Adaptação acústico-prosódica entre falantes

Vera Cabarrão[1, 2], Helena Moniz[1, 2], Fernando Batista[1, 3],
Isabel Trancoso[1, 4], Ana Isabel Mata[2]

[1]L2F, INESC-ID, Lisboa, Portugal
[2]FLUL/CLUL, Lisboa, Portugal
[3]ISCTE-IUL – Instituto Universitário de Lisboa, Lisboa, Portugal
[4]Instituto Superior Técnico, Universidade de Lisboa, Portugal

Abstract:

This paper presents a global analysis of entrainment in map-task dialogues in European Portuguese, including 48 dialogues, between 24 speakers. Our main goal is to analyze the acoustic-prosodic similarities between speaker pairs, namely if there are global entrainment cues displayed in the dialogues, if entrainment is manifested in distinct sets of features shared amongst the speakers, if entrainment depends on the gender and role of the speaker (giver or follower), and if speakers tend to entrain more with specific interlocutors regardless of the role. Results show that globally speakers tend to be more similar to their partners than to their own speech in the majority of the analyzed features, a strong evidence for entrainment. Moreover, almost all the pairs of speakers display cues of global entrainment, even though in different degrees (speakers entrain but in distinct features). Additionally, the role and gender effects tend to be less striking than the specific interlocutor effect. Our results support the fact that all prosodic parameters are monitored by the speakers in our corpus, contrarily to studies for other languages, which indicate that the main cues are energy related.

Keywords: entrainment, acoustic-prosodic features, dialogues

Palavras-chave: adaptação, parâmetros acústico-prosódicos, diálogos

1. Introdução

A adaptação entre falantes, também designada por acomodação, sintonia ou mesmo sincronismo (do inglês *entrainment*), tem sido descrita como a capacidade de os seres humanos se ajustarem, tanto a nível comportamental como discursivo, ao seu interlocutor (Brennan & Clark, 1996; Beňuš, 2014). Esta estratégia tem sido estudada sob diversas perspetivas, quer para compreender quais os mecanismos linguísticos, psicológicos e sociais (Levitan *et al.*, 2012; Beňuš, 2014) que a motivam, quer para replicar este comportamento, tipicamente humano, em sistemas de diálogo automáticos (Levitan, 2014).

Estudos recentes mostram que a adaptação entre falantes desempenha um papel crucial tanto na resolução de tarefas, bem como na flexibilidade dos falantes, o que os caracteriza como mais atrativos e agradáveis socialmente para os seus interlocutores (Beňuš, 2014), levando mesmo a interações mais bem sucedidas. Do mesmo modo, esta estratégia tem impacto no sucesso de diálogos entre humanos e sistemas automáticos, na medida em que os falantes recorrem a estratégias semelhantes às utilizadas em diálogos humano-humano (Coulston *et al*., 2002; Levitan, 2014), para resolução de tarefas conjuntas.

Apesar de estudado em línguas como o Inglês ou Mandarim, em Português Europeu (PE), este tópico só agora começa a ser alvo de análise. O presente estudo surge, assim, como uma primeira tentativa de

compreender se esta estratégia, de facto: (i) ocorre em fala espontânea em PE, (ii) se se manifesta em propriedades acústico-prosódicas, (iii) se todos os parâmetros prosódicos são monitorizados pelos falantes ou se apenas alguns (como no Inglês) o são, (iv) se é influenciada por fatores sociolinguísticos, nomeadamente pelo papel desempenhado pelos falantes ou mesmo o interlocutor específico com quem se interage.

Como tal, o principal objetivo deste trabalho é analisar as semelhanças entre falantes ao nível do diálogo (*global entrainment*), nomeadamente se existem diferentes graus de adaptação relacionados com os parâmetros acústico-prosódicos partilhados pelos falantes, se estes dependem do género e/ou do papel desempenhado no diálogo (dador ou seguidor) e se os falantes mostram tendência para se adaptar a determinados interlocutores independentemente do papel que ambos desempenham. As pistas acústico-prosódicas utilizadas já provaram ter impacto neste tipo de tarefas (*e.g.*, duração, *pitch* (f_0), energia, medidas de articulação de fala, medidas de qualidade da voz) e permitem uma comparação entre línguas diferentes (*e.g.*, Levitan & Hirschberg, 2011; Levitan, 2014; Gravano *et al.*, 2014). Com base nestas medidas, e seguindo o trabalho de Levitan & Hirschberg (2011) e Levitan (2014), foi calculada a semelhança entre falantes num mesmo diálogo (pares de falantes), a semelhança entre falantes que não participam no mesmo diálogo (não--interlocutores) e a semelhança entre um mesmo falante em diálogos diferentes. A comparação permitirá, em última análise, discutir criticamente se todos os parâmetros prosódicos são usados no processo de adaptação entre falantes.

Este artigo está organizado da seguinte forma: a Secção 2 apresenta uma breve revisão bibliográfica de estudos sobre adaptação entre falantes, privilegiando a adaptação acústico-prosódica; a Secção 3 descreve os dados, as pistas prosódicas e as métricas utilizadas; a Secção 4 apresenta os resultados obtidos acerca da adaptação entre pares de falantes e, finalmente, a Secção 5 apresenta as conclusões obtidas e direções de trabalho futuro.

2. Estado da arte

Em diálogos espontâneos, os falantes naturalmente convergem ou divergem nas suas opiniões ao longo da interação, embora tenham também a capacidade de se adaptar ao comportamento e discurso do seu interlocutor (Grice, 1975; Giles *et al.*, 1987; Giles *et al.*, 1991; Brennan & Clark, 1996). Na Teoria da Acomodação, Giles (1991) descreve esta estratégia como um conjunto organizado e complexo de múltiplas alternativas disponíveis aos falantes nos diálogos, que lhes permitem uma aproximação ao outro ou, pelo contrário, um distanciamento. Concretamente, os falantes têm a capacidade de moldar o seu discurso e a forma como a mensagem é veiculada, de modo a se adaptarem ou não ao interlocutor.

