

Hipóteses acerca do reconhecimento de acento lexical em palavras isoladas¹

Fernanda Consoni¹, Pablo Arantes² Plínio A. Barbosa³, Waldemar Ferreira Netto⁴

¹ FFLCH -USP
fernanda.consoni@gmail.com

² IEL- UNICAMP
pabloarantes@gmail.com

³ IEL-UNICAMP

⁴ FFLCH- USP

Abstract. *In this paper we draw hypotheses about how listeners recognize isolated spoken words out of acoustic cues concerning lexical stress position in Brazilian Portuguese. An experiment was carried out in which two syllable fragments containing incomplete stress cues were auditorily presented to subjects. Their task was to identify from which word the fragment heard was excised. All words were chosen from paradigms similar to the one illustrated by hábito/habito/habitar. Error rates in the identification task were analysed. Duration, fundamental frequency and intensity were also analysed in order to investigate in what extent they play a role in the identification task.*

Keywords.

Resumo. *O artigo traz hipóteses acerca das estratégias utilizadas pelos ouvintes no reconhecimento de palavras isoladas a partir de pistas acústicas sobre a posição do acento lexical. O experimento consistiu em solicitar que sujeitos reconhecessem palavras, retiradas de paradigmas como hábito/habito/habitar, que foram previamente cortadas e deram origem a fragmentos com diferentes combinações de sílabas tônicas e átonas. Para determinar quais parâmetros acústicos serviriam de guia para o julgamento dos sujeitos, a duração das unidades V-V, a frequência fundamental e intensidade foram medidas nos estímulos experimentais. O confronto entre as respostas de identificação das palavras e as pistas ajudaram a elucidar de que maneira e em que extensão os parâmetros acústicos exercem influência sobre o reconhecimento de palavras.*

Palavras-chave. *Fonologia; Fonética; Acento Lexical.*

Introdução

Com o objetivo de investigar a relação entre a produção e o julgamento de acento lexical em português, variante de São Paulo, em palavras isoladas, propôs-se o experimento relatado nesse artigo. Em português brasileiro há uma distinção fonológica expressa pela distribuição dos graus de acentuação ao longo palavra, de modo que a posição do acento é um fator distintivo de significado (compare-se as palavras sábia,

sabia e sabiá, p.e.), como observado por vários autores, Gonçalves Vianna (1883), Câmara Jr (1979), Delgado Martins (1988), Massini-Cagliari (1999), entre outros.

A pergunta chave do experimento que foi proposto era: se forem apresentados fragmentos de palavras retiradas de paradigmas como máquina/maquina/maquinar ou pátina/patina/patinar contendo apenas parte da informação sobre a localização do acento lexical, poderiam os falantes reconstruir o restante da informação e reconhecer com acuidade as palavras que originaram os fragmentos?

1. Proposta de experimento

Propôs-se um teste de identificação de palavra no qual os sujeitos, após ouvirem as duas sílabas iniciais de palavras com três tipos acentuais, são solicitados a identificar qual era a palavra apresentada. Para tanto, criou-se um conjunto de setenta e cinco palavras com seqüências idênticas de fonemas nas duas primeiras sílabas e obedecendo ao seguinte arranjo: primeira sílaba átona e segunda tônica, doravante PT, (afônico, patina, p.e.), primeira sílaba tônica e segunda átona, doravante TA, (áfono, pátina, p.e.), primeira e segunda sílabas átonas, doravante PP, (afonia, patinar, p.e.).

A hipótese levantada aqui era que, mesmo na falta do sinal acústico completo, o ouvinte fosse capaz de reconhecer palavras a partir da audição das seqüências PT e TA, uma vez que o acento lexical está presente no sinal acústico emitido. Por outro lado, nos questionamos se o mesmo seria capaz de fazer a identificação correta para seqüências como PP. Na ausência dos parâmetros acústicos característicos do acento lexical, qual seria a estratégia do falante para fazer o julgamento?

