

PERCEPÇÃO E PRODUÇÃO DOS PADRÕES DE VOT DO INGLÊS POR APRENDIZES BRASILEIROS: O PAPEL DE MÚLTIPLAS PISTAS ACÚSTICAS SOB UMA PERSPECTIVA DINÂMICA

Ubiratã Kickhöfel ALVES*
Márcia Cristina ZIMMER**

- RESUMO: Neste trabalho, a partir de uma concepção dinâmica de aquisição fonético-fonológica de L2, investigamos a percepção (identificação e discriminação) e a produção dos padrões de *Voice Onset Time* (VOT) das plosivas iniciais do inglês por 32 aprendizes do Sul do Brasil. Partimos da premissa de que, sobretudo entre aprendizes com nível básico de proficiência, o VOT não se mostra como pista acústica prioritária para as distinções funcionais de sonoridade. Os resultados dos testes de percepção mostram que, independentemente do nível de proficiência dos aprendizes (básico ou avançado), o VOT tomado unicamente não se faz suficiente para a distinção entre /p/, /t/, /k/ e /b/, /d/, /g/. Tais resultados, que exercem influência sobre os dados de produção, corroboram uma visão dinâmica de aquisição de L2, a partir da qual múltiplas pistas acústicas agem em conjunto nas distinções entre sons, cabendo ao aprendiz saber selecionar aquelas pistas com caráter mais primordial no sistema a ser adquirido.
- PALAVRAS-CHAVE: Percepção de L2. Produção de L2. Língua inglesa. VOT. Pistas Acústicas.

Introdução

O processo de aquisição fonético-fonológica de uma Segunda Língua (L2)¹ caracteriza-se como complexo e dinâmico por natureza. Muitas variáveis, agindo em conjunto, mostram-se fundamentais para o entendimento desse complexo quadro. No que diz respeito à percepção e à conseqüente produção dos sons da língua-alvo, múltiplas pistas acústicas podem agir em conjunto no estabelecimento das diferenças funcionais entre os sons a serem adquiridos. Nesse sentido, adquirir uma LE implica a capacidade do aprendiz de detectar, em termos de percepção, aquelas pistas acústicas que se mostram produtivas no

* UFRGS – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Letras. Porto Alegre – RS – Brasil. 91501-970 – ukalves@gmail.com.

** UNIRITTER – Centro Universitário Ritter dos Reis. Programa de Pós-Graduação em Letras. Porto Alegre – RS – Brasil. 90840-440 – marcia.zimmer@gmail.com.

¹ No presente trabalho, não se faz distinção entre os termos Segunda Língua e Língua Estrangeira, ou entre os termos 'Aquisição' e 'Aprendizagem'.

sistema-alvo, e, posteriormente, no que concerne à produção, fazer uso majoritário dessas pistas acústicas, de caráter não redundante, para o estabelecimento das diferenças fonológicas do sistema de LE.

Um exemplo bem claro, a ser discutido no presente trabalho, diz respeito à aquisição dos padrões de *Voice Onset Time* (VOT) do inglês por aprendizes brasileiros. No inglês, as plosivas surdas /p/, /t/, /k/ são produzidas com um intervalo de VOT longo, também chamado de VOT Positivo (aspiração). Esse aspecto fonético constitui a pista principal para a distinção entre segmentos surdos e sonoros, na língua alvo (SCHWARTZHAUPT; ALVES; FONTES, 2013). Entretanto, em um estudo piloto prévio, Alves e Zimmer (2012) sugeriram que a duração do VOT, entre aprendizes brasileiros, não parecia ser a pista fundamental para a distinção entre segmentos surdos e sonoros do inglês, ao contrário do que ocorre com falantes nativos dessa língua, em que a presença/ausência de aspiração se revela como a pista majoritária para que os segmentos sejam identificados como surdos ou sonoros. No caso do falante de Português Brasileiro, outras pistas acústicas, tais como a intensidade da explosão do segmento plosivo e o valor de F0 da vogal seguinte, podem estar exercendo um papel majoritário nessas diferenças funcionais. Com base em tal possibilidade, seria possível explicar, também, o fato de que aprendizes brasileiros, mesmo em nível avançado de proficiência, acabam por não produzir valores de VOT semelhantes àqueles encontrados no padrão nativo (ALVES; SCHWARTZHAUPT; BARATZ, 2011).

Em outras palavras, seguindo-se premissa estabelecida no estudo piloto de Alves e Zimmer (2012), é possível que a pista acústica VOT, ao ser tomada isoladamente, não seja suficiente para a distinção, em termos perceptuais, entre /p, t, k/ e /b, d, g/ do inglês. Casos semelhantes têm sido discutido em estudos recentes, como os de Sundara (2005), Oh (2011) e Kong, Beckman e Edwards (2012). Nos trabalhos em questão – que se voltam à investigação do francês canadense, do coreano e do japonês, respectivamente – pistas acústicas adicionais, tais como a intensidade de explosão do segmento plosivo e o valor de F0 da vogal seguinte, mostram-se determinantes, como os detalhes acústicos principais, na distinção entre os segmentos surdos e sonoros.

Tal constatação tem implicações diretas no que diz respeito ao entendimento do processo de aquisição de uma língua estrangeira (LE). Ao considerarmos que os segmentos constituem conjuntos de múltiplas pistas acústicas, falantes das diferentes línguas do mundo (ou ainda, em escala menor, dos seus dialetos) tendem a selecionar algumas pistas acústicas prioritárias para caracterizar as distinções funcionais entre os segmentos naquele dado sistema. Em termos de modelos perceptuais de aquisição de L2, tal constatação pode ser explicada satisfatoriamente pelo *Perceptual Assimilation Model-L2* (PAM-L2 – BEST;

TYLER, 2007). Conforme explicam Antoniou et al. (2011), tal proposta tem por base o modelo da Fonologia Gestual de Browman e Goldstein (1992, 1993, 2000). Considerando-se o arcabouço da Fonologia Gestual, seguindo-se Goldstein e Fowler (2003), pode-se pensar no gesto como “moeda comum” de análise entre conhecimento fonológico, percepção e produção. Nesse sentido, “[...] ao adquirirem uma L2, os aprendizes estão sendo expostos a um novo conjunto de gestos articulatórios, incluindo novas relações de faseamento e padrões de coordenação entre estes gestos.” (ANTONIOU et al., 2011, p.560).

A partir da premissa de que “[...] os átomos fonológicos são ações públicas, então eles são diretamente responsáveis pela estrutura nos sinais acústicos de fala, que, por sua vez, provêm informação diretamente a respeito dos átomos fonológicos.”, (GOLDSTEIN; FOWLER, 2003, p.179), verificamos, neste trabalho, de que modo a exposição a uma pista acústica, que constitui evidência de um gesto de status distintivo na língua-alvo, pode exercer modificações na percepção e na produção do sistema de interlíngua do aprendiz.

Dessa forma, no presente trabalho, com base em testes de percepção e produção, discutimos o possível caráter redundante do VOT para a distinção entre plosivas surdas e sonoras do inglês por parte de aprendizes brasileiros. O trabalho apresenta, portanto, os seguintes objetivos principais: (i) Avaliar se os aprendizes, em dois diferentes níveis de proficiência, se mostram capazes de distinguir diferentes padrões de VOT, produzidos por falantes nativos do inglês; (ii) Investigar se os aprendizes dos dois grupos de proficiência produzem padrões de VOT que se mostrem semelhantes àqueles encontrados na língua-alvo; (iii) Com base nos resultados dos dois objetivos anteriores, discutir o papel do VOT como uma pista acústica suficiente, entre aprendizes brasileiros de inglês, para determinar distinções funcionais entre plosivas surdas e sonoras.

