

Análise qualitativa da consoante /p/ produzida por falantes bilíngues de português e inglês

Qualitative analysis of the /p/ consonant by bilingual speakers of Portuguese and English

Otávio Augusto Rodrigues Bernado-Silva

Universidade Federal do Paraná

Denise Cristina Kluge

Universidade Federal de Santa Catarina

Maria Lúcia de Castro Gomes

Universidade Tecnológica Federal do Paraná

Otávio Augusto Rodrigues Bernado-Silva

Mestre em Linguística pela UFPR e Doutorando em Linguística pela UFPR/UIC. <https://orcid.org/0009-0000-6817-9164>

Denise Cristina Kluge

Doutora em Letras (UFSC) e docente da Universidade Federal do Rio de Janeiro e da Universidade Federal do Paraná. <https://orcid.org/0000-0003-4656-7902>

Maria Lúcia de Castro Gomes

Docente aposentada da Universidade Tecnológica Federal do Paraná (UTFPR). <https://orcid.org/0000-0002-6375-936X>

Recebido em:
19/12/2022

Aceito em:
28/06/2023

MAI / AGO 2023
ISSN 2317-9945 (ON-LINE)
ISSN 0103-6858
P. 19-32

RESUMO

A presente pesquisa buscou averiguar a viabilidade de se analisar qualitativamente produções de consoantes /p/ por falantes bilíngues de português e inglês em um contexto de comparação de locutores. Partindo de uma visão de língua como Sistema Dinâmico Complexo (SDC) e adotando pressupostos teóricos que vão desde a interdependência de seus agentes até a emergência de comportamentos e a fronteira entre o caótico e o não-caótico, busca-se comparar as produções de duas informantes diferentes utilizando espectrogramas. Após análises quantitativas realizadas em um trabalho anterior, foram selecionadas uma informante estadunidense (A3) e uma informante brasileira (B15) para que suas produções das palavras “*pie*” e “*pai*” fossem comparadas visualmente, através das observações de formas de onda, barras de vozeamento, amplitude de ondas. As comparações foram feitas na perspectiva interfalante e intrafalante, além de serem comparadas com os dados quantitativos anteriormente obtidos. Essa abordagem qualitativa, aliada à quantitativa, parece poder contribuir para o ramo da Comparação de Locutores, uma vez que a análise do período de oclusão e do VOT de um indivíduo podem ajudar a individualizar o locutor em questão.

PALAVRAS-CHAVE

Bilinguismo. Análise acústica. Consoantes oclusivas. Comparação de Locutor. VOT

ABSTRACT

The present research aims to verify whether there is a reason to analyze the productions of /p/ consonants in a qualitative manner, produced by bilin-

gual speakers of English and Portuguese in a Speaker Comparison context. Departing from a Complex Dynamic Systems (CDS) view of language and adopting the theoretical background of CDSs, such as the interdependence between agents and emergence of behaviors, among others, it is aimed to compare the productions of two speakers. After a quantitative analysis carried out in a previous work, we selected an American speaker (A3) and a Brazilian speaker (B15) to compare them, through a closer examination of the spectrograms, focusing on features like differences in wave shape, voicing bars, and amplitude of the acoustic waves, as well as a comparison with the quantitative data. The current approach used in this research might provide data for the Speaker Comparison field, as the analysis of the occlusion and Voice Onset Time (VOT) periods of a speaker may help in individualizing the speaker in question.

KEYWORDS

Bilingualism; Acoustic analysis; Stop Consonants; Speaker Comparison; VOT.

1. Introdução

O presente trabalho se propõe a expandir a discussão qualitativa acerca dos dados analisados na dissertação de mestrado do primeiro autor, uma pesquisa orientada pelas coautoras. Partindo de uma visão de língua como Sistema Dinâmico Complexo (SDC), procuramos averiguar a variabilidade da língua para além das visões tradicionais, que simplificam a linguagem como algo previsível e equacionável de modo exato. A opção por nos debruçar sobre espectrogramas de informantes individuais visa observar detalhes que podem ser ignorados ao se considerar apenas as médias gerais de um grupo de informantes. Essa atenção especial aos detalhes pode ser de grande valor para áreas como a Comparação de Locutor (CL), que é um ramo que busca selecionar as características acústicas da fala com o intuito de fornecer um perfil de indivíduos e suas eventuais características individualizantes (NOLAN, 2001; HOLLIEN, 2002). O objetivo principal desta pesquisa é verificar se a análise qualitativa dos espectrogramas corrobora os dados estatísticos e as medições feitas por software, além de verificar se essa análise é condizente com os pressupostos teóricos dos Sistemas Dinâmicos Complexos (SDCs).

O texto apresenta, primeiramente, as características dos SDCs, para introduzir, em seguida, um breve histórico dos ramos da Fonética Forense e da CL. Em seguida, são apresentadas as características articulatórias e acústicas da consoante /p/ e da metodologia do trabalho. Por fim, são discutidos os resultados e apresentadas as considerações finais.