Na maioria dos estudos sobre adaptação entre falantes, esta estratégia não é analisada *per se*, mas na sua relação, por exemplo, com o sucesso do diálogo (Nenkova *et al.*, 2008; Reitter & Moore, 2007, com variáveis sociais (Chartrand & Bargh, 1999; Beňuš *et al.*, 2012), nomeadamente um falante ser considerado mais flexível e inteligente pelo seu interlocutor, ou em relações de poder (Danescu *et al.*, 2012; Beňuš, 2014), assim como com a resolução de problemas ou tarefas específicas, tal como em diálogos em formato map-task, jogos de cartas (Levitan & Hirschberg, 2011; Gravano *et al.*, 2014); terapia de casais (Lee *et al.*, 2010) ou relações românticas (Weidman *et al.*, 2016). Os estudos apontam, assim, para uma estreita relação entre o processo de acomodação entre falantes e a vertente de socialização.

A nível linguístico, este tópico tem sido analisado sob diversas perspetivas, nomeadamente a nível acústico-prosódico (Pardo, 2006; Levitan & Hirschberg, 2011; Levitan, 2014; Gravano *et al.*, 2014) e lexical e sintático (Ward & Litman, 2007; Nenkova *et al.*, 2008; Lopes *et al.*, 2013). Há ainda estudos sobre a multimodalidade na adaptação, nomeadamente através de gestos ou expressões faciais (Chartrand & Bargh, 1999).

Pardo (2006) apresenta evidências de proximidade fonética entre falantes, ao comparar as semelhanças fonéticas e prosódicas de palavras produzidas durante o diálogo com as mesmas palavras ditas antes ou depois deste por cada um dos interlocutores individualmente. O autor também encontrou proximidade fonética no início e no fim dos diálogos, mostrando que a convergência entre falantes aumenta à medida que o diálogo flui. Quanto à variação por género e por papel desempenhado, os resultados deste estudo indicam que os falantes masculinos convergem mais do que os femininos, assim como os dadores convergem mais do que os seguidores. Num estudo perceptual mais recente, Pardo *et al.* (2010) confirmam que o género dos pares de falantes e o papel desempenhado por estes têm maior influência no grau de convergência fonética do que os atributos acústico-fonéticos de cada falante.

Estudos mais recentes (Levitan *et al.*, 2011; Levitan, 2014) analisam a adaptação entre falantes para o Inglês Americano no *Columbia Games Corpus* (Gravano, 2009). Este *corpus* compreende vários jogos colaborativos (*e.g.*, jogos de cartas, identificação de objetos) entre pares de falantes que não se conhecem. O que se pretende é que ambos colaborem para resolver os jogos, de modo a obterem um maior número de pontos e, consequentemente, uma compensação monetária no final da tarefa. A nível global, entenda-se ao nível do diálogo completo, os resultados mostraram que os falantes são mais semelhantes aos seus interlocutores do que aos falantes com quem nunca interagiram (não-interlocutores) nos parâmetros média e máximo de energia e débito de fala. Ao comparar pares de falantes e estes com o seu próprio discurso noutros diálogos, as autoras mostraram que os falantes são mais semelhantes ao seu próprio discurso do que aos seus interlocutores nos seguintes parâmetros: média de f_0, *jitter* (perturbações de f_0), *shimmer* (perturbações de energia), *Noise-to-Harmonic-Rate* (medida em dB da proporção entre um som harmónico e ruído na voz) e débito de fala. Levitan (2014) e Xia *et al.* (2014) também verificaram que falantes de Mandarim (num *corpus* de jogos colaborativos semelhante ao *corpus* do Inglês) se assemelham mais aos seus interlocutores do que aos não-interlocutores nos mesmos parâmetros dos falantes de Inglês, diferindo apenas no máximo de f_0. Os autores mostram ainda que pares de falantes de géneros diferentes (masculino-feminino) apresentam maiores graus de adaptação do que os pares masculino-masculino e feminino-feminino, em ambas as línguas.

Relativamente a variáveis sociolinguísticas, Levitan *et al.* (2012) já tinham observado, também no *corpus* de jogos, que a adaptação prosódica entre pares feminino-feminino se correlacionava com variáveis sociais, nomeadamente com a tentativa de agradar e de dar encorajamento para completar a tarefa partilhada. Tal acontecia mais nestes pares do que em pares mistos. Ainda assim, relativamente à fluidez do diálogo, esta adaptação revelou-se mais importante entre falantes do género masculino.

Seguindo esta linha de análise, Gravano *et al.* (2014) estudaram a adaptação de acentos tonais, contornos entoacionais e tons de fronteira, de acordo com o sistema de anotação ToBI (Beckman *et al.*, 2005). Os autores analisaram a relação entre estes eventos prosódicos e variáveis sociais manualmente anotadas, como por exemplo, se a conversa flui naturalmente, se os falantes se mostram aborrecidos e/ou com problemas em se compreender mutuamente, se estão concentrados na tarefa, se parecem gostar ou não do interlocutor, de que forma estão a contribuir para o sucesso da tarefa, *inter alia*. Os resultados mostraram que, quando os falantes utilizam o mesmo contorno entoacional do interlocutor, estes estão a contribuir para o sucesso da interação, pois tornam o seu discurso mais claro, dão mais encorajamento à conclusão da tarefa, mostram que colaboram com o interlocutor e que estão energicamente a jogar.

Nº 4 – 09/2018 | 18-33 | https://doi.org/10.26334/2183-9077/rapln4ano2018a28

3. Metodologia

3.1. *Corpus*

Neste estudo foi utilizado o *corpus* CORAL[1] (Trancoso *et al.*, 1998), que contém 64 diálogos em formato *map-task* entre 32 falantes. Cada diálogo ocorre entre dois falantes que desempenham papéis diferentes, dador de informação e seguidor. O primeiro dispõe de um mapa com uma rota e alguns marcos e tem como tarefa guiar o seguidor, para que este reconstrua o mesmo caminho no seu mapa incompleto (controlado de modo a desencadear estratégias de resolução colaborativas). O CORAL é equilibrado em termos de género e de papel desempenhado pelos falantes, sendo que cada um desempenha o papel de dador e de seguidor duas vezes com interlocutores diferentes. O corpus tem 7 horas de fala ortograficamente transcritas, com um total de 61 mil palavras.

