análises post hoc de Bonferroni, revelaram que os parâmetros assinalados na Tabela 1 diferenciam fala lida e narrada em PB e em francês. Em PE e alemão, não há diferenciação dessas duas falas por nenhum dos oito parâmetros analisados.

**Tabela 1** Parâmetros que discriminaram entre fala lida (LE) e narrada (NR) em PB e FR. Mediana da taxa de  $F_0$  (dF0med) em st/s; Assimetria de Pearson da taxa de  $F_0$  (dF0as); Taxa de picos de  $F_0$  (F0t) em picos/s. Valores de p para o teste de Kruskal-Wallis para  $\alpha = 5 \%$ .

Língua	Parâmetro	p <	Valores médios
PB	dF0med	$10^{-5}$	-0,02 (NR) -0,15 (LE)
	dF0as	$10^{-4}$	-0,02 (NR) -0,39 (LE)
	F0t	0,01	0,89 (NR) 0,94 (LE)
FR	dF0med	0,0002	0,05 (NR) -0,05 (LE)

Para o parâmetro dF0med, não há interação entre sexo e estilo em PB, mas há diferença entre os sexos com as seguintes médias: -0,19 (LE, M), -0,067 (NR, M), -0,13 (LE, H) e 0,02 (NR, H). Quanto ao francês, no mesmo parâmetro, diferença entre os sexos apenas para o estilo de leitura com valores médios de -0,1 (M) e -0,02 (H).

Para o parâmetro dF0as, não há interação entre sexo e estilo em PB, mas há diferença entre os sexos com as seguintes médias: -0,55 (LE, M), -0,16 (NR, M), -0,23 (LE, H) e 0,12 (NR, H). Já para o parâmetro F0t em PB, há diferença significativa entre os sexos. Quanto aos estilos, há diferença média nesse parâmetro apenas nas mulheres: -1,01 (LE, M), 0,92 (NR, M), 0,88 (H, nos dois estilos).

Antes de comparar o uso dos parâmetros acústicos entre as línguas, convém explicar para que tipo de aspectos da melodia e do ritmo apontam esses resultados. O parâmetro dF0med está relacionado à media das taxas de variação de  $F_0$ , ou seja, das subidas e descidas melódicas. Para o PB, o valor médio negativo

nos dois estilos, indica maior proporção de descidas do que de subidas. Mas a proporção de subidas é maior na narração, relacionado a subidas mais rápidas e frequentes do que a leitura. Esse dispositivo é ainda maior na narração masculina, que tem proporção total de subidas maior que de descidas. O mesmo se dá em francês, mas apenas na leitura, como homens com maior proporção de subidas que as mulheres. Na narração, nessa língua, a proporção total de subidas é maior que de descidas nos dois sexos.

O parâmetro  $dF_0as$  está relacionado à cauda da distribuição das taxas de variação de  $F_0$ , com valores negativos indicando distribuição assimétrica à esquerda, correlato de uma maior tendência a descidas de  $F_0$ . Em PB, esse valor é quase nulo na narração, indicando que essa tendência da assimetria negativa existe com maior grau na leitura, pela maior proporção de instâncias de terminalidade. O parâmetro  $F_0t$ , por sua vez, é a taxa de produção de picos de  $F_0$  na curva melódica filtrada a 1,5 Hz (para preservar apenas os picos relevantes linguisticamente). Em PB, esse valor é significativamente menor na narração, indicando maior distância temporal entre as unidades salientes acusticamente, o que parece ser característico da narração, relacionado também à presença de pausas silenciosas mais longas do que na leitura.

Quanto às diferenças no uso dos parâmetros entre as línguas, relacionadas ou não a um estilo em particular, a Tabela 2 apresenta os resultados. O que se deduz do exame da tabela é que o PB se aproxima em leitura em francês para dois parâmetros acústicos ( $dF_0as$  e  $F_0t$ ), enquanto se aproxima, independentemente de estilo do PE para  $F_0med$  e  $F_0dp$ . Também se vê que o PB, independentemente do estilo, é distinto do alemão em todos os casos. Que PB e PE não sejam distintos na mediana de  $F_0$  pode indicar um amálgama de aspectos fisiológicos e linguísticos, que fazem que portugueses e brasileiros soem mais agudo.