2. Os Sistemas Dinâmicos Complexos

A Teoria dos Sistemas Dinâmicos Complexos (TSDC) teve origem nas ciências físicas e matemáticas (BARANGER, 2000; HOLLAND, 2006), mas logo se observou sua aplicabilidade em diversas áreas do conhecimento, como

economia, fisioterapia e, como apresentado aqui, também na linguística. Baranger (2000) explica que, na área da física, o paradigma do cálculo de Newton e Leibniz já estava consolidado e considerado irrefutável, o que fez com que muitos teóricos e alunos não enxergassem a possibilidade de um paradigma adjacente, o dos sistemas complexos. Baranger (2000) exemplifica os sistemas complexos com fractais: figuras geométricas que se repetem indefinidamente à medida que são analisadas cada vez mais de perto. Elementos naturais como o corpo humano, colônias de insetos e as ondas do mar também podem ser observados com uma complexidade irreduzível. Holland (2006) argumenta que cada sistema complexo é formado por agentes interdependentes que atuam em paralelo em uma rede constante de comunicação. Adaptações, emergências e evoluções podem ser observadas nos SDCs. Tomando como exemplo a língua humana, podemos observar milhões de falantes de uma mesma língua enviando e recebendo sinais o tempo todo, alterando gradualmente o idioma falado, inserindo novas palavras no léxico e abandonando outras, e mudando a gramática ao longo dos séculos. Baranger (2000) também lista algumas características necessárias de um SDC, como a relação entre o caos e o não caos (um SDC nunca é totalmente caótico, pois, se fosse, não haveria ordem alguma, mas sim opera na fronteira, uma vez que padrões imprevisíveis podem emergir), relações entre cooperação e competição, e a interação não linear entre os agentes. Chan (2001) também adiciona algumas características aos SDCs, como controle distribuído (ou seja, não existe uma hierarquia centralizada de ordens e regras), distância do equilíbrio e dependência sensível às condições iniciais. Na linguística, linguistas do continente americano, representados por trabalhos como o *position paper* de Larsen-Freeman et al. (2013) que utilizavam a terminologia Sistemas Adaptativos Complexos, e europeus, representados por pesquisas de autores como De Bot et al. (2007), chegaram ao consenso do termo Sistemas Dinâmicos Complexos (SDC), que é adotado na presente pesquisa. As principais características que justificam a adoção dos SDC em pesquisas na linguagem é o reconhecimento de que os processos de aprendizagem e aquisição não são ergódicos, ou seja, não contêm variações uniformes. Segundo Lowie e Verspoor (2019), existe alta variabilidade no desenvolvimento dos aprendizes de uma língua estrangeira. Neste sentido, Verspoor et al. (2008) afirmam que é normal que exista grande variabilidade nos *caminhos* que cada falante toma no aprendizado com o passar do tempo, resultando até mesmo em ordens diferentes de aquisição de certos tópicos de uma língua estrangeira.

Na seção seguinte, o tópico da Comparação de Locutores é apresentado e discutido. A TSDC pode contribuir para a CL por suportar análises não apenas entre falantes, mas também comparações de produções partindo de um mesmo locutor: além de pistas necessárias para discriminar dois falantes, pode ser útil um arsenal de pistas que possam individualizar um determinado locutor.

3. Comparação de Locutores

A fonética é uma área da linguística que tem diversas aplicabilidades

práticas. É possível, por exemplo, treinar atores para representarem sotaques de personagem com mais acurácia e músicos para conseguirem cantar de forma mais fidedigna em línguas estrangeiras. A CL é outra área de importância para a atuação de foneticistas. Para começar a discussão acerca disso, faz-se necessária uma breve revisão de como a Fonética Forense se desenvolveu através dos anos.

Segundo Nolan (2001), um exemplo do quão complexa a área da fonética forense pode ser está na plasticidade e no potencial de mudança que o sistema fonológico de um indivíduo carrega em si. O indivíduo pode adquirir novas línguas, códigos, alterar seu sotaque e expandir seu léxico, além, é claro, do efeito da mudança na própria voz ao longo do tempo. Hollien (2002) adiciona que o estresse, o período do dia e as emoções sentidas podem alterar a voz de uma pessoa. Esse autor expõe que existe um problema de continuidade na fala das pessoas porque os órgãos da fala possuem outras funcionalidades primárias (como respirar e digerir) e pode haver importantes mudanças em diversos componentes do trato vocálico.