Metodologia

Participantes

O estudo contou com 34 participantes, todos naturais da cidade de Porto Alegre – RS. Os aprendizes foram organizados, de acordo com o *Oxford Placement Test Online*², em dois níveis de proficiência: 24 aprendizes de nível básico (níveis A1 e A2, de acordo com o Marco Comum Europeu) e 10 avançados (níveis C1 e C2). Nenhum participante havia recebido instrução formal sobre pronúncia até a

² O *Oxford Placement Test Online* é um teste de nivelamento validado, realizado de forma online a partir do *site* <www.oxfordenglishtesting.com>. Para maiores informações a respeito do teste, vejam-se os artigos de Pollitt (2007) e Purpura (2007).

data de realização dos testes. Todos os informantes realizaram testes de Percepção (Identificação e Discriminação) e Produção.³

Testes de Percepção

Os estímulos foram gravados, em um estúdio profissional, por seis falantes nativos de inglês norte americano (3 homens e 3 mulheres), que se encontravam vivendo no Sul do Brasil há menos de 6 meses. Esses seis falantes leram uma lista contendo três pares mínimos do inglês (*bit – pit, dick – tick, gill – kill*), sendo cada par iniciado por um diferente ponto de articulação, seguido por uma vogal alta, contexto que facilita a produção de intervalos de VOT mais longos (YAVAS; WILDERMUTH, 2006). Para que fosse garantida a qualidade dos estímulos de áudio, solicitamos, a cada um dos falantes, que a lista de palavras-alvo fosse lida três vezes, para que os melhores *tokens* pudessem ser escolhidos para a construção dos testes de percepção.

As plosivas produzidas pelos seis falantes nativos de inglês apresentaram três diferentes padrões de VOT. As plosivas surdas (*pit, tick, kill*) foram sempre realizadas com VOT Positivo (aspiração), enquanto que /b, d, g/ (*bit, dick, gill*) foram produzidos com pré-vozeamento, ou seja, vibração de pregas vocais durante a etapa da closura (VOT Negativo) e com VOT Zero (sem pré-vozeamento ou aspiração, de modo que o vozeamento do segmento vocálico seguinte iniciasse logo após a explosão), uma vez que esses dois padrões ocorrem, variavelmente, na produção de tais plosivas do inglês (LISKER; ABRAMSON, 1964; ABRAMSON; LISKER, 1973; DOCHERTY, 1992; SIMON, 2010).

Além de termos incluído esses três padrões de VOT nas tarefas perceptuais, contamos, também, com estímulos manipulados no software *Praat – Version 5.3.48* (BOERSMA; WEENINK, 2013), para que pudéssemos obter o padrão que denominamos de VOT Zero Manipulado. Para a construção desse padrão, foram reduzidos os intervalos de VOT dos segmentos aspirados (VOT Positivo), de modo que obtivéssemos uma plosiva que apresentasse a mesma duração de VOT de um segmento sonoro do inglês (padrão VOT Zero), mas que, ao mesmo tempo, mantivesse as outras pistas acústicas que caracterizam uma plosiva surda da língua inglesa. Esse padrão se mostra de grande importância, pois, do contraste de tal padrão manipulado com o Zero Natural, poderemos obter insumos que nos permitam avaliar se o VOT constitui a pista acústica majoritária, empregada por aprendizes brasileiros, na distinção entre plosivas surdas e sonoras.

³ Com relação aos aspectos éticos da pesquisa, cabe informar que todos os participantes preencheram um Termo de Consentimento Livre e Esclarecido, a partir do qual os aprendizes foram informados dos procedimentos de coleta de dados, bem como dos riscos e benefícios da pesquisa, além de terem sido informados acerca do fato de que poderiam retirar sua participação em qualquer uma das etapas do desenvolvimento do estudo.

Tanto o teste de Identificação quanto o de Discriminação contaram, portanto, com os quatro padrões de VOT supracitados: VOT Negativo, VOT Positivo, VOT Zero Natural e VOT Zero Manipulado. Ambos os testes de percepção foram elaborados no *software Praat* versão 5.3.48 (BOERSMA; WEENINK, 2013). Nas seções que seguem, descreveremos cada um dos testes.

Teste de Identificação

No Teste de Identificação, os aprendizes foram expostos a estímulos que consistiam em palavras individuais (um dos membros de um dos três pares mínimos descritos acima) e eram convidados a clicar em um botão que indicasse a consoante inicial da palavra que haviam escutado (/p/, /b/, /t/, /d/, /k/ or /g/). Os estímulos foram apresentados em ordem aleatória. A tarefa contava com 48 questões de identificação, sendo que cada um dos padrões de VOT testados (Negativo, Positivo, Zero Natural e Zero Artificial) era apresentado em 12 questões cada (com 4 questões para cada um dos pontos de articulação).

Teste de Discriminação

O Teste de Discriminação apresentava o formato AxB. Nessa tarefa, os estímulos apresentados aos aprendizes eram caracterizados por tríades de palavras. Os participantes deviam indicar se a consoante inicial da segunda palavra é a mesma daquela apresentada na primeira (ex. *bit – bit – pit*) ou na terceira palavra (*bit – pit – pit*), ou se as três palavras eram iniciadas com a mesma consoante (ex. *pit – pit – pit*).

Três tipos de contrastes entre os padrões foram testados na tarefa AxB: (i) VOT Negativo vs. Zero VOT Manipulado (12 questões – 4 para cada ponto de articulação); (ii) VOT Negativo vs. VOT Positivo (12 questões); e (iii) VOT Zero Manipulado vs. VOT Positivo (12 questões). Outros possíveis contrastes, tais como VOT Zero Não Manipulado vs. VOT Negativo, bem como VOT Zero Não-Manipulado vs. VOT Zero Artificial, não foram incluídos no experimento por fins de delimitação, visto que estudos anteriores (ALVES; SCHWARTZHAUPT; BARATZ, 2011) já haviam evidenciado que os aprendizes tendem a discriminar o último desses dois contrastes. Além dos três tipos de contrastes empregados, o teste também contou com nove questões do tipo *catch trial* (três para cada tipo de articulação).⁴

⁴ Os *catch trials* são questões que apresentam a mesma consoante inicial na tríade (ex. *pit – pit – pit*) e têm o objetivo de testar a atenção dos aprendizes à tarefa. Como as respostas obtidas para os *catch trials* apresentaram altos índices de acurácia, indicando, dessa forma, o envolvimento dos participantes com a tarefa, os resultados dessas questões não serão discutidos neste artigo, por fins de delimitação.

Testes de Produção

Os mesmos aprendizes que participaram dos testes de percepção realizaram dois testes de produção, em língua portuguesa e em língua inglesa, cujas descrições serão apresentadas no que segue.

Produção de palavras em português brasileiro

Os aprendizes foram solicitados a ler palavras dissilábicas, apresentadas isoladamente, iniciadas pelos segmentos /p/, /k/, /b/, /g/ e seguidas por uma vogal anterior alta, correspondente ao mesmo contexto fonético-fonológico utilizado nos testes de percepção. Palavras iniciadas por /t/ e /d/ não foram incluídas no instrumento pelo fato de o dialeto de L1 dos aprendizes apresentar o fenômeno de palatalização da alveolar frente a [i] como praticamente categórico (KAMIANECKY, 2002). Para fins de delimitação, neste trabalho, apresentaremos os resultados referentes às plosivas surdas /p/ e /k/.

O teste contou com dois *types* para cada uma das consoantes-alvo (além de 8 *types* de palavras distratoras, iniciadas por segmentos que não os plosivos). Cada um dos *types* era produzido duas vezes, totalizando 4 *tokens*, produzidos por cada participante, para cada uma das consoantes. As palavras eram apresentadas em um arquivo do Microsoft PowerPoint (.ppt), cada uma em um slide diferente. A gravação em áudio foi realizada com um microfone *headset* Philips modelo SHM 3550, em um laptop da marca Sony Vaio, modelo PCG-31311X, com o uso do *Software Audacity* (2015) – Versão 2.0.5.