2. Metodologia

Para que se pudesse investigar a relação entre a produção e o julgamento de acento lexical em PB, elaborou-se um teste no qual seqüências de sons eram apresentadas aos sujeitos, que deveriam escolher em um formulário de papel qual palavra julgavam ter ouvido entre três opções, sempre sobre o mesmo paradigma. (Modelo: Qual palavra você ouviu? Prática, Pratica, Praticar.). Foram apresentadas setenta e cinco seqüências de sons com intervalo, para resposta, de três segundos. Os testes foram realizados em ambientes silenciosos, com aparelho mini disk do fabricante Panasonic, modelo S-SL160. Os sujeitos selecionados para realizarem o experimento eram de ambos os sexos, com idade entre 20 e 40 anos e escolaridade em nível superior.

A gravação do corpus foi realizada por uma locutora. Para evitar a leitura isolada das palavras-alvo, montou-se a frase-veículo: "As palavras são _____, _____, _____ e jubilo". A palavra "jubilo" foi inserida no final de cada frase-veículo para evitar que os efeitos acústicos do declínio de intensidade e alongamento final incidissem sobre as palavras-alvo. Cada uma das palavras-alvo apareceu de forma aleatória em cada uma das três posições livres da frase-veículo, tendo sido evitado o aparecimento de seqüências como "pátina, patina, patinar". Escolheu-se entre as três versões gravadas de cada palavra aquela que figurava na posição medial, a menos que esta apresentasse algum ruído que comprometesse a audição.

As seqüências de sons foram obtidas através da edição de setenta e cinco palavras por meio do programa Cool Edit 96 de forma a conservarem apenas as duas primeiras sílabas. Tal edição permitiu criar um conjunto de seqüências idênticas de fonemas, alternando apenas a posição do acento para o caso de seqüências PT e TA, e a

ausência do acento para o caso PP. Houve a preocupação de selecionar palavras que contivessem a mesma consoante no ataque da terceira sílaba para evitar que a antecipação do movimento para a produção da terceira sílaba influísse no reconhecimento da palavra, de maneira que a identificação pudesse ser realizada sem que a informação sobre proeminência fosse julgada. Desse modo, apenas através das informações acústicas relativas ao acento é que os sujeitos poderiam fazer o julgamento.

O sujeito-locutor foi gravado numa sala silenciosa nas dependências da FFLCH/USP em sessão única, utilizando um microfone unidirecional headset. A digitalização do sinal acústico foi feita com taxa amostral de 44 kHz por meio do programa Cool Edit 96.

3. Resultados

A Tabela 1 resume os resultados obtidos no experimento de identificação de palavras. Cada combinação foi julgada 300 vezes. Cada coluna identifica uma combinação apresentada aos sujeitos e as linhas que vêm embaixo indicam os votos recebidos. Nas células preenchidas de cinza tem-se a quantidade de votos acertados. As duas células restantes na coluna indicam, os votos em cada padrão errado.

Tabela 1: dados de acerto/erro na tarefa de identificação de palavra

	<i>PP</i>	<i>TA</i>	<i>PT</i>
<i>PP</i>	115 (38,3 %)	9 (3 %)	61 (20, %)
<i>TA</i>	54 (18%)	260 (86, 7%)	20 (6,7 %)
<i>PT</i>	131 (43,7%)	31 (10,3)	219 (73 %)

3.1. Análise Acústica²

Os parâmetros acústicos que têm sido tradicionalmente investigados como correlatos dos fenômenos de acento, duração, intensidade e frequência fundamental (Fernandes, 1976, Delgado Martins 1986, Massini-Cagliari, 1992) foram obtidos através do programa *Praat* (Boersma e Weenink, 2005) para as palavras empregadas no experimento. O objetivo foi tentar deprender quais parâmetros influenciam o reconhecimento das palavras e em que medida eles o fazem.

Uma das principais hipóteses do trabalho é que a presença do acento em lugares diferentes nas combinações TA e PT e sua ausência no caso da combinação PP devem afetar a identificação da palavra que origina cada trecho. Os testes χ^2 para o padrão de acertos/erros conduzidos separadamente em cada combinação resulta significativo em todos os casos. No caso das combinações TA e PT, o teste χ^2 confirma o que indicam as porcentagens de acerto. Os sujeitos mostram uma tendência sistemática (i.e., acima do aleatório) de identificar corretamente a combinação apresentada. A significância no caso da combinação PP, pelo contrário, revela uma tendência sistemática de erro que representa 61,6 % do total de votos recebidos pela combinação. Estes resultados parecem mostrar que, no caso das combinações TA e PT, os parâmetros acústicos presentes nos trechos apresentados aos sujeitos são suficientes para garantir a identificação segura da palavra de onde eles foram retirados.