Os termos “Verificação de Locutor” e “Identificação e locutor” nem sempre são adotados atualmente, como explicado por Gonçalves (2013), sendo o termo preferível “Comparação de Locutor”. Para compreender como esse consenso foi alcançado, é necessário observar o desenvolvimento das pesquisas ao longo do tempo. Nolan (1999) dividia a área em Verificação de Locutor (VL) e Identificação de Locutor (IL). A VL era aplicada em casos nos quais o locutor desejava ser identificado para reconhecimento de voz remoto, acesso a cofres ou sistemas de segurança por voz, por exemplo. Nesses casos, a voz do locutor poderia ser comparada com as vozes presentes em um banco de dados para tomar uma decisão sobre a validade da voz. No caso da IL, de acordo com Nolan (1999), estava mais relacionada à identificação de indivíduos envolvidos em crimes com base em evidências de arquivos de áudio (atualmente, pode-se exemplificar com ameaças enviadas por meio de serviços de mensagens virtuais, como o WhatsApp ou Telegram). A maior dificuldade nesses casos seria, segundo o autor, a possível existência de disfarces de voz e a implementação de estratégias para enganar possíveis peritos. Hollien (2002) também menciona que, em casos de IL, há a dificuldade de comparar a voz do suspeito com a voz virtualmente de um conjunto de pessoas. Gonçalves (2013) afirma que, após o ano de 2002, ambos os termos, VL e IL, foram englobados pelo termo “Comparação de Locutor”. A autora afirma que, em casos práticos de investigações forenses, ocorre a coleta de dados por meio da leitura de frases de referência pelos suspeitos. Nesse sentido, é válido investigar as características acústicas dos sons consonantais /p/ nas línguas portuguesa e inglesa, uma vez que são consoantes muito comuns em ambas as línguas, e as variáveis do bilinguismo e das diferenças de duração da oclusão e do tempo de vozeamento podem fornecer pistas importantes na individualização dos falantes.

4. A Consoante /p/

Para definir teoricamente a consoante /p/, foco da atual pesquisa, é importante discutir o status de consoantes e das consoantes oclusivas, especificamente. Segundo Ladefoged e Johnson (2011), as consoantes oclusivas

(ou plosivas) são articuladas após o ar expelido dos pulmões ser impedido de passar totalmente por algum articulador. Entende-se por articuladores órgãos como a lâmina da língua ou o lábio inferior, por exemplo. No caso da consoante /p/, ela é classificada como bilabial não vozeada, pois a interrupção no fluxo de ar acontece pelo fechamento e contato entre os articuladores lábio inferior e lábio superior e não há vibração das pregas vocais. De acordo com Kent e Read (2011), é fundamental também notar, em uma análise acústica, a barra de vozeamento, que será mencionada nas análises qualitativas. A barra de vozeamento é uma seção na parte inferior de um espectrograma acústico que, caso esteja preenchida (ou seja, visualmente escura na imagem), aponta que há vozeamento – as pregas vocais estão vibrando no momento da produção do som. Segundo Schwartzhaupt (2015), é possível dividir a consoante oclusiva em três partes: a oclusão, ou obstrução do ar pelos articuladores, a soltura do ar (caracterizada pela explosão de ar) e, por fim, as primeiras ondas do som vocálico seguinte. Uma diferença importante entre as consoantes oclusivas não vozeadas (/p/, /t/, k/) do português em comparação ao inglês está no *Voice Onset Time* (VOT). Segundo Kupske (2016) e Ladefoged e Johnson (2011), na língua inglesa o tempo entre a soltura da explosão de uma plosiva e o início da vogal seguinte é mais longa e essa etapa vem acompanhada de um ruído turbulento e aspirado. Este período é denominado de VOT e é visível como ruído no espectrograma em uma análise acústica.

O primeiro período de uma consoante oclusiva é a oclusão propriamente dita. Cristófar-Silva *et al.* (2019) também chamam esse período de “ausência de energia”, uma vez que pode ser observado como uma área em branco em uma análise acústica. Durante a oclusão, os articuladores se fecham e há uma interrupção total, porém breve, do fluxo de ar. Crystal e House (1988) apresentam, em uma revisão de literatura, as médias de durações de oclusão de oclusivas em língua inglesa encontradas por outros autores. Estes valores variam entre 50ms e 86ms para consoantes vozeadas e 50ms e 140ms para consoantes não-vozeadas.

Com relação ao VOT, diversos estudos vêm sendo desenvolvidos desde o trabalho inaugural de Lisker e Abramson (1964). Lofredo-Bonatto e Silva (2018) afirmam que línguas como o inglês requerem a presença do VOT nas consoantes não vozeadas para que haja a diferenciação entre suas contrapartes vozeadas (como a diferença entre /p/ de /b/, por exemplo). Isso levanta questionamentos de como falantes nativos de línguas românicas (como português, espanhol e italiano), idiomas que não possuem um VOT longo como característica de suas consoantes plosivas no ataque silábico, produziram estes segmentos em línguas como o inglês.

5. Metodologia

Os dados da presente pesquisa são provenientes do *corpus* produzido para a pesquisa de Gomes (2018) para análise de ditongos. Esses dados foram coletados a partir de dois instrumentos: texto e sentenças-veículo. O texto era dividido em seis partes e era lido pelas informantes uma vez em português e uma vez em inglês. As sentenças-veículo totalizavam 40 e eram lidas três vezes por cada informante. Além das palavras-alvo da

pesquisa de Gomes (2018) e das escolhidas palavras *pie* (inglês) e *pai* (português) para a presente pesquisa, havia também distratores. A escolha do par *pie/pai* se deu pela semelhança de contexto fonológico. Em ambas as línguas, produz-se a consoante /p/ com o ditongo /aɪ/ em sequência. Nota-se, ademais, que a ordem em que cada sentença-veículo aparecia para as informantes era alterada a cada repetição. Para a presente pesquisa foram escolhidos do *corpus* 16 informantes, sendo oito mulheres estadunidenses e oito mulheres brasileiras.