Produção de palavras em inglês

Neste instrumento, também caracterizado pela leitura de palavras-alvo apresentadas individualmente em slides de um arquivo .ppt, as palavras-alvo eram iniciadas pelos segmentos /p/, /t/, /k/, /b/, /d/, /g/ e seguidas por uma vogal frontal alta (ex. *pit*, *tip*, *kit*). Tendo em vista os objetivos do presente estudo, neste artigo, reportaremos apenas os valores de VOT das palavras iniciadas pelas plosivas surdas /p/, /t/, /k/.

Além das palavras distratoras, o teste contava com 3 *types* para cada consoante. Cada *type* era produzido duas vezes, o que totalizava 6 *tokens* para cada consoante por participante. Assim como no teste em português, as gravações foram feitas num computador Sony Vaio PCG-31311X, com um microfone *headset* Philips, modelo SHM 3550.

Hipóteses

Apresentamos, no que segue, as hipóteses que serviram de base para o estudo. Todas as hipóteses seguem a premissa de que, em um nível mais básico de proficiência, os aprendizes não apresentam o VOT como a principal pista acústica para a distinção entre plosivas surdas e sonoras, ao passo que, em um nível avançado de adiantamento de LE, tal pista já tenha assumido um status prioritário.

As hipóteses serão apresentadas de acordo com o teste a que se referem (Identificação, Discriminação ou Produção).

Hipóteses do Teste de Identificação

H1: No que diz respeito à tarefa de identificação dos padrões VOT Negativo (/b/, /d/ /g/ do inglês) e VOT Positivo (/p/, /t/, /k/ do inglês, com aspiração), não haverá diferenças significativas entre os resultados obtidos dos dois grupos de proficiência.

Motivação para a hipótese: mesmo que não guiados pelo VOT, os aprendizes de nível básico identificarão o padrão de VOT Negativo como vozeado e o VOT Positivo como surdo, uma vez que tais aprendizes baseiam suas respostas em outras pistas acústicas, tais como a intensidade da explosão (plosivas surdas apresentam intensidade mais forte), que os leva a uma correta caracterização do vozeamento dessas consoantes. Os aprendizes mais avançados, por provavelmente seguirem o VOT como pista majoritária, também identificarão os dois padrões como caracterizadores de segmentos surdos e sonoros, respectivamente.

H2: Com relação à identificação do padrão de Zero VOT Não Manipulado, não haverá uma diferença significativa entre os grupos básico e avançado.

Motivação: alunos de nível mais básico de proficiência não tenderão a identificar tais consoantes como surdas, uma vez que eles podem estar sendo guiados, na identificação, pela pista acústica referente à intensidade da explosão (*burst*). Aprendizes de nível mais avançado também estarão sendo guiados pelo padrão de VOT da L2, de modo que a pequena duração do intervalo de VOT os leve a identificar essas consoantes como vozeadas.

H3: No que concerne à identificação do padrão de VOT Zero Manipulado, haverá uma diferença significativa entre os estudantes nos dois níveis de proficiência.

Os aprendizes de nível elementar basearão a identificação nas pistas que se mostram prioritárias na L1 (pistas outras que não o VOT) e, portanto, não identificarão tal padrão como surdo, uma vez que, com exceção da redução do intervalo de aspiração, todas as outras propriedades acústicas desse som manipulado levam à identificação desse padrão como /p/, /t/, /k/; por sua vez, os participantes com um grau mais avançado de proficiência irão identificar tais segmentos como vozeados, uma vez que tomarão o VOT como pista acústica prioritária na identificação, e seguirão o padrão de VOT do inglês, de acordo com o qual plosivas com VOT Zero caracterizam segmentos sonoros.

Hipóteses do Teste de Discriminação

H4: No que diz respeito à discriminação entre os padrões VOT Negativo vs. VOT Positivo, hipotetizamos que não haverá uma diferença estatisticamente significativa entre os resultados obtidos dos participantes de nível básico e avançado de proficiência.

Motivação: os aprendizes dos dois grupos de proficiência serão capazes de discriminar entre *tokens* de /b/, /d/, /g/ e /p/, /t/, /k/, mesmo que por meio de diferentes pistas acústicas, uma vez que hipotetizamos que apenas os aprendizes avançados fazem uso dos padrões de VOT.

H5: Quanto à discriminação entre VOT Negativo e VOT Zero Manipulado, hipotetizamos que haverá uma diferença estatisticamente significativa entre os resultados obtidos dos dois grupos de proficiência.

Motivação: esperamos que os aprendizes de nível básico apresentem altos índices de discriminação, uma vez que eles estarão sendo guiados por pistas outras que o VOT, o que os levará a considerar o padrão de VOT Negativo (com intensidade de explosão fraca) como vozeado, e o VOT Zero Manipulado (com intensidade de explosão forte) como surdo. Os aprendizes avançados, por outro lado, serão guiados pela pista de VOT: uma vez que ambos os padrões são caracterizados por uma duração de VOT bastante curta, esperamos que os dois padrões não sejam discriminados por esses aprendizes.

H6: Com relação à discriminação entre VOT Zero Manipulado e VOT Positivo, também hipotetizamos que haverá uma diferença estatisticamente significativa entre os resultados obtidos dos dois grupos de proficiência.

Motivação: os aprendizes com proficiência básica apresentarão baixos índices de discriminação, uma vez que eles não tomam o VOT como o principal fator em suas respostas. Já os aprendizes de nível de proficiência alta, guiados pelo VOT como pista prioritária, discriminarão esses dois padrões, uma vez que

o primeiro apresenta um VOT curto, enquanto que o segundo é caracterizado pela aspiração.

Hipóteses referentes ao teste de produção

H7: Em cada um dos grupos (considerados separadamente), haverá uma diferença significativa entre as durações de VOT de /p/ e /k/ encontradas em português e em inglês.

Motivação: ainda que os aprendizes possivelmente não tenham atingido valores de VOT semelhantes àqueles encontrados na língua inglesa, em função, inclusive, de o VOT não ser tomado como pista majoritária em termos perceptuais, os participantes já fazem uso parcial da pista de VOT para sinalizar, com uma duração maior, as plosivas surdas do inglês.

H8: Não haverá diferenças significativas entre as durações de VOT produzidas pelos dois grupos no que diz respeito à produção do padrão de VOT Positivo.

Motivação: a hipótese em questão segue a literatura prévia (ALVES; SCHWARTZHAUPT; BARATZ, 2011), que tende a mostrar que, independentemente do nível de proficiência dos aprendizes, os valores de VOT não chegam a ser produzidos de acordo com o padrão nativo. Embora tanto aprendizes de nível básico quanto de nível avançado já consigam identificar plosivas aspiradas como surdas (conforme estabelecido em nossa primeira hipótese), e mesmo que o VOT já seja a pista prioritária apresentada (pelo menos entre os aprendizes de nível avançado), partimos da hipótese de que tais fatos não necessariamente se mostram como condições que garantam valores de VOT próximos do padrão nativo, de modo que não haja diferenças significativas nos valores produzidos por ambos os grupos.

Descrição e discussão dos dados

Esta seção está subdividida em três partes que tratam da descrição e discussão dos dados de identificação, discriminação e produção, respectivamente.