A amostra utilizada neste trabalho contém 48 diálogos[2] entre 24 falantes (12 homens e 12 mulheres). Os diálogos ocorrem entre pares mistos e pares do mesmo género. O grau de familiaridade entre interlocutores varia desde pares que não se conhecem a um par de gémeas idênticas (s21 e s24) (Tabela 1). Esta amostra está dividida em constituintes similares a frases (*sentence-like units* - SU), com um total de 42 mil palavras.

Diálogos (N)	Falantes
Familiaridade	24 Conhecidos
	24 Desconhecidos
Género	16 Pares mistos
	16 Pares masculino-masculino
	16 Pares feminino-feminino

Tabela 1. Distribuição de diálogos por grau de familiaridade e género dos falantes

3.2. Parâmetros prosódicos e métricas

Para medir o grau de adaptação entre falantes, foram extraídos dois grupos de parâmetros acústico-prosódicos para cada SU. O primeiro grupo corresponde às pistas knowledge-based (Moniz *et al.*, 2014), pistas que já provaram ter impacto neste tipo de tarefas, nomeadamente duração de fala com e sem silêncios, f_0, valores normalizados de máximo, mínimo, média, mediana e desvio padrão de energia, bem como declives de f_0 e de energia. As medidas de duração contemplam o rácio de articulação (número de fones ou sílabas por duração de fala com e sem silêncios internos); rácio de fala (número de fones ou sílabas dividido pela duração da fala com e sem silêncios internos) e rácio de fonação (100% da duração da fala sem silêncios internos divididos pela duração da fala incluindo os silêncios internos). O segundo grupo corresponde a pistas tipicamente adotadas em tarefas paralinguísticas, as eGeMAPS (Extended Geneva Minimalistic Acoustic

[1] O corpus CORAL está disponível no ELRA *Catalogue of Language Resources*, com a referência ELRA - S0367.
[2] Os restantes 16 diálogos não puderam ser utilizados devido ao facto de o áudio não estar disponível em canais separados, o que impossibilita a extração de parâmetros acústico-prosódicos por falante.

Nº 4 – 09/2018 | 18-33 | https://doi.org/10.26334/2183-9077/rapln4ano2018a28

Parameters for Voice Research and Affective Computing), apresentada por Eyben *et al.* (2016). Estas consistem num conjunto de 88 parâmetros acústico-prosódicos que incluem: pistas de f_0, energia e parâmetros de voz (*jitter*, *shimmer* – medidas de aspereza na voz) com valores absolutos (designadas por Eyben *et al.* (2010) de descritores de baixo nível) e funcionais extraídos a partir desses valores, ou seja, cálculos aplicados sobre os valores absolutos (estatísticas, regressão polinomial, transformadas), dos quais resultam, por exemplo, os parâmetros slopeV0-500 e slopeUV0-500 - declives de f_0 em regiões vozeadas e não vozeadas entre duas bandas. Para cada SU, os valores de média de cada parâmetro foram calculados por falante em cada diálogo.

Com base nestas medidas, e seguindo o trabalho de Levitan & Hirschberg (2011) e Levitan (2014), foi calculada a semelhança acústico-prosódica entre falantes num mesmo diálogo (pares de falantes – Equação 1), a semelhança entre falantes que não participam no mesmo diálogo (não-interlocutores – Equação 2) e a semelhança entre um mesmo falante em diálogos diferentes (Equação 3). Nestas equações, *s* refere-se ao falante de um diálogo, *s_f* corresponde à média de cada parâmetro acústico-prosódico desse falante no diálogo em causa e *n* é o número de não-interlocutores. Na Equação 1, *_f* refere-se à média dos parâmetros do interlocutor no diálogo, na 2, *_f* remete para a média dos parâmetros de um falante que não o do mesmo diálogo e na 3, *_f* diz respeito à média dos parâmetros do mesmo falante, mas num diálogo diferente. Adicionalmente, e uma vez que todos os falantes desempenham o papel de dador e de seguidor duas vezes com interlocutores diferentes, também é possível observar se se comportam de modo distinto tendo em conta o papel que desempenham e o interlocutor com quem interagem. Em suma, a Equação 1 permite extrair os valores de acomodação entre pares, a Equação 2 permite analisar a outra face da moeda, com quem não se interage, e a Equação 3 pode ser considerada uma medida de limiar para se entender quão similar é um falante ao longo de toda a sua participação no *corpus*.

Em suma, depois de calculadas as equações, foi obtido um valor de semelhança para cada parâmetro acústico-prosódico entre pares de falantes (Equação1), não-interlocutores (Equação 2) e o mesmo falantes em diálogos diferentes (Equação 3). Estes mesmos valores foram depois comparados com um *paired t-test*. Este teste estatístico permite verificar a existência de diferenças estatisticamente significativas entre os valores obtidos de cada equação. Após determinar se há diferenças estatisticamente significativas entre grupos (por exemplo, pares vs. não pares; dadores vs. seguidores), a polaridade do valor t indica em que grupo as semelhanças entre falantes são maiores ou menores. Desta forma, um valor negativo indica que dentro do primeiro grupo há maiores semelhanças entre falantes, enquanto um valor positivo indica que essa semelhança é maior no segundo grupo em análise. Assim, é possível comparar os dois grupos e observar em qual deles e em que parâmetros os falantes são mais semelhantes entre si.

$$(1) \qquad ENT(s,f) = -|s_f - s_f^i|$$

$$(2) \qquad ENTX(s,f) = -\frac{\sum_{x=0}^{n-1} |s_f - s_f^x|}{n}$$

$$(3) \qquad ENT_{self}(s,f) = -|s_f - s_f'|$$

Nº 4 – 09/2018 | 18-33 | https://doi.org/10.26334/2183-9077/rapln4ano2018a28

4. Resultados

Uma nota prévia sobre a apresentação dos resultados. Por questões de legibilidade, não é possível apresentar os resultados para todos os parâmetros acústico-prosódicos analisados, tendo sido feita uma seleção dos mesmos em função da sua significância. A presente secção dará conta da distribuição dos resultados relativos à adaptação global, de forma genérica, e a factores sociolinguísticos, em particular do papel desempenhado pelos falantes, género do falante, falante específico com quem se interage.