**Tabela 2** Parâmetros que discriminam entre as línguas. Além dos referidos na Tab. 1, Mediana de  $F_0$  ( $F_0med$ ) em st; Desvio-padrão de  $F_0$  ( $F_0dp$ ) em st; Assim. de Pearson de  $F_0$  ( $F_0as$ ); Desvio-padrão da taxa de  $F_0$  ( $dF_0dp$ ) em st/s; Ênfase espec. (enf) em dB.

Parâmetro	pelo menos, $p <$	Valores médios
$F_0med$	0,01	89 (PB=PE) > 87 (FR=AL)
$F_0dp$	0,02	1,5(FR) < 2,2 (PB=PE) < 3,0 (AL)

(continua)

**Tabela 2** Parâmetros que discriminam entre as línguas. Além dos referidos na Tab. 1, Mediana de  $F_0$  ( $F_{0med}$ ) em st; Desvio-padrão de  $F_0$  ( $F_{0dp}$ ) em st; Assim. de Pearson de  $F_0$  ( $F_{0as}$ ); Desvio-padrão da taxa de  $F_0$  ( $dF_{0dp}$ ) em st/s; Ênfase espec. (enf) em dB. (continuação)

Parâmetro	pelo menos, p <	Valores médios
$F_{0as}$	$10^{-5}$ (LE) 0,002 (NR)	LE: PB (-0,08) < FR = PE (0,11) < AL (0,28) NR: PB (0,02) < FR = AL = PE (0,22)
$dF_{0med}$	0,04 (LE)	LE apenas: -0,15 (PB) < FR=PE=AL (-0,05)
$dF_{0dp}$	0,05	FR (2,5) < PB (3,2) < PE (4,11) < AL (5,4)
$dF_{0as}$	0,02 (LE)	LE apenas PB = FR (-0,3) < PE = AL (0,2)
enf	$10^{-5}$	FR (0,1) < PE = AL (0,8) < PB (2,6)
$F_{0t}$	0,002	LE: PB = FR (0,9) > AL = PE (0,8) NR: PB (0,9) > AL (0,7)

A fala dos alemães é a mais variável melodicamente entre todos e a do francês a menos. Sobre o primeiro fato já apontamos essa diferença entre PB e alemão em estudo anterior (Barbosa, Mixdorff, Madureira, 2011). A assimetria de  $F_0$  revela em PB uma tendência na narração a uma frequência superior de valores de  $F_0$  mais baixos do que a média na leitura. Essa tendência a uma frequência maior de valores de  $F_0$  mais altos nas demais línguas vale para os dois estilos. O PB tende também a ter taxas de  $F_0$  ( $dF_0$ ) em média mais negativas, privilegiando assim descidas mais rápidas do que nas demais línguas com variabilidade ( $dF_{0dp}$ ) intermediária em realizar isso, com francês a menos variável nesse parâmetro. PB é também a língua que faz o maior esforço vocal ao falar (enf), além da maior taxa de picos de  $F_0$ , o que aponta para uma maior vivacidade (*liveliness* em inglês). Em relação ao PE, PB tem mesmo valor mediano de  $F_0$ , com diferença apenas para as narrativas femininas: 94 st (PB) > 91 st (PE).

### 3.2 Estilos profissionais: discurso jornalístico e político

A Tabela 3 aponta as diferenças entre os dois estilos profissionais nas línguas que os diferenciam do ponto de vista melódico e rítmico. Em alguns casos somente as mulheres exibem a diferença entre os estilos, mas não os homens.

**Tabela 3** Parâmetros que discriminam entre discurso jornalístico (JN) e político (PL) entre línguas pertinentes. Mediana de  $F_0$  (F0med) em st; Ênfase espec. (enf) em dB; Taxa de picos de  $F_0$  (F0t) em picos/s. Valores de p para o teste post hoc de Bonferroni  $\alpha = 5\%$ .