Identificação

Apresentamos, nesta seção, os dados referentes aos testes de Identificação. Como se observa na Tabela 01, independentemente do nível de proficiência dos aprendizes o padrão VOT Negativo (pré-vozeamento) é identificado, quase que categoricamente, como vozeado (99,31%, $m=12^5$ – nível básico e 96,67%,

⁵ m =mediana;

m=12 – nível avançado). Testes de Mann-Whitney não apontaram diferença significativa entre os dois grupos nas identificações dos segmentos como surdos (U=108,00, p= ,121) ou sonoros (U=112,5; p=,487). Tal fato não surpreende, uma vez que na própria língua materna dos aprendizes, o pré-vozeamento já corresponde a uma pista que indica a presença de um segmento vozeado. Dados recentemente coletados por nosso grupo de pesquisa evidenciam que, pelo menos no dialeto gaúcho do Português Brasileiro, o pré-vozeamento dos segmentos /b/, /d/, /g/ não se mostra como categórico, havendo, portanto, instâncias de produção desses segmentos exibindo o padrão VOT Zero. Isso fortalece o argumento de que, ainda que importante, o pré-vozeamento talvez não seja uma pista autossuficiente para a distinção entre segmentos surdos e sonoros nesse dialeto. Estudos sobre a identificação de plosivas /b/, /d/, /g/ sem vozeamento de PB (ALVES; ZIMMER, 2012) são fundamentais para o aprofundamento das discussões a respeito do efetivo papel exercido por essa pista acústica nas distinções funcionais de sonoridade em nossa língua. Assim, mesmo que o VOT não venha a ser a única pista para a identificação de tais segmentos como vozeados, pistas adicionais, tais como a força de explosão do segmento plosivo, explicariam os resultados, pois segmentos vozeados, tanto em português quanto em inglês, são produzidos com uma força de explosão mais fraca (LISKER; ABRAMSON, 1964).

Tabela 1 – Resultados do Teste de Identificação⁶
(12 questões por padrão de VOT).

Padrões de VOT	Grupo Básico		Grupo Avançado	
	Surdo	Sonoro	Surdo	Sonoro
Negativo	0% 0/288	99,31% 286/288	3,33% 4/120	96,67% 116/120
Zero Natural	27,43% 79/288	69,1% 199/288	25% 30/120	71,67% 86/120
Zero Artificial	57,29% 165/288	39,93% 115/288	76,67% 92/120	16,67% 20/120
Positivo	91,67% 264/288	2,78% 8/288	98,33% 118/120	0% 0/120

Fonte: Elaboração própria a partir do banco de dados dos autores.

⁶ São apresentados aqui os índices referentes às respostas corretas com relação à escolha do ponto de articulação das consoantes. Não são apresentados os percentuais de escolha das alternativas [t], [d], [k], [g] em casos em que o estímulo consistia em uma bilabial surda ou sonora, por exemplo) Tal fato explica o porquê de a soma dos índices percentuais não corresponder a 100% das respostas fornecidas.

No que diz respeito à identificação do padrão VOT Positivo (aspiração), os resultados também confirmam o esperado: em ambos os níveis de proficiência, a identificação de tais segmentos como surdos se mostra praticamente categórica (básico – 91,67%, m=12; avançado – 98,33%, m=12). Testes de Mann-Whitney não demonstraram haver diferenças significativas entre os dois grupos, em suas identificações dos segmentos como surdos ($U=76,500$; $p=,051$) ou sonoros ($U=100,00$; $p=,175$). Este resultado confirma o observado em estudos prévios (ALVES; SCHWARTZHAUPT; BARATZ, 2011; ALVES; ZIMMER, 2012), que sugerem que segmentos aspirados do inglês são facilmente identificados pelos aprendizes. Ressaltemos, nesse sentido, que os segmentos do inglês, assim como as plosivas surdas de nossa língua, são produzidos com grande força de explosão – dessa forma, mesmo que o VOT não seja a pista prioritariamente usada pelos aprendizes, as respostas tenderão a apontar, praticamente de forma categórica, tais segmentos como surdos. Com base nesses resultados, confirma-se, assim, a Hipótese 1: independentemente de qual a pista majoritariamente seguida, não houve diferenças entre os dois grupos de proficiência no que diz respeito à identificação do vozeamento referente aos padrões de VOT Negativo e VOT Positivo.⁷

No que diz respeito ao padrão Zero Não Manipulado, verificamos que os aprendizes em ambos os níveis de proficiência preferem identificá-lo como sonoro (básico: 69,71%, m= 8,00; avançado: 71,67, m=10,00). Testes de Mann-Whitney demonstraram não haver uma diferença significativa entre os dois grupos, no que diz respeito à identificação dos segmentos como surdos ($U=111,00$; $p=,727$) ou sonoros ($U=103,00$; $p=,510$). Estes resultados vão ao encontro do que era por nós hipotetizado, uma vez que prevíamos que não haveria diferenças significativas entre os dois grupos na identificação do vozeamento das consoantes com esse padrão. Entretanto, comentários adicionais precisam ser feitos com relação à motivação da nossa hipótese inicial: conforme havíamos previsto, os dois grupos tenderiam a identificar tal padrão como vozeado por diferentes vias: os aprendizes de nível básico fundamentariam suas respostas em pistas que não o VOT (tal como a intensidade da explosão), enquanto que os participantes de nível avançado, por já atentarem ao VOT como pista principal e por já terem aprendido que o padrão de VOT Zero Não Manipulado equivaleria a consoantes sonoras no inglês, fariam uso dos padrões da L2 para responder à tarefa. Cabe, dessa forma, verificar se as respostas fornecidas pelos aprendizes fundamentam-se, efetivamente, nessas possibilidades levantadas. Para tal tarefa, a observação do padrão Zero Manipulado, a seguir, se mostrará de grande valia.

⁷ Nas discussões da identificação dos próximos padrões, retomaremos essa questão para sugerirmos que o VOT não é efetivamente seguido por nenhum dos dois grupos de proficiência.

Com relação ao padrão VOT Zero Manipulado, os dados da Tabela 01 nos mostram que, frente a esse padrão artificial, os aprendizes apresentam maior dificuldade ao tentarem identificar os segmentos como 'surdos' ou 'sonoros'. Essa indefinição mostra-se ainda mais clara no nível básico, em que 57,29% dos dados (m=6,00) são identificados como surdos, ao passo em que 39,95% dos *tokens* (m= 5,00) são identificados como sonoros. A preferência fica um pouco mais clara entre aprendizes de nível avançado, cujos 76,67% dos dados são identificados como surdos (m= 9,00). De acordo com a Hipótese 3, esperávamos que os aprendizes de nível avançado já seguissem a pista de VOT e, por conseguinte, já identificassem tais segmentos como sonoros, diferentemente do que ocorreria com os aprendizes básicos, que identificariam o Zero Manipulado como surdo por seguir pistas como a intensidade de explosão do segmento. Isso não ocorreu em nossos dados; pelo contrário, aprendizes avançados apresentam, ainda que discretamente, índices mais altos de preferência pela identificação de tais segmentos como /p/, /t/, /k/. Nossa terceira hipótese, portanto, não foi confirmada. Mais do que isso, os dados aqui relatados sugerem, também, pistas outras, além do VOT, que exercem ação na identificação das consoantes surdas e sonoras, de modo que os participantes não tomem por base unicamente a presença ou ausência de aspiração para fornecer suas respostas. Cabe mencionar que o mesmo experimento, quando realizado com falantes nativos de inglês, obteve altos índices de identificação do padrão Zero Manipulado como sonoro, confirmando a tendência de que os falantes nativos de inglês se baseiam na ausência/presença de aspiração para identificar sonoridade, mesmo frente a segmentos de natureza híbrida (SCHWARTZHAUPT; ALVES; FONTES, 2013).