4.1. Adaptação global

Na comparação entre pares de falantes num mesmo diálogo e falantes que não interagem, os resultados mostram evidências de adaptação global entre os primeiros nos parâmetros declive de f_0, duração de fala, com e sem silêncio interno, e rácio de fonação. Nestes parâmetros, observam-se diferenças estatisticamente significativas ($p<0,05$ representado com ** e $p<0,01$ representado com *) entre os pares e os não-pares de falantes (Tabela 2), bem como valores de *t* que mostram que os primeiros apresentam menos diferenças entre si do que os segundos. Numa análise aos diálogos do *corpus* CORAL, Moniz *et al.* (2014) já haviam observado que os falantes não apresentavam diferenças significativas de débito de fala, produzindo sempre a mesma média de sílabas (6 sílabas por segundo). A título de exemplo, observem-se as Figuras 1 e 2, duas interações em que foi detetada adaptação acústico-prosódica entre falantes. Em ambos os exemplos, ilustra-se a natureza dos dados, nomeadamente pares pergunta-resposta e trocas de informações sobre a posição no mapa. Nestes dados, quer o dador quer o seguidor produzem interações curtas (respostas afirmativas ou negativas) ou longas (instruções, informações), pelo que é possível equacionar que há uma medida de diálogo aceitável em termos do parâmetro duração e que os falantes tendem a respeitá-la.

Parâmetros		Pares vs. não-pares	Pares vs. mesmos falantes
		t/Sig.	t/Sig.
f_0	f_0_amean	4,302 *	-7,890 *
	f_0_pctlrange0_2	0,963	-2,940 *
	f_0_meanRisSlope	-2,775 *	-3,745 *
	f_0_meanFallSlope	1,279	-2,406 *
	slopeV0_500	4,306 *	-5,864 *
	slopeV500_1500	1,950 **	-6,373 *
	slopeUV0_500	5,657 *	-1,070
	slopeUV500_1500	1,605	-5,389 *

Nº 4 – 09/2018 | 18-33 | https://doi.org/10.26334/2183-9077/rapln4ano2018a28

Energia	loudness_amean	9,612 *	1,862
	loudness_pctlrange0_2	5,169 *	-0,679
	loudness_meanRisSlope	6,725 *	0,522
	loudness_meanFallSlope	1,955 **	-2,934 *
	loudnessPeaksPerSec	2,139 **	-5,833 *
Qualidade de voz	Jitter	1,684	-2,675 *
	Shimmer	2,907 *	-5,474 *
	HNR	3,059 *	-7,886 *
Vozeamento/ Desvozeamento	VoicedSegmentsPerSec	-0,872	-3,872 *
	MeanVoicSegLengthSec	2,942 *	-5,650 *
	MeanUnvoiSegLength	2,834 *	-1,276
Duração	dur speech (with sil)	-8,195 *	-15,001 *
	dur (without sil)	-8,645 *	-15,204 *
	articulation rate phone	-1,497	-6,736 *
	rate of speech phone	-0,415	-5,754 *
	phonation ratio	-2,019 **	-6,750 *
	articulation rate syl	-0,058	-4,703 *
	rate of speech syl	0,120	-4,808 *

Tabela 2. *T-tests*: diferenças entre pares vs. não-pares (df = 95) e pares vs. mesmos falantes (df = 95)

Os resultados obtidos não vão totalmente ao encontro dos de Levitan (2014) para o Inglês e Mandarim, na medida em que os parâmetros energia (média e máximo) e débito de fala são os que apresentam maiores semelhanças entre falantes. No presente estudo, é a duração o parâmetro com mais evidências de adaptação global, o que leva a equacionar a hipótese de que esta estratégia pode ocorrer de forma mais automática, e independentemente da língua, neste parâmetro, mas não em pistas como a energia ou f_0.

Na comparação entre pares de falantes *vs.* mesmos falantes com o seu próprio discurso noutro diálogo, os resultados mostram maiores semelhanças entre interlocutores na maioria dos parâmetros analisados, o que prova que há uma forte adaptação acústico-prosódica. Apenas nos parâmetros de energia, os falantes são mais

Nº 4 – 09/2018 | 18-33 | https://doi.org/10.26334/2183-9077/rapln4ano2018a28

próximos ao seu próprio discurso. Estes resultados diferem dos obtidos por Levitan (2014), na medida em que a autora observou mais semelhanças entre interlocutores nas pistas média e máximo de energia, sendo que nos restantes parâmetros (f_0 e qualidade de voz) estes se aproximavam mais do seu próprio discurso.

Assim, é possível concluir que há evidências de adaptação entre falantes, mas expressa em diferentes graus: os falantes são mais semelhantes aos seus interlocutores do que ao seu próprio discurso na maioria dos parâmetros acústico-prosódicos e, na comparação entre pares e não-pares, esta adaptação verifica-se maioritariamente em parâmetros de duração.