Língua	Parâmetro	p <	Valores médios
PE	F0med	0,006 (M)	92 (JN) < 99 (PL)
	enf	0,0005	2,0 (JN) < 4,2 (PL)
FR	F0t	0,0002 (M)	0,8 (JN) < 1,2 (PL)
	enf	0,05	0,8 (JN) < 1,6 (PL)
AL	F0med	0,008	88 (JN) < 95 (PL)
	enf	0,02 (M)	0,4 (JN) < 1,9 (PL)

A ausência do PB na tabela se deve ao fato de ser a única língua que não distingue os dois estilos para os parâmetros considerados. Para as demais, a ênfase espectral assinala esforço maior no discurso político, embora em alemão essa diferença esteja restrita às mulheres. A mediana de  $F_0$  é superior no discurso político feminino português e, independentemente de sexo, também superior em alemão. O francês distingue pela taxa de picos de  $F_0$ , indicador de vivacidade, com maior valor para discurso político feminino.

**Tabela 4** Parâmetros que discriminam o discurso jornalístico entre as línguas. Vide tabelas anteriores para indicação dos parâmetros. Valores de p para teste de Bonferroni para  $\alpha = 5\%$ .

Parâmetro	peelo menos, p <	Valores médios
F0med	0,04	88 (AL) < 91 (PB=PE=FR)
F0dp	0,03	2,0 (FR=AL) < 3,0 (PB=PE)
dF0dp	0,008	4,0 (FR=AL) < 6,2 (PB=PE)
enf	0,007	0,6 (FR=AL) < 3,4 (PB=PE)
F0t	0,02	somente 0,9 (PB) < FR (1,1)



Quanto a diferenças de parâmetros entre as línguas, somente a fala de jornalistas é distinta entre as línguas, por isso, a Tabela 4 só se refere a esse estilo, o que é um resultado muito interessante. Assim, o discurso político tem mesma média para todos os parâmetros examinados entre as línguas. Outro resultado interessante é que os parâmetros no discurso jornalístico têm mesmas médias entre PB e PE contra a mesma média para francês e alemão.

Além disso, a tabela 4 mostra que PB e PE empenham maior esforço vocal que francês e alemão no estilo jornalístico. Que a variabilidade tanto de  $F_0$  quando de sua taxa é superior em PB e PE contra francês e alemão, tornando o discurso jornalístico mais vivaz nas primeiras. O Alemão tem em média o menor valor de  $F_0$ , soando mais grave que as demais, portanto. Em compensação, quando comparado ao francês, o PB tem menor taxa de produção de picos de  $F_0$ .

3.3 Estilos profissionais e não profissionais

A Tabela 5 apresenta uma comparação entre os estilos profissionais e não profissionais para as duas variedades do português.

**Tabela 5** Parâmetros que discriminam entre os estilos profissionais e não profissionais em duas variedades do português. Para todos os casos,  $p < 0,05$  em um teste post hoc não paramétrico de Bonferroni.

Língua	Parâmetros	
PB	F0med	ns
	F0dp	(PL=JN) > (LE=NR)
	F0as	(PL=JN) > NR > LE
	dF0med	NR > LE
	dF0dp	(PL=JN) > (LE=NR)
	dF0as	NR > LE
	enf	ns
	F0t	PL > LE, NR < LE
PE	F0med	PL > (JN =LE=NR)
	F0dp	JN>LE
	F0as	(PL=JN) < (LE=NR)
	dF0med	ns
	dF0dp	PL > (LE=NR), JN=PL
	dF0as	ns
	enf	JN>PL>(LE=NR)
	F0t	(PL=JN) > (LE=NR)

Para o PB, depreende-se da Tabela 5 que os estilos profissionais se agrupam com valores superiores aos dos estilos não profissionais para F0dp, F0as e dF0dp, que se relacionam à variabilidade da curva de F<sub>0</sub>. No estilo político, a taxa de picos de F<sub>0</sub> é maior do que na leitura. Já no PE, com exceção da ausência de significância nas variáveis ligadas à taxa de variação de F<sub>0</sub>, dF0med e dF0as, para as demais os estilos político (F0med), jornalístico (F0dp e enf) ou ambos (F0as, dF0dp, F0t) têm valores superiores ao estilos de leitura (F0dp) ou leitura e narração. Nas duas variedades do português, a variação de F<sub>0</sub> e sua taxa (dF0), além da taxa de picos de F<sub>0</sub> é superior nos estilos profissionais.