As informantes estadunidenses eram imigrantes ou intercambistas provenientes das costas Nordeste e Leste dos Estados Unidos da América e viviam no Brasil há meses ou anos, enquanto as informantes brasileiras eram estudantes de letras ou professoras de inglês, todas nascidas e/ou moradoras de longa data da região metropolitana de Curitiba, Paraná, Brasil. As integrantes de ambos os grupos afirmavam usar ambas as línguas diariamente. Optou-se por selecionar apenas mulheres nesta pesquisa para não lidar com a variável sexo. A média de idade das informantes brasileiras foi de 29 anos, enquanto a média para as estadunidenses foi de 35 anos.

Conforme em Gomes (2018), o microfone utilizado foi o Zoom H1 Handy Recorder, e as gravações se deram em um estúdio de acústica em uma universidade no estado do Paraná. As informantes faziam um treinamento e, então, sozinhas na cabine de gravação, liam folhas impressas contendo os textos e as sentenças-veículo. Os dados, etiquetados para a pesquisa de mestrado, foram então analisados pelo software *Praat* e armazenados em planilhas virtuais para melhor organização e realização de testes estatísticos. O número total de *tokens* de oclusão, VOT, e duração total da consoante foi de 1764 após o descarte de alguns dados em que a informante pronunciou a palavra trocando, por confusão, a consoante inicial ou a vogal subsequente, ou quando houve corrupção de arquivos, em apenas dois casos.

Para a etiquetagem dos espectrogramas na análise acústica, o VOT foi medido desde a primeira forma de onda da soltura da oclusão da plosiva até o primeiro pulso vocálico do ditongo seguinte.

Com relação aos testes estatísticos aplicados nas sessões quantitativas, que serão brevemente mencionados antes da análise qualitativa, optou-se pela realização de testes não paramétricos Mann-Whitney, uma vez que a variabilidade (apontada por testes de normalidade) entre os dados era grande. Os valores de *p* iguais ou inferiores a 0,05 foram considerados como estatisticamente significativos.

6. Resultados e discussão

A seguir, faz-se um breve resumo dos resultados quantitativos obtidos durante a pesquisa de mestrado, para, em seguida, proceder-se às análises qualitativas do conjunto de dados.

Apesar do presente recorte estar focado na análise qualitativa dos espectrogramas que estarão dispostos na seguinte seção, julga-se importante apresentar as durações totais e os principais resultados estatísticos obtidos nas análises do mestrado no sentido de, quando necessário, referir-se às

durações dos segmentos de oclusão, VOT e consoante pelos valores averiguados na dissertação. Os principais resultados dos dados quantitativos podem ser observados na Tabela 1, disposta a seguir.

Tabela 1: Dados quantitativos obtidos durante a pesquisa de mestrado

Part.	Tarefa	VOT		OCLUSÃO		CONSOANTE	
		ING	PORT	ING	PORT	ING	PORT
AM	SENT	82,7 (18,4)	25,9 (5,5)	110,8(24,9)	131(30,2)	193,4(43,2)	156,9(35,7)
	TEXT	60,5 (17,4)	26,1 (7,1)	75,7 (21,9)	78,8(22,6)	136,2(40,1)	104,9(29,7)
BRAS	SENT	49,2 (11,1)	13,9 (3,4)	127,3(28,4)	155,8(37,1)	176,5(39,5)	169,5(40,6)
	TEXT	46,7 (12,0)	15,9 (5,4)	105,8(27,8)	85(28,6)	152,5(39,8)	100,8(34)

Fonte: Produção própria.

Na Tabela 1, pode-se observar os valores encontrados para VOT, oclusão e consoantes produzidos por estadunidenses (referidas com AM na tabela) e brasileiras nas duas línguas. Após a realização de testes estatísticos Mann-Whitney, constatou-se que as principais diferenças significativas foram: duração absoluta da oclusão¹ em inglês entre estadunidenses (92,2ms) e brasileiras (115,2ms), $U=0$; $p<0,001$; duração absoluta do VOT em inglês entre estadunidenses (72,5ms) e brasileiras (47,6ms), $U=8$; $p=0,041$; duração absoluta de VOT para estadunidenses entre inglês (72,6ms) e português (26ms), $U=0$; $p=0,016$; duração absoluta de VOT para brasileiras entre inglês (47,6ms) e português (14,9ms), $U=2$, $p=0,047$; durações absolutas das consoantes das brasileiras entre inglês (162,8ms) e português (134,3ms). É possível concluir, a partir dos dados quantitativos, que estadunidenses produziram oclusões significativamente mais curtas do que brasileiras; estadunidenses produziram VOTs em inglês significativamente mais longos que brasileiras; estadunidenses e brasileiras produziram VOTs diferentes nas duas línguas e significativamente mais longos em inglês do que em português; brasileiras produziram consoantes mais longas em inglês do que em português. A partir destes dados, propõe-se verificar se é possível observar esses resultados quantitativos nos espectrogramas das produções.