Tal constatação leva-nos a retomar, ainda, os resultados referentes ao Padrão Zero Natural, para o qual não havíamos previsto diferença significativa entre grupos, ainda que os aprendizes de diferentes proficiências baseassem suas respostas em diferentes pistas acústicas. De fato, considerando-se que o padrão VOT Zero se encontra presente nas consoantes /p/, /t/ e /k/ do sistema de L1 dos aprendizes,⁸ poderíamos ter previsto uma preferência pela identificação das consoantes com padrão Zero Não Manipulado como surdas, o que tampouco foi verificado nos dados. A identificação desse padrão como sonoro também serve, de antemão, como um argumento para a possibilidade de os aprendizes estarem sendo guiados por outros aspectos além do VOT, ao basearem suas respostas. Ademais, os dados do padrão Zero Manipulado parecem indicar que, independentemente do grau de proficiência dos aprendizes, a duração de VOT não é a pista considerada para as respostas providas pelos aprendizes. Nesse sentido, a ausência de diferenças significativas entre os dois grupos talvez

⁸ E, conforme já expresso anteriormente, pelo menos em menor grau, em alguns casos de produção de /b/, /d/ e /g/, no dialeto gaúcho, o que representaria uma evidência adicional para o fato de o VOT não constituir, tampouco nos seus dialetos de L1, a pista principal entre esses aprendizes.

não se deva ao fato de que os aprendizes estejam seguindo pistas prioritárias diferentes, que, ainda assim, resultem em um mesmo padrão de identificação, mas, sim, que, independentemente do nível de proficiência dos aprendizes, o VOT não constitui a pista prioritária para a distinção entre segmentos surdos e sonoros. Frente a essa constatação, concluímos, também, que tampouco os padrões VOT negativo e VOT positivo são identificados como sonoro e surdo, respectivamente, através do VOT, independentemente do nível de proficiência dos aprendizes.

Evidências adicionais para a constatação de que o VOT não corresponde à pista prioritariamente atentada por nenhum dos dois grupos de proficiência poderão ser verificadas na seção que segue, ao tratarmos dos dados de Discriminação.

Discriminação

Apresentamos, nesta seção, os resultados referentes ao Teste de Discriminação, detalhados na tabela 02. No que diz respeito ao contraste VOT Negativo vs. VOT Positivo, os dados da Tabela 02 nos mostram altos índices de discriminação nos níveis básico (acuidade=76,74%, $m=9,00$; igualdade=9,72%, $m=.50$) e avançado (acuidade=93,33%, $m=11,50$; igualdade=2,5%; $m=0$). Testes de Wilcoxon não apontaram diferenças significativas entre os grupos (Acuidade: $U=76,500$, $p=.087$; Igualdade: $U=79,000$, $p=0,082$). De fato, conforme havíamos previsto, independentemente se fosse por meio da mesma pista acústica ou por pistas acústicas diferenciadas (ou seja, em que apenas o grupo avançado atentaria à pista acústica do VOT), os resultados se mostrariam os mesmos, de modo que os padrões de VOT Negativo e VOT Positivo fossem altamente discriminados.

No que tange ao contraste entre o VOT Negativo vs. Zero Manipulado, esperávamos encontrar uma diferença significativa entre grupos, uma vez que, ao seguirem o VOT como pista prioritária, os aprendizes de nível avançado não diferenciariam entre os dois padrões (pois, no falar nativo, tanto o VOT Negativo quanto o Zero Natural são usados nas produções das plosivas sonoras), ao passo que os aprendizes de grau básico, ao serem guiados por pistas tais como a intensidade do *burst*, discriminariam entre o VOT Negativo (com explosão fraca) e o Zero Manipulado (que, apesar de ter um curto intervalo de VOT, apresentava uma explosão forte). Contudo, os resultados da tabela 02 parecem ir de encontro à quinta hipótese, pois apontam para índices de respostas de igualdade ainda mais baixos para aprendizes de nível avançado do que para os básicos.

Tabela 2 – Resultados do teste de discriminação (12 questões por contraste).⁹

Padrões de VOT	Grupo Básico		Grupo Avançado	
	Acuidade	Igualdade	Acuidade	Igualdade
Negativo x Zero Artificial	45,49% 131/288	29,51% 85/288	64,17% 77/120	25% 30/120
Negativo x Positivo	76,74% 221/288	9,72% 28/288	93,33% 112/120	2,5% 3/120
Zero Artificial x Positivo	34,03% 98/288	45,49% 131/288	38,33% 46/120	50,83% 61/120

Fonte: Elaboração própria a partir do banco de dados dos autores.

De fato, ambos os grupos (com uma supremacia do grupo avançado – 64,17%, $m=8,00$) tendem a julgar os dois padrões como diferentes. Tal fato constitui argumento adicional para a proposta de que o VOT não está sendo tomado como pista prioritária para a distinção entre segmentos surdos e sonoros, e ainda deixa claro que tal fato não abrange unicamente aprendizes de nível básico, mas, também, participantes com grau avançado de proficiência.

Finalmente, observemos os dados referentes ao contraste VOT Zero Manipulado vs. VOT Positivo. Os dados da Tabela 02 evidenciam que esse foi o contraste para o qual os aprendizes se mostraram mais inseguros em suas respostas, ao considerarmos tanto os aprendizes de nível básico (acuidade = 34,03%, $m= 4,00$; igualdade = 45,49%, $m= 5,00$) quanto os de nível avançado (acuidade = 38,33%, $m = 5,50$; igualdade = 50,83%, $m= 6,50$), sendo que ambos os grupos apresentam uma leve preferência por considerar esses dois padrões como iguais. Testes de Mann-Whitney demonstraram não haver uma diferença estatisticamente significativa entre os dois grupos (acuidade: $U=104, 500$, $p=,555$; igualdade: $U=105,000$, $p=,569$). É contrariada, assim, nossa Hipótese 6, que previa que haveria uma diferença significativa entre os dois grupos, uma vez que esperávamos que, ao passo em que os aprendizes avançados, guiados pelo VOT, discriminassem entre os dois padrões, os participantes básicos apontassem um alto índice de igualdade entre os padrões. Entretanto, o que verificamos é que os aprendizes de ambos os níveis de proficiência, frente a tal contraste,

⁹ Na Tabela 02, 'acuidade' corresponde aos índices de respostas corretas fornecidas na tarefa AxB (por exemplo, em [p]at, [p]at and [b]at, os aprendizes deveria responder que X apresenta a mesma consoante que em A, não B); 'Igualdade1' corresponde às ocorrências em que os aprendizes não discriminaram X de A ou B, de modo a considerarem que as consoantes iniciais das três consoantes eram iguais. Nesta tabela, não apresentamos os índices de respostas incorretas, que correspondem àquelas escolhas em que os aprendizes escolheram a opção A ao invés da resposta correta B, ou vice-versa.

tendem a apresentar a mesma preferência, de modo a preferirem a responder que os dois padrões se mostram iguais.

Em suma, os resultados do teste de Discriminação corroboram os índices previamente verificados nos resultados do teste de Identificação, de modo a corroborar que o VOT não se mostra como a pista prioritariamente seguida por aprendizes de nível básico e, nem mesmo, pelos aprendizes de grau avançado na LE, para os quais pensávamos que, assim como verificado entre falantes nativos (SCHWARTZHAUPT; ALVES; FONTES, 2013), a presença/ausência de aspiração corresponderia ao fator preponderante para as respostas por eles fornecidas. Na seção que segue, verificaremos as possíveis implicações de tais resultados nos dados de produção dos segmentos.

Teste de Produção

Os resultados do Teste de Produção em Português Brasileiro (PB) são apresentados na Tabela 03, na qual são reportados valores de VOT surpreendentemente altos para o Português Brasileiro. Ainda que estudos prévios (GEWEHR-BORELLA; ZIMMER; ALVES, 2011; VEIGA-FRANÇA, 2011; ALVES; SCHWARTZHAUPT; BARATZ, 2011; SCHWARTZHAUPT, 2012) já tenham evidenciado a possibilidade de 'semiaspiração' da plosiva velar no dialeto gaúcho, surpreende-nos, sobretudo, os valores médios de 24,33 ms ($m = 24,00$) no nível básico e 33,00 ($m = 33,5$) no nível avançado, no que diz respeito à produção de palavras iniciadas pelo segmento bilabial /p/.

Tabela 3 – Resultados do teste de produção de palavras em PB

Consoante	Básico (24)		Avançado (10)	
	<i>Tokens</i>	Média (DP)	<i>Tokens</i>	Média (DP)
/ p /	133	24,13 (5,44)	51	33,00 (7,18)
/ k /	139	55,71 (20,09)	57	51,7 (23,09)

Fonte: Elaboração própria a partir do banco de dados dos autores.