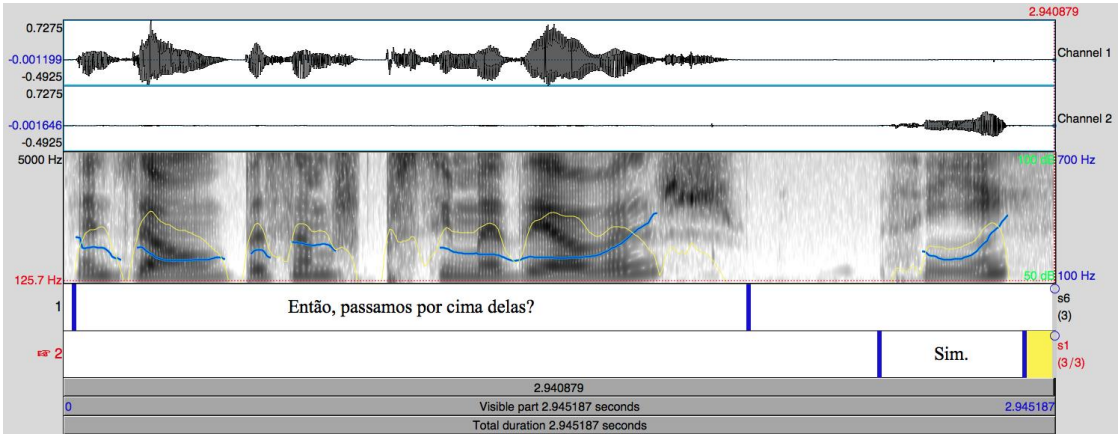

Figura 1. Exemplo de adaptação entre os falantes s6-s1

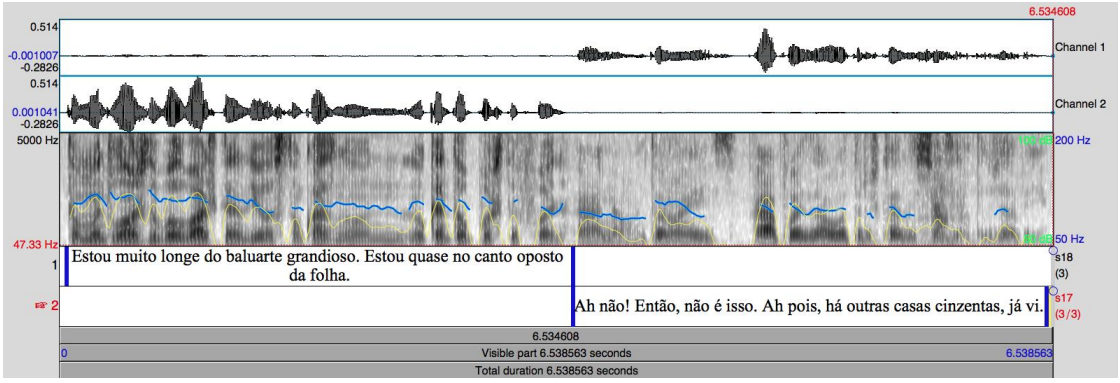

Figura 2. Exemplo de adaptação entre os falantes s18-s17

4.2. Adaptação global por papel desempenhado

Quanto à comparação dos falantes a desempenhar diferentes papéis (dador *vs.* seguidor), apenas os parâmetros *slope UV0_500* (declive de f_0) e duração sem silêncio interno apresentam diferenças estatisticamente significativas, sendo os falantes mais semelhantes a si próprios no papel de dadores da informação (Tabela 3). Ainda assim, os resultados também mostram que os dadores tendem a ser mais semelhantes em parâmetros relacionados com energia e declives de f_0, enquanto os seguidores apresentam

Nº 4 – 09/2018 | 18-33 | https://doi.org/10.26334/2183-9077/rapln4ano2018a28

menores diferenças em medidas de qualidade da voz e duração. Ao analisar cada falante individualmente, verifica-se que não há uma influência clara do papel desempenhado no comportamento dos falantes, na medida em que cerca de metade é mais consistente no seu discurso enquanto dador (N=11) e a outra metade como seguidor (N=13). Estes resultados levam-nos a questionar se a adaptação entre falantes poderá estar mais relacionada com a ligação ao interlocutor do que propriamente com o papel desempenhado no diálogo, *vide* secção 4.3. para um desenvolvimento sobre esta questão.

Parâmetros		Dador vs. seguidor
		t/Sig.
f0	*f0*_amean	0,869
	*f0*_pctlrange0_2	0,875
	*f0*_meanRisSlope	-0,676
	*f0*_meanFallSlope	-0,352
	slopeV0_500	-0,842
	slopeV500_1500	-1,852
	slopeUV0_500	-2,471 **
	slopeUV500_1500	0,105
Energia	loudness_amean	-0,720
	loudness_pctlrange0_2	-1,007
	loudness_meanRisSlope	-0,527
	loudness_meanFallSlope	-1,693
	loudnessPeaksPerSec	0,738
Qualidade de voz	jitter	0,572
	shimmer	0,966
	HNR	1,355
Vozeamento/ Desvozeamento	VoicedSegmentsPerSec	1,807
	MeanVoicSegLengthSec	-0,115
	MeanUnvoiSegLength	0,677

Nº 4 – 09/2018 | 18-33 | https://doi.org/10.26334/2183-9077/rapln4ano2018a28

Duração	dur speech (with sil)	-1,978
	dur (without sil)	-2,279 **
	articulation rate phone	0,785
	rate of speech phone	1,011
	phonation ratio	0,089
	articulation rate syl	1,200
	rate of speech syl	1,916

Tabela 3. *T-tests*: diferenças entre dador vs. seguidor (df = 23)

4.3. Adaptação face à variável interlocutor

No corpus CORAL, todos os falantes interagem com dois interlocutores diferentes, duas vezes a desempenhar o papel de dador e outras duas, o de seguidor. Como tal, é possível observar se um falante se adapta mais a um determinado interlocutor do que a outro e qual a influência do papel desempenhado nessa relação. Para medir o grau de adaptação entre pares de falantes, foi utilizada a medida de semelhança entre interlocutores (Equação 1), tendo sido calculada a diferença entre cada par. Se um par é mais semelhante num maior número de parâmetros acústico-prosódicos, então esse mesmo par mostra um maior grau de adaptação.

Os resultados mostram que 10 falantes (s1; s4; s6; s8; s14; s17; s18; s21; s22; s24) apresentam maiores semelhanças com o mesmo interlocutor, independentemente do papel desempenhado, ainda que tal não ocorra sempre nos mesmos parâmetros (na Tabela 3, o X representa os parâmetros em que cada par apresenta maiores semelhanças). Estes resultados evidenciam que, apesar de alguns falantes mostrarem mais sensibilidade a determinados interlocutores, a adaptação acústico-prosódica não se manifesta sempre nos mesmos parâmetros.

Quanto aos restantes falantes, estes apresentam um comportamento semelhante com mais do que um interlocutor num número aproximado de parâmetros. É o caso do falante s12 que, enquanto dador, apresenta semelhanças com s13 e com s2, no mesmo número de pistas acústico-prosódicas. Apenas dois falantes (s15 e s20) se aproximam de um único interlocutor consoante o papel desempenhado (s15 assemelha-se a s16 enquanto dador e s20, como dador, é mais semelhante a s15 e, como seguidor, a s10).