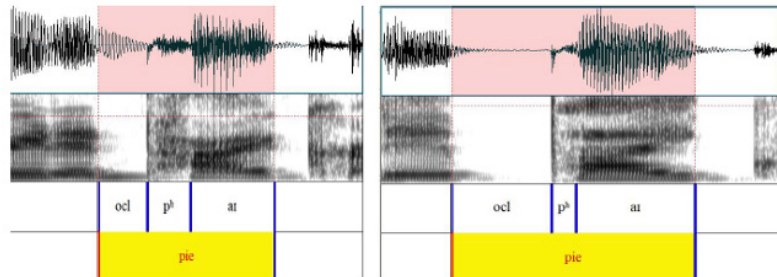
7. Análise qualitativa

A partir do argumento apresentado, na introdução, de fornecer um olhar mais individualizado para o conjunto de dados e para as premissas da análise acústica dos sons da fala, duas informantes foram selecionadas para a presente seção: a estadunidense com dados etiquetados como A3 e a brasileira com dados etiquetados como B15. Entre as características desejáveis deste conjunto de dados estão que os espectrogramas estavam limpos e sem muitos ruídos que pudessem comprometer a visualização das formas de

1 Durante a pesquisa de mestrado, também foram medidas durações relativas e as estatísticas correspondentes para cada medida. Todavia, como os resultados foram similares, optou-se por selecionar somente os valores absolutos de durações. A diferença é que, para o cálculo das durações relativas, divide-se a duração do segmento pela duração da palavra e multiplica-se por 100, para que possíveis interferências de variáveis, como velocidade de fala, por exemplo, sejam atenuadas.

onda; todos os dados foram processados e não foram descartados por eventuais erros na produção ou corrupção de arquivos. Dito isso, apresenta-se, na Figura 1, a comparação da produção da palavra *pie* na tarefa da leitura de texto em língua inglesa das informantes A3 e B15.

Figura 1: Produção de *pie* pelas informantes A3 e B15 na leitura de texto em inglês

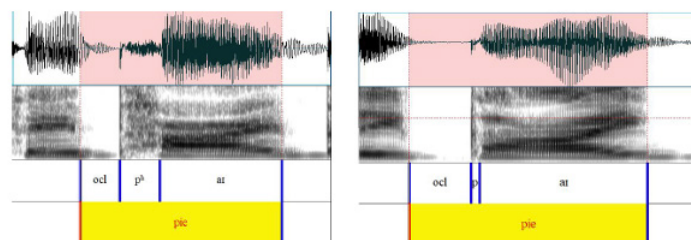


Fonte: Produção própria.

Na produção da palavra em inglês *pie*, as durações dos segmentos da informante A3, à esquerda, são: 75ms de oclusão, 74ms de VOT e 149ms do total da consoante. Para a informante B15, os valores são de 142ms de oclusão, 35ms de VOT e 177ms de total da consoante. É notada na imagem a grande diferença da duração do período de oclusão. A área em branco, que responde ao fechamento dos articuladores (lábios) antes da soltura, é visivelmente maior para a informante brasileira. Por outro lado, o VOT da informante estadunidense, numericamente com quase o dobro de duração em comparação à informante brasileira, também é muito maior no setor hachurado da imagem. A barra de vozeamento das imagens também é um ponto de destaque, uma vez que o período de oclusão da informante estadunidense é menor e com a barra de vozeamento presente em toda a extensão da oclusão, enquanto o resquíio da barra de vozeamento da produção da informante brasileira vai esmaecendo conforme o período de oclusão acontece, e o espaço da metade final do período de oclusão é totalmente limpo, ou seja, sem barra de vozeamento.

A seguir, na Figura 2, encontram-se as produções de *pie* em inglês pelas mesmas informantes, porém na tarefa de sentenças-veículo.

Figura 2: Produção de *pie* pelas informantes A3 e B15 nas sentenças-veículo em inglês

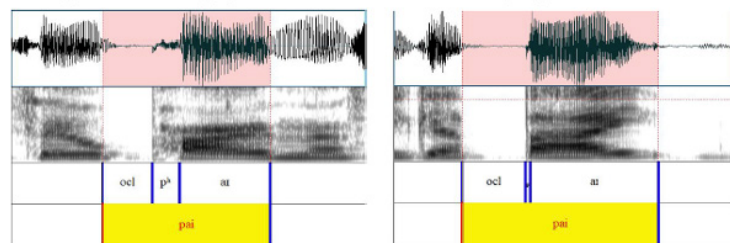


Fonte: Produção própria.

As durações dos segmentos da informante A3, à esquerda, são 77ms de oclusão, 77ms de VOT e 154ms do total da consoante. Já para B15, à direita, os valores são 143ms de oclusão, 12ms de VOT e 155ms de total da consoante. Para a informante brasileira, o VOT dessa repetição é notadamente

curto, tendo apenas 12ms e sendo muito diminuto na imagem, ao contrário do VOT da estadunidense, que se mantém longo e discernível da oclusão e do ditongo. Com relação às barras de vozeamento, elas seguem padrões muito parecidos com os padrões encontrados na Figura 1: o período de oclusão da estadunidense segue quase totalmente ocupado pela barra de vozeamento, enquanto a produção da brasileira demonstra esmaecimento e o período de oclusão termina limpo e sem barra de vozeamento visível. Outro fator interessante nesse *sample* é que a duração total da consoante é praticamente igual para as duas informantes, porém com composições de duração de oclusão e VOT totalmente discrepantes. Por fim, para finalizar as comparações de espectrogramas entre as informantes na língua inglesa, apresenta-se a Figura 3, com dados de A3 e B15 para a tarefa de textos na língua portuguesa.