**Tabela 6** Parâmetros que discriminam entre os estilos profissionais e não profissionais em francês (FR) e alemão (AL).  
Para todos os casos,  $p < 0,05$  em um teste post hoc não paramétrico de Bonferroni.

Língua	Parâmetros	
FR	F0med	(PL=JN) > (LE=NR)
	F0dp	(PL=JN) > (LE=NR)
	F0as	JN < NR, LE < NR
	dF0med	ns
	dF0dp	(PL=JN) > (LE=NR)
	dF0as	ns
	enf	(PL=JN) > (LE=NR)
	F0t	JN > (RE=PL=NR)
AL	F0med	PL > (JN =LE=NR)
	F0dp	JN < NR
	F0as	PL < JN < (LE=NR)
	dF0med	ns
	dF0dp	ns
	dF0as	ns
	enf	PL > (LE=NR=JN)
	F0t	(PL=JN) > (LE=NR)

A Tabela 6 apresenta uma comparação entre os estilos profissionais e não profissionais para o francês e o alemão. Para o francês, depreende-se da Tabela 6 que os estilos profissionais também tendem a se agrupar com valores superiores aos dos estilos não profissionais para F0med, F0dp, dF0dp e enf, que se relacionam ao valor mediano de F<sub>0</sub> (F0med), à variabilidade da curva de F<sub>0</sub> e de sua taxa (F0dp e dF0dp) bem como ao esforço vocal (enf). No estilo jornalístico apenas a taxa de picos de F<sub>0</sub> é maior do que nos demais estilos. Como no PE, as variáveis

dF0med e dF0as não diferem entre os quatro estilos. Já no alemão, com exceção da ausência de significância nas três variáveis ligadas à taxa de variação de  $F_0$ , para as demais, os estilos político (F0med e enf) ou os dois estilos profissionais (F0t) têm valores superiores aos estilos de leitura e narração.

Depreende-se das duas tabelas que, em geral, excetuando-se o papel menor dos descritores da taxa de variação de  $F_0$  para o PE, francês e alemão e a mediana de  $F_0$  para o PB, as demais variáveis apontam para valores superiores para os estilos profissionais, diferenciando-os dos não profissionais.

## 4 Testes de percepção

Foram aplicados um teste de classificação e três testes de discriminação. O teste de classificação com 94 estímulos foi feito para reconhecer um dos quatro estilos de elocução nas 4 línguas. De forma a impedir o reconhecimento imediato de um político ou jornalista brasileiro, aplicamos o algoritmo PURR (SONN-TAG; PORTELE, 1998) para deslexicalizar todos os estímulos e fazer com que o ouvinte prestasse atenção apenas na prosódia para a classificação.

Os estímulos separados para os testes de discriminação consideraram os dois sexos de forma igual com pares intra e intersexos, as duas ordens de apresentação entre os estilos na mesma língua e em línguas distintas e a ordem entre as línguas, todas comparadas entre si. Isso foi feito para que não fossem fatores que pudessem afetar os julgamentos. Foram montados três testes: o primeiro, com 56 estímulos não deslexicalizados para discriminação dos estilos leitura e narração, considerando todas as combinações entre línguas, ordens de apresentação e combinação dos sexos. O segundo, com 78 estímulos não deslexicalizados para discriminação do estilo jornalístico em relação aos estilos político e de narração, pois são os estilos em princípio que mais se aproximam de uma fala para um interlocutor imediato. Esse teste excluiu o PB, pois o reconhecimento do profissional enviesaria as respostas. Por conta disso, um terceiro teste de discriminação com áudio deslexicalizado para estímulos do PB foi montado com 66 estímulos (em testes comparando o estilo jornalístico com narração e discurso político, para pares de línguas em que uma delas é o PB num desses estilos). A pergunta ao ouvinte, quando o bloco continha um trecho de estilo jornalístico, foi qual dos dois estímulos separados por um tom puro, era discurso jornalístico. Quando o bloco continha um trecho de leitura contrastado à narração, foi qual dos dois estímulos era leitura. Assim, 294 estímulos foram considerados para os testes, realizados com pelo menos 10 ouvintes brasileiros universitários em cada caso.