Um outro aspeto relevante nesta análise é o facto de alguns falantes só se terem conhecido na gravação dos dados, enquanto outros já se conheciam previamente, ou como colegas de universidade, laboratório, ou como amigos, havendo até um par de gémeas idênticas. Considerando o grau de familiaridade entre os falantes, observa-se que, de entre os 10 pares que mostram mais adaptação em ambos os diálogos em que participam, 6 eram desconhecidos (s1-s17; s6-s18; s14-s22; s12-s13; s15-s16; s10-s20) e apenas 4 (s4-s8; s21-s24; s12-s3; s20-s15) já se conheciam previamente. Não foram considerados aqui os falantes que mostram adaptação a mais do que um interlocutor diferente, pois esses mostram semelhanças independentemente do grau de familiaridade.

Estes resultados levam-nos a colocar a hipótese de que os falantes despendem um maior esforço em ajustar o seu discurso a estranhos do que a alguém que já lhes é familiar. Quando há uma relação prévia entre interlocutores, já foi estabelecido, em algum momento, um conhecimento comum (*common ground*) de como ambos se costumam comportar, sendo talvez por isso menos monitorizado o fluir do diálogo. Um claro

exemplo são os diálogos das irmãs gémeas, os mais breves de todo o *corpus* a atingir o objetivo com sucesso: no diálogo entre s24-s21, a tarefa é concluída com apenas 36 SU da parte da dadora e 11 da seguidora (Figura 3); no segundo diálogo, com os papéis inversos, há apenas 30 e 38 SU, respetivamente. Estas falantes, que já se conhecem tão bem, não necessitam de falar muito para completar a tarefa e serem bem sucedidas, o que aponta para um nível máximo de adaptação pré-existente entre elas.

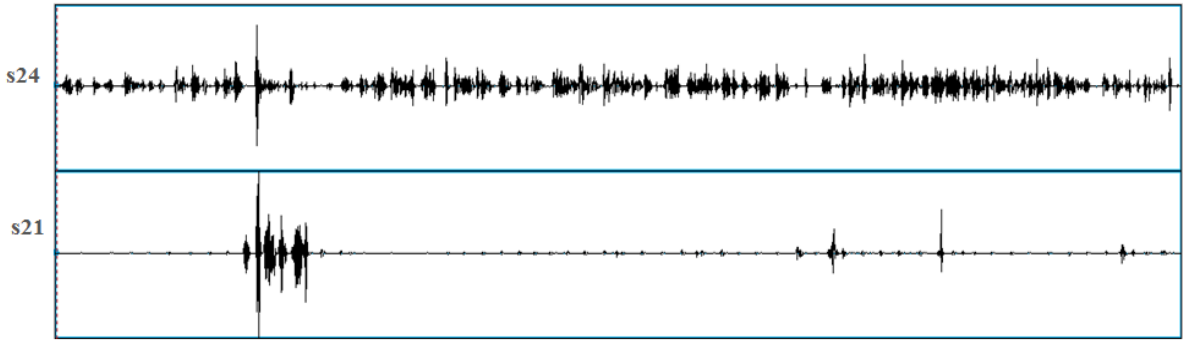

Figura 3. Exemplo de diálogo entre as irmãs gémeas (s24 é a dadora e s21, a seguidora)

Ainda assim, e tendo em conta o seu grau de familiaridade, era expectável encontrar uma clara adaptação acústico-prosódica entre ambas na maioria das pistas analisadas. Nos quatros diálogos em que as irmãs participam, observa-se que há pequenas diferenças entre elas em 66% dos parâmetros analisados quando comparadas com os outros dois falantes com quem interagem, nomeadamente s11 e s9. Quanto às pistas acústico-prosódicas, observa-se na Tabela 3 que estas são semelhantes principalmente em pistas relacionadas com níveis de f_0. Observa-se ainda que há mais semelhanças entre elas quando s24 é a dadora e s21, a seguidora, ainda que estas não sejam significativas (s24-s21 são mais semelhantes em 52% das pistas e s21-s24, em 48%).

Parâmetros		s1	s17	s4	s8	s6	s18	s14	s22	s21	s24
		s17	s1	s8	s4	s18	s6	s22	s14	s24	s21
f_0	f_0_amean	X	X	X	X	X	X	X	X	X	X
	f_0_pctlrange0_2		X		X				X	X	X
	f_0_meanRisSlope	X	X		X	X		X		X	X
	f_0_meanFallSlope		X	X	X	X				X	X
	slopeV0_500	X	X	X	X	X	X	X	X	X	
	slopeV500_1500		X	X		X				X	
	slopeUV0_500		X	X	X	X		X	X	X	
	slopeUV500_1500		X	X		X		X		X	X

Nº 4 – 09/2018 | 18-33 | https://doi.org/10.26334/2183-9077/rapln4ano2018a28

Energia	loudness_amean	X	X	X	X	X					
	loudness_pctlrange0_2	X			X	X		X	X	X	
	loudness_meanRisSlope				X	X		X	X	X	X
	loudness_meanFallSlope	X	X		X	X			X	X	X
	loudnessPeaksPerSec	X		X	X	X	X	X	X		X
Qualidade De voz	Jitter		X	X		X	X	X	X		X
	Shimmer	X	X	X	X	X	X	X	X		X
	HNR	X	X	X	X	X	X	X	X		X
Vozeamento/ Desvozeamento	VoicedSegmentsPerSec		X			X		X		X	X
	MeanVoicSegLengthSec		X	X	X	X		X	X	X	X
	MeanUnvoiSegLength	X	X	X	X	X	X		X		X
Duração	dur speech (with sil)	X			X	X	X	X	X		X
	dur (without sil)	X			X	X	X		X		X
	articulation rate phone		X		X	X	X		X	X	X
	rate of speech phone		X		X	X		X	X	X	X
	phonation ratio			X	X		X	X	X	X	X
	articulation rate syl	X		X	X	X		X	X		X
	rate of speech syl	X			X	X		X	X		X

Tabela 3. Adaptação por pares de falantes (dador-seguidor). O X marca os parâmetros com maiores semelhanças.