Figura 3: Produção de *pai* pelas informantes A3 e B15 no texto em português

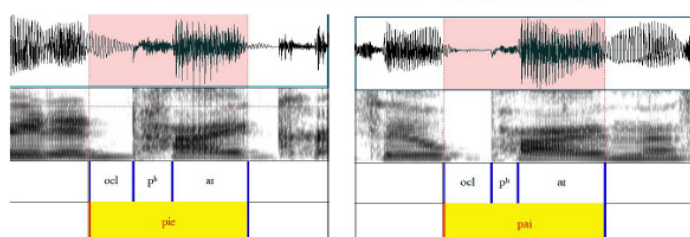


Fonte: Produção própria.

Os valores de duração para a informante A3 no espectrograma da Figura 3 são 82ms de oclusão, 45ms de VOT e 127ms de consoante total. Já para a informante B15, os valores foram 146ms de oclusão, 12ms de VOT e 158ms de consoante total. Analisando o VOT entre as informantes é visível a diferença de praticamente três vezes a duração de VOT de A3 em relação à B15. Assim, como já observado na Figura 2, a oclusão da B15 continua muito maior do que a da A3, porém as barras de vozeamento dessa repetição em específico não possuem a característica do esmaecimento presente nos exemplos anteriores. Talvez por se tratar da língua portuguesa nesse caso. Nos testes estatísticos da pesquisa quantitativa, as diferenças significativas de maior peso haviam sido encontradas na língua inglesa, indicando que a língua portuguesa realmente possuía mais uniformidade entre as nacionalidades.

Com o intuito de se comparar a mesma informante em línguas diferentes, apresenta-se, primeiramente, a Figura 4, contendo a informante A3 produzindo *pie* em inglês à esquerda e produzindo *pai* em português à direita.

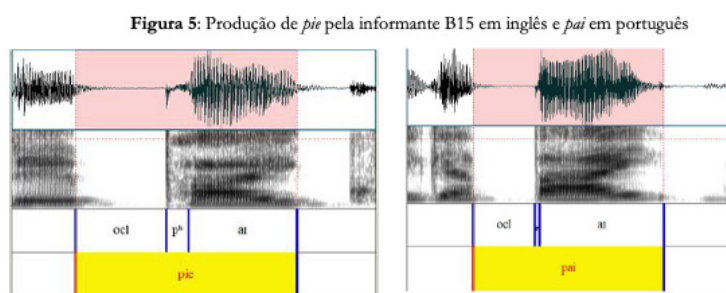
Figura 4: Produção de *pie* pela informante A3 em inglês e *pai* em português



Fonte: Produção própria.

Na Figura 4, os valores de durações para as produções de A3 em inglês foram de 78ms para oclusão, 71ms para VOT e 149ms para a duração total da consoante. Já no português, os valores foram de 82ms para oclusão, 45ms para VOT e 127ms para o total da consoante. Observa-se, pelo espectrograma e pelas médias de duração, durações absolutas de VOT maiores para o inglês do que para o português, conforme é amplamente descrito pela literatura (observa-se, no entanto, que esse VOT da A3 em português é mais longo do que o padrão nativo de falantes brasileiras). O setor específico do VOT parece com mais hachuras e as ondas com maior amplitude em inglês do que no português, talvez indicando pela imagem a aspiração mais forte na articulação do segmento em língua inglesa. Comparando especificamente essas duas imagens também é possível verificar nitidamente a diferença entre um período de oclusão vozeado no inglês (presença da barra de vozeamento na porção inferior da imagem) do que um período de oclusão não vozeado no português.

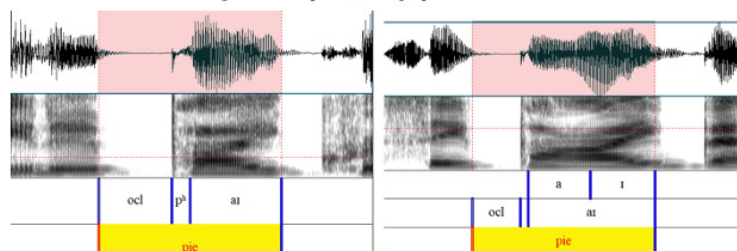
Com o mesmo objetivo da última análise apresentada pela Figura 4, ou seja, observar a mesma informante nas duas línguas, apresenta-se B15 produzindo *pie* em inglês à esquerda e *pai* em português à direita na Figura 5.



Fonte: Produção própria.