Para o teste de classificação, o único estilo que teve reconhecimento indistinto do aleatório foi a leitura. Isto é, a leitura não foi um estilo reconhecido.

Quanto aos demais estilos, eles são todos três diferentes do acerto aleatório (25 %) com  $p$  pelo menos  $10^{-5}$ . Os percentuais de acerto foram: (PL = 47 %, JN = NR = 40 %), significativo ao considerar esses três estilos apenas com  $X^2 = 5,9$ , e  $p = 0,015$  e poder de teste de cerca de 67 %, isto é, a probabilidade de rejeitar uma  $H_0$  falsa. Assim, o estilo discurso político é o mais facilmente identificado. A coerência entre os ouvintes na classificação, embora significativa, é de apenas 0,08 pelo teste Kappa de Fleiss.

A matriz de confusão da Tabela 7 permite visualizar o número de respostas para cada estilo.

**Tabela 7** Matriz de confusão para o testes de classificação.

↓ Resp.\Estímulos →	LE	NR	JN	PL
LE	84 (29%)	78 (28%)	42	47
NR	90	112 (40%)	61	54
JN	76	47	111 (40%)	45
PL	38	39	62	130 (47%)

Percebe-se na matriz que os estímulos de leitura, por ser leitura de história, provavelmente, são mais confundidos com os de narração e muito próximos aos do discurso jornalístico também. Embora a narração também seja confundida em parte com a leitura, a confusão é menor e os discursos profissionais ficam mais à parte. Quanto a esses, o discurso político é o que tem mais acertos e é confundido em menor grau de forma semelhante com os outros três. Já o estilo jornalístico é confundido mais com o discurso político e a narração. Esses resultados são semelhantes ao de Obin et al. (2011) para o discurso jornalístico. Quanto ao fato de o discurso político ser melhor identificado aqui, mas não no estudo com o francês, é porque esse último inclui o sermão religioso, mais confundido com o discurso político. Mas ambos foram discriminados do jornalístico, conforme apontam os autores.

Quanto às línguas, o teste de proporções para  $\alpha = 5\%$  apontou todas as proporções iguais estatisticamente a 39 % de acerto (significativamente distinto do acerto aleatório). Não há assim efeito de língua no teste de classificação. As percentagens de acerto baixas são devidas ao fato de que o teste usa fala deslexicalizada. Passemos agora a examinar os acertos dos ouvintes quanto à discriminação entre os estilos de elocução.

Para prever as categorias do teste de classificação a partir dos parâmetros acústicos, empregamos uma Análise Discriminante Linear (LDA, na sigla em inglês). Considerando os 517 trechos empregados para a análise acústica apresen-

tada anteriormente, obtém-se com um modelo LDA cuja predição de categorias se encontra na Tabela 8.

**Tabela 8** Matriz de confusão para do modelo LDA.

↓ Resp.\Estímulos →	LE	NR	JN	PL
LE	165 (76%)	77 (44%)	29 (48%)	18 (28%)
NR	44 (20%)	91 (52%)	8 (13%)	18 (28%)
JN	7 (3,5%)	5 (3%)	14 (23%)	2 (3%)
PL	1 (0,5%)	2 (1%)	9 (15%)	27 (41%)

Observa-se, pela comparação das tabelas 7 (classificação via percepção) e 8 (classificação via acústica), que a discriminação entre os estilos é melhor feita pelo modelo LDA a partir dos parâmetros acústicos para leitura e narração com percentagens de acerto superiores ao feito pelos ouvintes para esses estilos. Mas, para esse mesmo modelo, o estilo narrativo é mais confundido com a leitura do que na percepção. No que toca ao estilo jornalístico, os ouvintes acertam 40% contra apenas 23% via LDA, onde ele é mais confundido com a leitura. O desempenho da classificação do estilo político é semelhantes na classificação automática e via ouvintes.