Frente a esses dados, não se pode negar a possibilidade de que, sobretudo entre os aprendizes de nível avançado, a fala em L1 dos aprendizes esteja sofrendo transferência dos padrões de VOT da L2 (SANCIER; FOWLER, 1997; COHEN, 2004), de modo que os intervalos de VOT em L1 expressos na Tabela 03 não reflitam as durações efetivamente produzidas por falantes monolíngues do dialeto do Sul do Brasil. Tal possibilidade, de fato, não se mostra imprevisível frente a uma perspectiva dinâmica de aquisição de linguagem, de acordo com

a qual qualquer mudança em um dos sistemas linguísticos do falante podem implicar modificações substanciais em todos os demais sistemas linguísticos, refletindo-se, portanto, em termos de influência não somente da L1 para a L2, mas, também, da L2 para a L1, ou sobre as diversas línguas utilizadas pelo aprendiz (L3 e assim por diante), (DE BOT; LOWIE; VERSPOOR, 2007; BECKNER et al., 2009; BLANK, 2013).

A observação dos dados da Tabela 4, referente à produção de palavras da língua inglesa, deixa claros valores mais altos para /p/ (básico: 45,04, m=45,50; avançado: 34,4, m=31,5) e /k/ (básico: 68,87, m=67,50; avançado: 79,8, m=82,5) na L2 do que na L1. Corrobora-se, parcialmente, a nossa Hipótese 7, que previa que, em cada um dos níveis de proficiência, haveria uma diferença estatisticamente significativa entre os valores de VOT encontrados na L2 e aqueles encontrados na L1. De fato, testes de Wilcoxon demonstraram haver uma diferença significativa nas produções tanto no nível básico ($Z=-2,702$, $p=,007$) quanto em nível avançado ($Z=-2,193$, $p=,028$) nas produções de /k/, resultado esse não plenamente verificado no caso de /p/, em que uma diferença significativa foi encontrada, apenas, na comparação dentro do nível básico (/p/: $Z=-4,03$, $p=,000$), provavelmente pelo fato de os intervalos de VOT dessa consoante se mostrarem bastante altos já na própria L1 dos aprendizes de grau avançado.

Tabela 4 – Resultados do teste de produção de palavras em língua inglesa

Consoante	Básico (24)		Avançado (10)	
	<i>Tokens</i>	Média (DP)	<i>Tokens</i>	Média (DP)
/ p /	136	45,04 (16,71)	53	34,40 (15,25)
/ t /	131	59,04 (13,56)	57	58,40 (17,52)
/ k /	131	68,87 (21,42)	60	79,80 (14,95)

Fonte: Elaboração própria a partir do banco de dados dos autores.

Questionamentos poderiam ser feitos a respeito das diferenças significativas encontradas. Afinal, já que o VOT, conforme discutido nos resultados referentes aos testes de percepção, não se mostra como a pista prioritariamente adotada pelos aprendizes, como explicar o fato de as produções em L2 já se encontrarem com intervalos de aspiração mais longos do que aqueles encontrados na L1? Como resposta a esse questionamento, julgamos necessário considerar, primeiramente, o fato de os participantes não tomarem o VOT como pista acústica prioritária, conforme discutido nas seções anteriores. Isso não implica que tais aprendizes não possam perceber e reconhecer a pista em questão como um aspecto da fonologia do inglês. Em outras palavras, é possível que a aspiração já seja percebida como um

detalhe alofônico, necessário para a produção de fala “sem sotaque” do inglês, mas que não seja tomada como pista acústica funcionalmente prioritária na distinção fonológica entre segmentos surdos e sonoros, conforme já diferenciamos. Dessa forma, os aprendizes poderiam, até certo grau, produzir a aspiração como um detalhe de redução de sotaque, porém uma necessidade de produção de tal pista como necessária para o estabelecimento de distinções fonológicas ainda não se faria sentir. Tal distinção funcional estaria sendo instanciada, dessa forma, a partir da ação conjunta de múltiplas pistas acústicas, corroborando uma concepção dinâmica de aquisição de linguagem, ou, ainda, a partir de uma pista prioritária outra que não o VOT.

Além disso, também é necessário levar em conta que, apesar de apontadas tais diferenças estatisticamente significativas, os dados sugerem que, ainda que as durações de aspiração na L2 se mostrem mais longas do que aqueles intervalos de tempo encontrados nos dados de L1, as produções em L2 não são, ainda, semelhantes ao padrão nativo. Conforme explicam Cho e Ladefoged (1999), os valores de VOT padrão, na língua inglesa, equivalem a 55 ms para /p,/ 70 ms para /t/ e 80 ms para /k/. Ao verificarmos que, ainda que os aprendizes já não mais produzam os valores de VOT que realizam na L1, os intervalos de duração de aspiração apresentam, na L2, um valor intermediário entre L1 e L2, o que serve como argumento adicional para o fato de que o VOT possui, entre os aprendizes, um caráter de importância alofônica, mas não prioritário, para a distinção funcional entre segmentos surdos e sonoros.

Foi justamente o valor intermediário de VOT, já vastamente descrito pela literatura de interlíngua português-inglês (ALVES; SCHWARTZHAUPT; BARATZ, 2011), que nos motivou elaborar nossa oitava e última hipótese. De acordo com a Hipótese 8, não haveria diferenças significativas entre as durações de VOT produzidas por aprendizes básicos e avançados de inglês. Ao formularmos tal hipótese, encontramos embasamento não somente nos dados empíricos dos trabalhos supracitados, mas, também, na própria concepção dinâmica de aquisição de linguagem que norteia o presente estudo. Considerando-se que, embora o aprendiz seja capaz de perceber e discriminar os sons da língua-alvo, a produção dos segmentos da língua-alvo exige que o aprendiz consiga abandonar o *timing* dos articuladores da L1, de modo a ser capaz de orquestrar os articuladores de acordo com a temporalidade e o ritmo da L2 (ZIMMER; ALVES, 2012). Assim, parecia-nos possível que, a despeito do fato de os aprendizes proficientes já seguirem o VOT como pista prioritária, diferenciando segmentos surdos de sonoros em função da presença/ausência de aspiração, esses participantes com proficiência apropriada ainda não teriam adquirido a temporalidade do VOT na língua estrangeira, uma vez que a percepção dos padrões de VOT implicaria um requisito necessário, mas não suficiente, para a produção de aspiração.

Testes de Mann-Whitney demonstraram não haver, realmente, diferenças significativas entre níveis de proficiência para /p/ (U=74,000, p=,082), /t/ (U=114,500, p=,835) ou /k/ (U=84,000, p=,173). A discussão já realizada sobre os dados de percepção na seção anterior, entretanto, nos possibilita considerar outra explicação para esses dados: uma vez que, conforme visto, o VOT não se mostra como a pista prioritária para a identificação e a discriminação de segmentos em nenhum dos dois níveis de proficiência, tal fato provavelmente esteja sendo refletido nos próprios dados de produção. Em outras palavras, a pouca diferença entre os dois níveis de proficiência, no que concerne aos intervalos de VOT produzidos, não diz respeito unicamente à dificuldade de adquirir a temporalidade da L2, mas, sobretudo, provavelmente se deve ao fato de, mesmo em termos perceptuais, a distinção entre segmentos surdos e sonoros estar sendo instanciada por pistas acústicas outras além do VOT. Uma vez que o VOT, pelo menos entre os aprendizes aqui investigados, não parece ser a pista prioritária para a percepção dos sons, tampouco tal detalhe fonético será o fundamental para a distinção entre os sons produzidos por esses aprendizes, de modo que percepção e produção se encontrem altamente relacionados.