Na Figura 5, os valores de durações da informante B15 em inglês foram de 142ms para oclusão, 35ms para VOT e 177ms para a duração total da consoante. Já para o português, os valores foram de 146ms para oclusão, 12ms para VOT e 158ms para o total da consoante. Nota-se a grande diferença de VOT da brasileira entre as línguas. Na Figura 4, apontamos que a produção de VOT da A3 em português teve duração muito acima da média das nativas brasileiras – ora, ela foi quase a mesma duração da própria brasileira B15 na língua inglesa. Já o VOT diminuto em português corresponde ao padrão nativo encontrado na literatura (mas não exatamente com a média encontrada na presente pesquisa). As oclusões da B15 em ambas as línguas são quase iguais e a barra de vozeamento começa ativada, porém esmaece em ambos os casos. Na sequência, nas Figuras 6 e 7, analisam-se dados da mesma informante na mesma língua. Na figura 6, é possível observar duas produções em inglês pela informante B15.

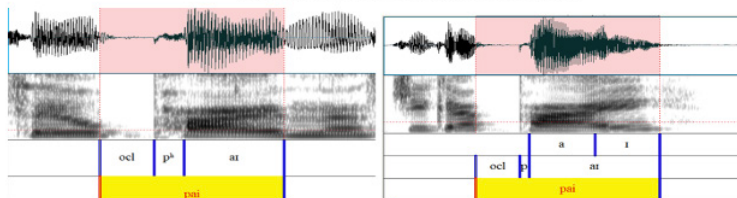
Figura 6: Duas produções de *pie* pela informante B15



Fonte: Produção própria

Na Figura 6, os valores de duração da oclusão da brasileira em inglês, na segunda produção da tarefa do texto, localizada à esquerda, foi de 142ms de oclusão, 35ms de VOT e 177ms de duração total da consoante. Já na terceira repetição das sentenças-veículo pela mesma informante em inglês, à direita, indicou valores de duração de oclusão de 134ms, VOT 21ms e duração total da consoante de 155ms. É notável a diferença de um VOT de 35ms ao lado de um VOT de 21ms, uma vez que a duração maior fica explícita. No oscilograma da soltura da oclusão, pode-se observar que as formas de onda são parecidas entre as duas produções. As trajetórias de F1 e F2 das produções também possuem uma abertura um tanto acentuada, mas isso foge do escopo da pesquisa atual. Nota-se também que ambas as barras de vozeamento se encontram esmaecidas ao início da articulação de /p/. A seguir, na Figura 7, duas produções em português pela informante A3.

Figura 7: Duas produções de *pai* pela informante A3



Fonte: Produção própria.

Na Figura 7, a produção à esquerda representa a segunda repetição de *pai* em português pela informante A3 na tarefa de texto. Os valores encontrados foram de 82ms para oclusão, 45ms para VOT e 127ms para total da consoante. Já a produção à direita corresponde à primeira repetição de *pai* em português na tarefa de sentenças-veículo. Os valores encontrados foram de 108ms para oclusão, 23ms para VOT e 131ms para total da consoante. Destacam-se as duas barras de vozeamento esmaecidas durante os períodos de oclusão que, exceto pela diferença de 26ms de duração, são bem semelhantes. O VOT da figura da esquerda, assim como o observado da Figura 6, é bem discernível pela maior duração, ao se comparar com o VOT da figura à direita. Inclusive, esse VOT de 45ms fica bem mais próximo às produções de VOT da língua inglesa, mas não portuguesa.

Por meio dos espectrogramas apresentados, é possível afirmar que os dados quantitativos são observáveis mediante os espectrogramas. Os períodos de oclusões, geralmente mais longos nas informantes brasileiras, podem ser visualizados, bem como as diferenças de VOT (geralmente mais longos nas informantes estadunidenses). O F2 de transição, porém, é bem

mais difícil de ser observado visualmente. Além disso, notou-se que outras qualidades dos sons, como vogais e barras de vozeamento, podem ser observadas qualitativamente – parâmetros que podem ser analisados em futuros trabalhos.

8. Considerações finais

Com relação às aplicabilidades dos SDCs no contexto da presente pesquisa, percebeu-se a falta de um controle central (CHAN, 2001), uma vez que as produções analisadas simplesmente não seguiam um *manual* de como deveriam ser as durações de oclusão, VOT, consoante, e a variabilidade encontrada foi alta. Na porção quantitativa da pesquisa, foi possível observar que as produções de cada informante variaram muito no *range* de durações de VOT, entre 10ms e 150ms, então a adoção de um modelo de língua que adota como *esperado* um número grande de variações entre nacionalidades e intranacionalidades está de acordo com o que se espera de uma observação da língua em sua efetuação real, e não teórica.

Baranger (2000) afirma a distinta posição dos Sistemas Complexos entre o caos e a previsibilidade. Assim como um simples estado caótico não caracteriza um SDC, os dados da pesquisa, apesar de demonstrarem variabilidade, também indicam padrões como as durações mais longas de VOT por parte das estadunidenses em língua inglesa. Seria interessante incorporar análises de mais pares de informantes de ambas as nacionalidades para averiguar os espectrogramas qualitativamente, indo além das simples medidas médias de grupos.