Conclui-se então que, independentemente do papel desempenhado pelos falantes, estes tendem a adaptar-se a determinados interlocutores. Como tal, coloca-se a hipótese de existirem graus de monitorização da produção dos falantes, na medida em que o esforço de adaptação é ajustado consoante o interlocutor. Deste modo, coloca-se também a hipótese de que, no *corpus* em causa, a influência do interlocutor na acomodação acústico-prosódica é maior do que a do papel desempenhado.

No que diz respeito ao género dos falantes, os resultados mostram ainda que pares mistos tendem a apresentar maiores semelhanças entre si do que pares do mesmo género, independentemente do papel no diálogo: há quatro pares mistos nestas circunstâncias e apenas um par feminino-feminino (Tabela 4). Os falantes que mostram adaptação apenas a um interlocutor correspondem a dois pares masculino-masculino e os que mostram semelhanças a mais do que um falante correspondem, na maioria, a pares mistos. Estes resultados vão ao encontro dos obtidos para Inglês e Mandarim (Levitan, 2014), em que os pares mistos apresentam maiores graus de adaptação prosódica.

Pares Mistos	M-M	F-F
S4-S8	S10-S16	**S21-S24**
S1-S17	S16-S10	S2-S14
S6-S18	S20-S15	S9-S11
S14-S22	S7-S4	S11-S9
S3-S12	S10-S20	S24-S21
S5-S7	S12-S13	S9-S24
S7-S5	S13-S12	S11-S21
S8-S4	S15-S16	
S12-S3		
S13-S23		
S23-S13		
S2-S19		
S17-S1		
S18-S6		
S19-S2		
S22-S14		

Tabela 4. Distribuição dos pares de falantes com mais adaptação acústico-prosódica por género. Os falantes que se adaptam entre si independentemente do papel desempenhado estão destacados a negrito.

5. Conclusão

Este trabalho representa uma primeira análise do processo de adaptação entre falantes em diálogos (semi-)espontâneos em PE. Com base em medidas de proximidade entre parâmetros acústico-prosódicos de diferentes falantes, pretendeu-se aferir se estes se adaptam às propriedades de fala dos seus interlocutores. Adicionalmente, investigou-se a influência de variáveis sociolinguísticas, como género dos falantes, papel que desempenham no diálogo ou mesmo o interlocutor específico com quem interagem, na (in)existência de adaptação entre interlocutores e/ou, num olhar mais de perto, no grau de adaptação entre interlocutores.

Os resultados mostram que há evidências de adaptação nos diálogos, embora expressa em diferentes graus. Os falantes são mais semelhantes aos seus interlocutores do que ao seu próprio discurso noutros diálogos na maioria dos parâmetros acústico-prosódicos, sendo a energia o único parâmetro inalterado. Já na comparação entre pares e não-pares, esta adaptação verifica-se maioritariamente em parâmetros de duração.

O estudo mostra também que todos os parâmetros prosódicos (f_0, energia, duração, qualidade de voz) são monitorizados no processo de adaptação entre falantes, evidenciando um resultado para o PE que se diferencia dos obtidos por Levitan (2014), para o Inglês e Mandarim, uma vez que a autora observou adaptação global entre interlocutores principalmente em pistas de energia, quer na comparação com não-pares, quer com o mesmo falante noutro diálogo.

Relativamente ao papel desempenhado, novamente o estudo mostra um comportamento distinto do que é reportado por Pardo *et al.* (2010), uma vez que não há uma clara tendência para os falantes a desempenhar o papel de dadores se adaptarem mais aos seguidores. Observa-se que cerca de metade dos falantes é mais consistente no seu discurso enquanto dador (N=11) e a outra metade como seguidor (N=13). Tal pode dever-se ao facto de se tratar de um diálogo colaborativo, em que os falantes trabalham em conjunto para ser bem sucedidos na tarefa em causa. Estes resultados permitem também equacionar o facto de a adaptação entre falantes estar mais relacionada com o interlocutor, a sua postura e personalidade, do que com o papel que este desempenha no diálogo.

Esta hipótese é, de certa forma, confirmada pela comparação de diálogos em que os mesmos falantes interagem com diferentes interlocutores a desempenhar ambos os papéis. Aqui, os resultados mostram que 10 falantes apresentam semelhanças com o mesmo interlocutor, independentemente do papel desempenhado, ainda que tal não ocorra sempre nos mesmos parâmetros. Os restantes falantes apresentam um comportamento semelhante com mais do que um interlocutor num número aproximado de parâmetros e apenas dois falantes se aproximam apenas de um interlocutor consoante o papel desempenhado.

Quanto ao género, apurou-se que há uma maior tendência para adaptação entre pares mistos do que entre pares feminino-feminino e masculino-masculino, tal como verificado para o Inglês e para o Mandarim (Levitan, 2014).

Considerando que, nos dados analisados, os falantes interagem com interlocutores que já conheciam, bem como com desconhecidos, procurou-se aferir também se a adaptação acústico-prosódica está relacionada com estes graus de familiaridade. Os resultados evidenciam que, dos 10 pares com mais adaptação em ambos os diálogos em que participam, 6 eram desconhecidos e apenas 4 já se conheciam previamente. Os autores do presente estudo consideram que os falantes fazem um esforço adicional para monitorizarem o seu discurso e se ajustarem a alguém que não conhecem, de forma a que o diálogo flua com sucesso. Os diálogos em que as irmãs gémeas participam revelam isso mesmo, na medida em que já há uma adaptação prévia num grau máximo, o que leva a que o percurso seja concluído com o menor número de interações de todo o *corpus* (um número bastante reduzido de apenas 11 como dador). Ainda assim, dado que os valores em causa (6 pares de falantes desconhecidos vs. 4 pares que se conheciam previamente) não são assaz díspares, os autores procurarão verificar se o resultado se mantém noutros *corpora*.