Os aprendizes, então, podem estar assumindo que a não produção, ou a produção parcial, da duração longa do VOT no inglês não necessariamente desempenhará papéis detrimenais para a inteligibilidade. Tal fato pode ser reforçado em um contexto de comunicação entre brasileiros (o que é o caso da maior parte dos contextos de instrução em língua estrangeira em nosso país, em que os aprendizes tendem a compartilhar a mesma L1), que, ao compartilharem das mesmas pistas acústicas majoritárias, dispensariam da duração do VOT para a distinção entre /p/, /t/, /k/ e /b/, /d/, /g/. Uma efetiva necessidade do emprego majoritário de tais pistas somente se faria sentida em casos de comunicação entre falantes brasileiros e falantes nativos (ou aprendizes) de inglês que fazem efetivo uso da pista de VOT para o estabelecimento da distinção funcional entre os sons, ou, a partir de um trabalho de instrução explícita acerca do fenômeno (MOTTA; ALVES, 2013), cujos efeitos ainda precisam ser mais amplamente investigados.

Em suma, os resultados do teste de produção evidenciam que, na língua-alvo, os valores de VOT produzidos pelos aprendizes de ambos os níveis de proficiência já se mostram superiores aos da L1. Tais valores, entretanto, não se mostram ainda semelhantes aos encontrados no falar nativo, o que nos levou à sugestão de que, embora o aprendiz já reconheça a necessidade de produção de um VOT mais longo para a redução do sotaque, de modo que tal pista acústica adquira um caráter “alofônico”, tal detalhe fonético ainda não é tomado, pelos participantes, como o principal para a distinção entre segmentos surdos e sonoros.

Considerações finais

Neste trabalho, partimos da premissa inicial de que aprendizes básicos de inglês não tomam o VOT como pista primordial para a distinção entre segmentos surdos e sonoros do inglês. Todas as hipóteses elaboradas relacionavam-se à ideia de que haveria uma diferença entre aprendizes de nível básico e avançado de proficiência, uma vez que, enquanto o VOT não se caracterizaria como a pista principal para iniciantes na L2, participantes de nível avançado de proficiência, por sua vez, fariam uso da presença/ausência de aspiração para distinguir segmentos surdos de sonoros do inglês, tanto em termos de percepção quanto de produção.

Os resultados apresentados neste trabalho, entretanto, evidenciam que, apesar de nossas expectativas iniciais, nem mesmo os participantes com alto nível de proficiência parecem ter feito uso do VOT como pista principal para a distinção entre /p, t, k/ e /b, d, g/. Independentemente do grau de adiantamento do aprendiz, parece-nos que, sem instrução formal, os aprendizes continuam fazendo uso das pistas acústicas que se mostram prioritárias para a distinção de sonoridade em seus sistemas de L1. Tal fato refletiu-se, conforme evidenciado na seção anterior, tanto nos dados de percepção quanto nos de produção.

É preciso deixar claro, entretanto, que o objetivo do presente trabalho foi mostrar se o VOT poderia ser caracterizado, entre aprendizes de VOT, como uma pista suficiente para as distinções de vozeamento. Ao termos verificado a importante ação de outra(s) pista(s) acústica(s), estudos adicionais precisam, ainda, ser desenvolvidos, para que possamos determinar, claramente, quais aspectos acústicos estão em jogo. Interessa sobretudo determinar se a distinção de sonoridade, para tais aprendizes, se instancia através do somatório de diversas pistas acústicas que agem em conjunto, ou, ainda, se há uma pista acústica que se mostra prioritária e autossuficiente, que, tomada unicamente, se mostre capaz de dar conta de tais distinções. Conforme já mencionado, estudos de Sundara (2005), Oh (2011) e Kong, Beckman e Edwards (2012) levam-nos a sugerir que pistas como a intensidade da explosão e as transições de F0 desempenham um papel fundamental no que diz respeito a essa questão, em termos acústicos. Ao ressaltarmos a importância de estudos futuros sobre a questão, cabe-nos evidenciar a importância do presente estudo, por ter confirmado nossa hipótese de que o VOT, tomado isoladamente, não se mostra como uma pista suficiente para a distinção de sonoridade entre aprendizes brasileiros de inglês.

Os resultados aqui apresentados vão ao encontro de uma perspectiva gestual, de acordo com a qual a moeda comum, tanto da percepção quanto da produção, seria o gesto fonológico (GOLDSTEIN; FOWLER, 2003). Considerando-se o fenômeno analisado neste trabalho, a simples exposição à pista do VOT não se mostrou suficiente para que os aprendizes adquirissem o gesto da L2. Uma possível

explicação talvez resida no fato de que, conforme apontado por Goldstein e Fowler (2003), vários são os estudos, na própria aquisição de L1, que apontam para uma dificuldade maior de aquisição de contrastes referentes a ações de órgãos menos visíveis. No caso do presente estudo, o papel distintivo da laringe possivelmente não foi adquirido em função de o aprendiz basear suas distinções fonológicas em constelações gestuais e *timing* (duração ou tempo de execução/faseamento gestual) proceduralizados em sua L1. A formação gestual que resultaria na aspiração, para esses aprendizes, pode estar assumindo um caráter meramente alofônico, de modo que, frente à ação de outros órgãos envolvidos na pauta gestual, para os aprendizes, na língua-alvo, a ação da laringe não venha a assumir um caráter distintivo. É possível que, no caso da L2, a abstração dos movimentos usados para atingir determinada articulação na L2 sofra a influência da abstração gestual da L1, já automatizada pela memória procedimental. Assim, no caso do VOT, o aprendiz pode muito bem interpretar que uma duração mais longa da laringe com grau de constrição aberta não assumiria um caráter distintivo, mas, sim, estaria assumindo o mesmo papel frente a um único objetivo articulatório, que corresponde ao da sua língua materna.

Os resultados deste estudo chamam a atenção para uma série de outros artigos acerca do tema, cujas questões têm sido investigadas pelos membros de nosso grupo de pesquisa. Primeiramente, a observação dos dados de produção deste trabalho deixa claro que efeitos da L2 sobre a produção na língua materna precisam ser mais amplamente investigados. Além disso, não sendo o VOT a principal pista seguida por brasileiros, julgamos necessário, portanto, chamar a atenção do aprendiz para que ele passe a atentar para a importância da pista acústica do VOT na distinção entre plosivas surdas e sonoras na língua-alvo, estudos que versam sobre o papel da instrução explícita (MOTTA; ALVES, 2013) e do treinamento perceptual se fazem, portanto, de grande relevância. Nesse sentido, podemos mencionar outra contribuição da área de Aquisição Fonético-Fonológica de LE: contribuir, direta ou indiretamente, com a Linguística Aplicada de Ensino de Línguas Estrangeiras (ALVES, 2012). As conclusões obtidas no presente trabalho, portanto, abrem caminho para uma série de discussões e uma vasta agenda de investigações acerca do papel de pistas acústicas majoritárias, tais como o VOT, nas distinções funcionais em sistemas de língua materna e de interlíngua.

ALVES, U. K.; ZIMMER, M. C. Perception and production of English VOT patterns by Brazilian learners: the role of multiple acoustic cues in a DST perspective. **Alfa**, São Paulo, v.59, n.1, p.157-180, 2015.

- **ABSTRACT:** *In this study, departing from a dynamic conception of L2 phonetic-phonological acquisition, we investigate 32 Southern Brazilian learners' perception (identification and discrimination) and production of VOT patterns of initial stops in English. We initially*

hypothesized that, especially among learners with a basic level of L2 proficiency, VOT was not the main acoustic cue employed in the perception of voicing distinctions. Our results show that, regardless of the learners' proficiency level (basic or advanced), VOT is not a sufficient cue for the distinction between /p/, /t/, /k/ and /b/, /d/, /g/. These results, which have an influence on the lower VOT values found in our production data, conform with a dynamic view of L2 acquisition, according to which multiple acoustic cues play a role in language acquisition, forcing learners to tune in to the most important cue(s) in the target language.

- **KEYWORDS:** VOT. Second language acquisition. Acoustic cues.

REFERÊNCIAS

ABRAMSON, A.; LISKER, L. Voice-Timing Perception in Spanish Word-Initial Stops. **Journal of Phonetics**, London, n.1, p.1-8, 1973.