Entre as quatro línguas: (1) o estilo político é melhor classificado pelo LDA para o PE (em PB e FR ele é mais confundido com a narração), seguido do alemão com percentagem semelhante de confusão com a leitura.; (2) o estilo jornalístico é melhor classificado pelo LDA para o PE (em PB ele é confundido com o político e a leitura enquanto que em FR e AL com a leitura); (3) a leitura é melhor classificada pelo LDA em todas as línguas com destaque para o PB com melhor discriminação em relação aos demais; (4) a narração é melhor classificada pelo LDA em PE, AL e FR enquanto que em PB é confundida com a leitura, que foi a segunda percentagem de classificação de narrações pelos ouvintes para todas as línguas sem o léxico.

Os resultados do teste de discriminação, que têm resposta aleatória para 50% de acerto, uma vez que são feitos por pares de estímulos, revelaram que:

- (1) Entre leitura e narração, o acerto é superior a 90 % para a leitura, independentemente de língua;
- (2) Entre estilos jornalístico e político, o acerto é superior a 85 % para o primeiro estilo, independentemente de língua, excluído o PB;
- (3) Entre estilos jornalístico e narrativo, o acerto é superior a 85 % para o primeiro estilo, independentemente de língua, excluído o PB;

- (4) Entre estilos jornalístico e político, delexicalizado para incluir o PB, o acerto é superior a cerca de 65 % (sig. distinto de 50% para  $\alpha = 1\%$ ) para o primeiro estilo, independentemente de ser entre PB e outra língua ou somente o PB;
- (5) Entre estilos jornalístico e narrativo, delexicalizado para incluir o PB, o acerto é superior a cerca de 80 % para o primeiro estilo, independentemente ser entre PB e outra língua ou somente o PB.

## 5 Conclusão

Os resultados apresentados revelam que:

- (1) O conjunto de parâmetros acústicos extraídos, analisados e usados para classificação predizem em parte o comportamento dos ouvintes, que tendem a considerar a leitura apenas pela informação prosódica mais como narração do que a classificação automática, que a prediz a mais de 75 % corretamente. Quando pareada com a narração apenas, com o léxico, seu reconhecimento é superior a 90 %. O estilo jornalístico é melhor reconhecido pelo ouvinte do que ele é predito pela acústica. O estilo político é o melhor classificado entre os estilos profissionais via acústica ou percepção;
- (2) A coerência entre ouvintes no teste de classificação é baixa, uma vez também que o teste envolve itens deslexicalizados;
- (3) Os testes de discriminação revelam grande facilidade de discriminação entre dois estilos, com exceção dos estilos jornalístico e político, o que é um resultado distinto do teste de classificação pelos ouvintes, mas semelhante àquele via LDA;
- (4) Línguas distintas em geral não interferem na discriminação entre estilos, no entanto, narração e leitura são mais confundidas dentro do PB, apesar de a Tabela 1 mostrar parâmetros que os discriminam que, certamente, não são bem apreendidos pelos ouvintes do PB.
- (5) Os estilos profissionais são melhor discriminados em PE e em segundo lugar em alemão, o que está de acordo com parâmetros de  $F0_{med}$  e  $enf$  na Tabela 3.
- (6) Embora haja parâmetros que discriminem a fala profissional da não profissional, como há um certo grau de confusão entre esses estilos em cada língua, conclui-se que esses parâmetros nem sempre estão claramente distintos em cada par de estímulos examinados ou são de difícil apreensão pelo ouvinte.

## Agradecimentos

Este trabalho foi desenvolvido no quadro do projeto Universal CNPq “Análise e Modelamento Estatístico Inter-linguístico do Vínculo entre Produ-

ção e Percepção do Ritmo da Fala em Diferentes Estilos de Elocução”, Processo 476358/2013-2. O autor também é financiado pela bolsa de produtividade em pesquisa do CNPq, Processo 302657/2015-0. Agradecemos os colaboradores do projeto, Sandra Madureira, Philippe Boula de Mareüil e Hansjörg Mixdorff, o aluno Robson R. Monteiro, que ajudou com a preparação de áudios para o testes de percepção, bem como nossos locutores das tarefas de leitura e narração e os ouvintes das tarefas de percepção.