Este estudo mostra que a adaptação entre falantes em PE é um processo que implica a monitorização de todos os parâmetros acústico-prosódicos, embora com gradiência no uso dos respetivos parâmetros. Evidencia também que os falantes se adaptam a interlocutores específicos, independentemente do grau de familiaridade ou mesmo do papel desempenhado no diálogo.

Em jeito de síntese, os dados analisados correspondem a fala espontânea e mostram que a adaptação é um processo linguístico, suportado em parâmetros prosódicos que evidenciam padrões regulares a nível global. O processo de adaptação é influenciado por distintos fatores sociolinguísticos, papel desempenhado pelo falante, interlocutor específico com quem se interage e até mesmo o género do falante. O facto de a adaptação ser mais consistente com o falante específico com quem se interage do que propriamente com a familiaridade com o interlocutor poderá ser interpretado como um grau de monitorização superior. Resta saber se não terão também de ser considerados os traços de personalidade do interlocutor, visto tratar-se de um processo social e socializante, que rotula mesmo os interlocutores como mais flexíveis e inteligentes.

Num trabalho futuro, pretende-se explorar se a adaptação ao nível do diálogo e entre pares de falantes aqui encontrada também ocorre localmente, ou seja, enunciado a enunciado. Pretende-se ainda observar se essa adaptação local está relacionada com o tipo de frases, interrogativas ou declarativas, bem como com as estruturas que ocorrem no início do segundo enunciado (marcadores discursivos, constituintes afirmativos e disfluências).

Referências

Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel (2005) The original ToBI system and the evolution of the ToBI framework. In Sun-Ah (Eds) *Prosodic Typology – The Phonology of Intonation and Phrasing*, pp. 9-54.

Beňuš, Štefan (2014) Social aspects of entrainment in spoken interaction. *Cognitive Computation*, 6(4), pp. 802-813.

Beňuš, Štefan, Rivka Levitan & Julia Hirschberg (2012) Entrainment in spontaneous speech: the case of filled pauses in Supreme Court hearings. In *3rd IEEE Conference on Cognitive Infocommunications*.

Brennan, Susan E. & Herbert H. Clark (1996) Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22 (6), 1482-1493.

Chartrand, Tanya L. & John A. Bargh (1999) The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology* 76 (6), pp. 893-910.

Coulston, Rachel, Sharon Oviatt & Courtney Darves (2002) Amplitude convergence in children's conversational speech with animated personas. In *Seventh International Conference on Spoken Language Processing*, pp. 2689-2692.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang & Jon Kleinberg (2012) Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web*, pp. 699-708.

Eyben, Florian, Klaus R. Scherer, Björn W. Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y. Devillers *et al.* (2016) The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. In *IEEE Transactions on Affective Computing 7(2)*, pp. 190-202.

Eyben, Florian, Martin Wöllmer & Björn Schuller (2010) Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pp. 1459-1462.

Giles, Howard, Anthony Mulac, James J. Bradac, & Patricia Johnson (1987) Speech accommodation theory: The first decade and beyond. *Annals of the International Communication Association* 10(1), pp. 13-48.

Giles, Howard, Nikolas Coupland & Justine E. Coupland (1991) Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics* (1), pp. 1-68.

Gravano, Agustin (2009) *Turn-taking and affirmative cue words in task-oriented dialogue*. Dissertação de doutoramento, Universidade da Columbia.

Gravano, Agustín, Štefan Beňuš, Rivka Levitan & Julia Hirschberg (2014) Three ToBI-based measures of prosodic entrainment and their correlations with speaker engagement. In *Spoken Language Technology Workshop (SLT)*, IEEE, pp. 578-583.

Grice, Paul (1975) Logic and conversation. In Maite Ezcurdia, & Robert J. Stainton (eds.) *The semantics-pragmatics boundary in philosophy*. Broadview Press, pp. 41-58.

Lee, Chi-Chun, Matthew Black, Athanasios Katsamanis, Adam C. Lammert, Brian R. Baucom, Andrew Christensen, Panayiotis G. Georgiou & Shrikanth S. Narayanan (2010) Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In *Proceedings of Interspeech 2010*, pp. 793-796.

Levitan, Rivka (2014) *Acoustic-prosodic entrainment in human-human and human-computer dialogue*. Dissertação de doutoramento, Universidade da Columbia.

Levitan, Rivka & Julia Hirschberg (2011) Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Interspeech 2011*, pp. 3081-3084.

Levitan, Rivka, Agustín Gravano, Laura Willson, Štefan Beňuš, Julia Hirschberg & Ani Nenkova (2012) Acoustic-prosodic entrainment and social behaviour. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language*, pp. 11-19.

Lopes, Jose, Maxine Eskenazi & Isabel Trancoso (2013) Automated two-way entrainment to improve spoken dialog system performance. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8372-8376.

Moniz, Helena, Fernando Batista, Ana Isabel Mata & Isabel Trancoso (2014) Speaking style effects in the production of disfluencies. *Speech Communication* (65), pp. 20-35.

Nenkova, Ani, Agustin Gravano & Julia Hirschberg (2008) High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies*, Association for Computational Linguistics, pp. 169-172.

Pardo, Jennifer S. (2006) On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119(4), pp. 2382-2393.

Pardo, Jennifer S., Isabel Cajori Jay & Robert M. Krauss (2010) Conversational role influences speech imitation. *Attention, Perception, & Psychophysics* 72(8), pp. 2254-2264.

Reitter, David & Johanna D. Moore (2007) Predicting success in dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 808-815.

Trancoso, Isabel, Maria do Céu Viana, Inês Duarte & Gabriela Matos (1998) Corpus de Diálogo CORAL. In *PROPOR'98*, Porto Alegre, Brasil.

Xia, Zhihua, Rivka Levitan & Julia Hirschberg (2014) Prosodic Entrainment in Mandarin and English: A Cross-Linguistic Comparison. In *Proceedings of Speech Prosody*, pp. 65-69.

Ward, Arthur & Diane Litman (2007) Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In *Workshop on Speech and Language Technology in Education*.

Weidman, Sarah, Mara Breen & Katherine C, Haydon (2016) Prosodic speech entrainment in romantic relationships. In *Proceedings of Speech Prosody*, pp. 508-512.