ALVES, U. K. Pesquisa em aquisição de L2 e ensino: um relacionamento possível (mas não necessariamente garantido). In: LEFFA, V.; ERNST, A. (Org.) **Linguagens: Metodologias de Ensino e Pesquisa**. Pelotas: EDUCAT, 2012. P.233-252.

ALVES, U. K.; SCHWARTZHAUPT, B. M.; BARATZ, A. H. Percepção e produção dos padrões de VOT do inglês (L2) por aprendizes brasileiros. In: FERREIRA-GONÇALVES, G.; BRUM-DE-PAULA, M. R.; KESKE-SOARES, M. **Estudos em Aquisição Fonológica**. Pelotas: Ed. da UFPel, 2011. p.3-4.

ALVES, U. K.; ZIMMER, U. K. The Dynamics of Perception and Production of VOT Patterns in English by Brazilian Learners. In: MELLO, E.; PETTORINO, M.; RASO, T. (Ed.). **Proceedings of the VIIth GSCP International Conference: Speech and Corpora**. Firenze: Firenze University Press, 2012, p.223-227.

ANTONIOU, M. et al. Inter-Language Interference in VOT Production by L2-Dominant Bilinguals: Asymmetries in Phonetic Code-Switching. **Journal of Phonetics**, London, v.39, p.558-570, 2011.

AUDACITY. Software livre. Disponível em: <www.audacity.sourceforge.net>. Acesso em: 20 jan. 2015.

BECKNER, C. et al. Language is a Complex Adaptive System: Position Paper. **Language Learning**, Ann Arbor, v.59, suppl.1, p.1-26, 2009.

BEST, C. T.; TYLER, M. D. Nonnative and Second-Language Speech Perception: Commonalities and Complementarities. In: BOHN, O.-S.; MUNRO, M. J. **Language Experience in Second Language Speech Learning: Studies in Honor of James Emil Flege**. Amsterdam: John Benjamins, 2007. p.13-34.

BLANK, C. A. **A influência grafo-fônico-fonológica na produção oral e no processamento de priming em multilíngues**: uma perspectiva dinâmica.

2013. 226f. Tese (Doutorado em Letras) – Programa de Pós-Graduação em Letras, Universidade Católica de Pelotas, Pelotas, 2013.

BOERSMA, P.; WEENINK, D. **Praat**: Doing Phonetics by Computer. Version 5.3.48. 2013. Disponível em <www.praat.org>. Acesso em: 20 jan. 2015.

BROWMAN, C. P.; GOLDSTEIN, L. Competing Constraints on Intergestural Coordination and Self-Organization of Phonological Structures. **Bulletin de la Communication Parlee**, Cedex, n.5, p. 25-34, 2000.

_____. Dynamics and Articulatory Phonology. In: VAN GELDER, T.; PORT, R. F. (Ed.). **Mind as motion**. Cambridge: MIT Press, 1993, p.51-62.

_____. Articulatory Phonology: An overview. **Phonetica**, Basel, n.49, p.155-180, 1992.

CHO, T.; LADEFOGED, P. Variation and Universals in VOT: Evidence from 18 Languages. **Journal of Phonetics**, London, n.27, p.207-229, 1999.

COHEN, G. **The VOT Dimension**: a Bidirectional Experiment with English and Brazilian-Portuguese Stops. 2004. 96f. Dissertação (Mestrado em Língua Inglesa) – Universidade Federal de Santa Catarina, Florianópolis, 2004.

DE BOT, K.; LOWIE, W.; VERSPOOR, M. A Dynamic Systems Theory approach to second language acquisition. **Bilingualism: Language & Cognition**, Cambridge, v.10, n.1, p.7-21, 2007.

DOCHERTY, G. J. **The Timing of Voicing in British English Obstruents**. Berlin; New York: Foris Publications, 1992.

GEWEHR-BORELLA, S.; ZIMMER, M. C.; ALVES, U. K. Transferências grafo-fônico-fonológicas: uma análise de dados de crianças monolíngues (Português) e bilíngues (Hunrückisch-Português). **Gragoatá**, Niterói, v.30, p.201-219, 2011.

GOLDSTEIN, L.; FOWLER, C. A. Articulatory Phonology: a Phonology for Public Language Use. In: MEYER, A. S., SCHILLER, N. O. (Ed.). **Phonetics and Phonology in Language Comprehension and Production**: Differences and Similarities. Berlin: Mouton de Gruyter, 2003. p.159-207.

KAMIANECKY, F. **A palatalização das oclusivas dentais /t/ e /d/ nas comunidades de Porto Alegre e Florianópolis**: uma análise quantitativa. 2002. 114f. Dissertação (Mestrado em Letras) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2002.

KONG, E. J.; BECKMAN, M. E.; EDWARDS, J. Voice Onset Time is Necessary but not Always Sufficient to Describe Acquisition of Voiced Stops: The Cases of Greek and Japanese. **Journal of Phonetics**, London, v.40, p.725-744, 2012.

LISKER, L.; ABRAMSON, A. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. **Word**, New York, n.20, p.384-422, 1964.

MOTTA, C.; ALVES, U. K. Percepção de padrões de Voice Onset Time por aprendizes brasileiros de inglês: dados de discriminação e identificação. In: JORNADA DE JÓVENES LINGUISTAS, 2., 2013, Buenos Aires. **Resumos...** Buenos Aires, 2013. v.1. p.140.

OH, E. Effects of speaker gender on voice onset time in Korean stops. **Journal of Phonetics**, London, n.39, p.59-67, 2011.

POLLITT, A. **The meaning of OOPT Scores**. 2007. Disponível em: <https://www.oxfordenglishtesting.com/uploadedFiles/Buy_tests/oopt_meaning.pdf>. Acesso em: 20 jan. 2015.

PURPURA, J. **The Oxford Online Placement Test: What does it Measure and How?**. 2007. Disponível em: <http://www.oxfordenglishtesting.com/uploadedfiles/6_New_Look_and_Feel/Content/oopt_measure.pdf>. Acesso em: 26 ago. 2013.

SANCIER, M. L.; FOWLER, C. A. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. **Journal of Phonetics**, London, n.25, p.421-436, 1997.

SCHWARTZHAUPT, B. **Factors influencing Voice Onset Time: analyzing Brazilian Portuguese, English and Interlanguage data**. 2012. 65f. Trabalho de Conclusão de Curso (Graduação em Letras) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

SCHWARTZHAUPT, B.; ALVES, U. K.; FONTES, A. B. L.; O VOT como pista suficiente para a distinção surdo/sonoro: dados de falantes do inglês americano. In: BRUM DE PAULA, M. (Org.). **4º Seminário de Aquisição Fonológica: Resumos e Programação**. Pelotas: Ed. da UFPel, 2013. p.26.

SIMON, E. **Voicing in Contrast: Acquiring a Second Language Laryngeal System**. Ghent, Belgium: Academia Press, 2010.

SUNDARA, M. Acoustic phonetics of coronal stops: a cross-language study of Canadian English and Canadian French. **Journal of the Acoustical Society of America**, New York, n.118, p.1026-1037, 2005.

VEIGA-FRANÇA, K. V. **A aquisição da aspiração das plosivas surdas do inglês por falantes do Português Brasileiro: Implicações teóricas decorrentes de duas formas de descrição dos dados**. 2011. 100f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Letras, Universidade Católica de Pelotas, Pelotas, 2011.

YAVAS, M.; WILDERMUTH, R. The effects of place of articulation and vowel height in the acquisition of English aspirated stops by Spanish speakers. **IRAL**, Heidelberg, n.44, p.251-263, 2006.

ZIMMER, M. C.; ALVES, U. K. Uma visão dinâmica da produção da fala em L2: o caso da Dessonorização Terminal. **Revista da Abralín**, Brasília, v.11, n.1, p.221-272, 2012.

Recebido em janeiro de 2014

Aprovado em abril de 2014