## Referências

- Argente, J. A. (1992). From speech to speaking styles. *Speech Communication* 11, 325-335.
- Barbosa, P. A. (2006) *Incursões em torno do ritmo da fala*. Campinas: Pontes/Fapesp.
- Barbosa, P. A. (2012) Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. *Revista de Estudos da Linguagem* (UFMG). 20 (1), p. 11-27.
- Barbosa, P. A. (2015). Temporal parameters discriminate better between read from narrated speech in Brazilian Portuguese. *Proc. of the 18th International Congress of Phonetic Sciences*, Glasgow, Reino Unido: The Scottish Consortium for ICPhS 2015.
- Barbosa, P.; Madureira, S. (2016). Elicitation techniques for cross-linguistic research on professional and non-professional speaking styles. *Proc. Speech Prosody 2016*, 503-507.
- Barbosa, P. A., Mixdorff, H.; Madureira, S. (2011) Applying the quantitative target approximation model (qTA) to German and Brazilian Portuguese. *Proc. of Interspeech 2011*. Florença, Itália. Pp. 2065-2068.
- Barbosa, P. A.; da Silva, Wellington. (2012). A New Methodology for Comparing Speech Rhythm Structure between Utterances: Beyond Typological Approaches. In: H. Caseli et al. (Eds.): *PROPOR 2012*, LNAI 7243, pp. 329-337. Springer, Heidelberg.
- Barry, W. J. (1995). Phonetics and Phonology of speaking styles. *Proceedings of the International Congress of Phonetic Sciences*, Estocolmo, 1995. v. 2, p. 4-10.

- Boersma, P.; Weenink, D. (2016). Praat: doing phonetics by computer. <http://www.praat.org/>.
- Delgado, M. R.- Freitas, M. J. (1991). Temporal structures of speech: 'Reading news on TV'. In *Proceedings of the ESCA Workshop Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication*. Barcelona, 1991. p. 19.1 -19.5.
- Deshaies-Lafontaine, D. (1974). *A socio-phonetic study of a Québec French community: Trois-Rivières*. Tese de Doutorado, Londres: University College London.
- Eskénazi, M. (1993). Trends in speaking styles research. *Proceedings of the Eurospeech 1993*. Berlim, p. 501-509.
- Esling, J. H. (1991). Sociophonetic variation in Vancouver. In: J. Cheshire (Ed.) *English around the world*. Cambridge: Cambridge University Press. p. 123-133.
- Foulkes, P.; Docherty, G. J. (2006). The social life of phonetics and phonology. *Journal of Phonetics*. 34, 409-438.
- Henton, C.; Bladon, A. (1988). Creak as a sociophonetic marker. In: L. Hyman, & C. N. Li (Eds.). *Language, speech and mind: Studies in honor of Victoria A. Fromkin*. Londres: Routledge. p. 3-29.
- Higuchi, N.; Hirai, T.; Sagisaka, Y. (1997). Effects of speaking style on parameters of fundamental contour. In: van Santen, J. P. H.; Sproat, R. W.; Oliver, J. P.; Hirschberg, J. (Ed.) *Progress in Speech Synthesis*. Nova York: Springer-Verlag, p. 417-428.
- Irvine, Judith, T. (2001). Style as distinctiveness: the culture and ideology of linguistic differentiation. In: Eckert, Penelope; Rickford, John. R. *Style and sociolinguistic variation*. Cambridge: Cambridge University Press. p. 21-43.
- Llisterri, J. (1992). Speaking styles in speech research. In *ELSNET/SALT/ESCA Workshop Integrating Speech and Natural Language*, p. 17-37, University College Dublin.
- Obin, N.; Lanchantin, P.; Lacheret, A.; Rodet, X. (2011). Discrete/Continuous Modelling of Speaking Style in HMM-based Speech Synthesis: Design and Evaluation. *Proc. of Interspeech 2011*, Florença, Itália.



Sonntag, G.P.; Portele, T. (1998). PURR - a method for prosody evaluation and investigation. *Journal of Computer Speech and Language*, 12 (4), 437-451.

Traunmüller, Hartmut; Anders Eriksson (2000). Acoustic effects of variation in vocal effort by men, women, and children. *J. Acoust. Soc. Am.* 107: 3438-3451.