A partir da comparação dos espectrogramas apresentados, conclui-se válida a premissa de aliar a análise estatística com a análise qualitativa dos espectrogramas, uma vez que pode ser *apontado* visualmente nos espectrogramas o que exatamente as estatísticas estão dizendo com relação às durações dos segmentos. Um VOT de 12ms, por exemplo, fica quase inexistente ao lado de um VOT de 74ms, além da amplitude do ruído da aspiração nas consoantes com maiores VOTs que também pode ser visualizada. Além disso, outros achados da pesquisa residem em fatores como as diferenças entre barras de vozeamento e a presença de vogais com características de voz como a *creaky voice*. A análise desses aspectos pode ser útil para peritos em contextos de CL. Na Figura 1, por exemplo, a informante A3 apresenta a característica de voz crepitante, ou *creaky voice*, o que pode ser percebido no ditongo posterior à consoante de ataque na imagem. Contando com um *n* amostral maior, seria possível até mesmo criar uma hipótese referente a se a questão do esmaecimento da barra de vozeamento em alguns casos da informante A3 realmente acontece em diversos casos, como uma característica individual, ou se outras informantes também apresentam essa característica, indicando uma característica da língua. Em um contexto de comparação de suspeitos em uma análise forense, também é possível prestar atenção em fatores como a diferença de duração de VOT, uma vez que isso pode indicar bilinguismo, isto é, ser uma característica da fala da pessoa ou até mesmo que o locutor em questão possui outra língua materna.

Referências

BARANGER, M. **Chaos, Complexity and Entropy**: A physics talk for non-physicists. New England Complex Systems Institute. Abril 2000. Disponível em: <http://www.necsi.edu/projects/baranger/cce.html>. Acesso em: 27 jul. 2020.

CHAN, S. Complex Adaptive Systems. **Research Seminar in Engineering Systems**. 2001.

CRISTÓFARO-SILVA, T.; SEARA, I.; SILVA, A. H. P.; RAUBER, A. S.; CANTONI, M. **Fonética Acústica**: os sons do português brasileiro. São Paulo: Editora Contexto, 2019.

CRYSTAL, T.; HOUSE, A. **The duration of American-English stop consonants**: an overview. Communication Research Division, Institute for Defense Analyses, Princeton, Estados Unidos da América. 1988.

DE BOT, K.; LOWIE, W.; VERSPOOR, M. A Dynamic Systems Theory approach to second language acquisition. **Bilingualism: Language & Cognition**, p. 7-21, 2007.

GOMES, M. Normalização de frequência de formantes e análise com dados brutos na visão de língua como um Sistema Adaptativo Complexo. **Gradus-Revista Brasileira de Fonologia de Laboratório**, v. 3, n. 1, p. 81-102, 2018.

GONÇALVES, C. **Taxa de elocução e de articulação em corpus forense do português brasileiro**. Tese (Doutorado em Linguística) – Pontifícia Universidade Católica do Rio Grande do Sul. Porto Alegre, 2013.

HOLLAND, J. Studying Complex Adaptive Systems. **Journal of Systems Science and Complexity**. 2006.

HOLLIEN, H. Forensic Voice Identification. **Academic Press**. Estados Unidos, 2002.

KENT, R.D.; READ, C. **The Acoustic Analysis of Speech**. Connecticut, USA: Thomson, 2011.

KUPSKE, F.F. **Imigração, atrito e complexidade: a produção das oclusivas surdas iniciais do inglês e do português por sul-brasileiros residentes em Londres**. Tese (Doutorado em Linguística) – Universidade Federal do Rio Grande do Sul. Porto Alegre, 2016.

LADEFOGED, P.; JOHNSON, K. **A Course in Phonetics**. Cengage Learning, 2011.

LARSEN-FREEMAN, D. Complexity Theory: A New Way to Think. **RBLA**,

Belo Horizonte, v.13, n. 2, p. 269-373, 2013

LISKER, L.; ABRAMSON A.S. A Cross-Language Study of voicing in initial stops: Acoustical measurements. **Word**, 2003, p.384-422, 1964.

LOFREDO-BONATTO, M.; SILVA, M. Comparação de sons plosivos em crianças monolíngues e bilíngues por meio do parâmetro acústico voice onset time: relato de casos. **Revista CEFAC**, v. 20, p. 680-687, 2018.

LOWIE, W.; VERSPOOR, M. H. Individual differences and the ergodicity problems. **Language Learning**, v. 69, n. S1, p.184-206, 2019.

NOLAN, F. Speaker Recognition and forensic phonetics. *In: **The Handbook of Phonetics Science***. Blackwell Publishers, MA, Estados Unidos da América. 1999.

NOLAN, F. **Speaker identification evidence**: its forms, limitations, and roles. Reino Unido: University of Cambridge, 2001.

SCHWARTZHAUPT, B. M. **Testing intelligibility in English**: The effects of positive VOT and contextual information in a sentence transcription task. Dissertação (Mestrado em Linguística). Universidade Federal do Rio Grande do Sul. Porto Alegre, 2015.

VERSPOOR, M., LOWIE, W.; Van DIJK, M. Variability in second language development from a dynamic systems perspective. **Modern Language Journal**, 92, p. 214-231, 2008.