Nas combinações PT e TA pode-se observar que os erros somam apenas 27% e 13%, respectivamente, do total de julgamentos. Os testes envolvendo o padrão de acertos/erros, já relatados, nos permitem considerar que a tendência subjacente ao padrão no caso de PT e TA apontam para a existência de sistematicidade no acerto, permitindo considerar desprezível o erro. Apesar disto, os resultados dos testes χ^2 são significativos para as duas combinações – TA ($p = 0,005$) e PT ($p < 0,05$). O caso da combinação PP, ao contrário, impõe analisar com cuidado a preferência por um certo tipo de erro, uma vez que o teste envolvendo o padrão erros/acertos aponta uma tendência sistemática de erro, pois os julgamentos errôneos (61,6 %) representam a maioria dos votos. O teste realizado apenas considerando o número de erros divididos entre as combinações PT e TA é altamente significativo ($p < 0,05$). Considerando os 185 erros associados a esta categoria (PP), 70,8 % deles acontecem favorecendo a categoria PT. O número de votos recebidos pela combinação PT quando PP é apresentado, 131, é inclusive maior do que os votos corretos recebidos pela combinação PP, i.e., 115 votos. Há, claramente, um viés para confundir PP com PT. Há também a tendência menor, mas significativa ($p < 0,05$) mencionado acima de uma confusão no sentido contrário (confundir PP com PT). 75,3 % (61) dos erros associados à categoria PT (81) são votos em PP. Em resumo, dada a categoria PP o mais provável é que os sujeitos respondam erradamente que ouviram a palavra de onde se retirou a combinação PT. A confusão não é perfeitamente simétrica porque, dada a combinação PT, não há tendência significativa para identificações erradas. Mas quando o erro ocorre há favorecimento da confusão PP por PT. Deve-se buscar as bases para a explicação destas sistematicidades nos parâmetros acústicos das duas categorias confundíveis.

3.1.2 Análise dos parâmetros acústicos

Com o objetivo de verificar em que medida o comportamento dos parâmetros acústicos pode explicar os padrões de acertos/erros observados no teste de reconhecimento da palavra-fonte das combinações apresentadas, obteve-se medidas acústicas de duração (vogal, sílaba e unidade V-V), f_0 e intensidade nos trechos cruciais das 75 palavras empregadas no experimentoser Helvetica ou Arial, fonte 10, em negrito, com espaçamento de 06 pts antes e depois de cada legenda.

3.1.3.Duração

Os parâmetros duração das vogais e das sílabas são capazes de criar um contraste que pode ajudar a distinguir as combinações PT e TA, o que poderia explicar (i) a tendência significativa dos sujeitos de reconhecerem corretamente as duas combinações, ajudados talvez pela diferença de direção na mudança do parâmetro, ascendente no caso de PT e descendente no caso de TA e (ii) a tendência de não confundir estas duas categorias entre si.

TA é julgado como sendo PT em 10,3% dos casos e o contrário ocorre apenas 6,7%. O que estes dois parâmetros não esclarecem é porque há tantos erros no sentido de confundir PP com PT, erro que ocorre em 43,7,6% dos julgamentos errôneos de PP. Se apenas a duração de vogal ou de sílaba servissem de pista, o erro no julgamento de PP deveria estar melhor dividido entre as categorias PT e AT, dado que a duração das posições das vogais e das sílabas nas posições 1 e 2 não são diferentes entre si, no entanto esta distribuição homogênea não ocorre (conferir Tabela 1, primeira coluna de dados).

O parâmetro duração das unidades V-V³ pode ajudar a explicar o padrão de erros que a duração das vogais e sílabas não consegue. A derivada da duração de V-V da combinação PT é claramente ascendente e a de TA é estatisticamente nula. PP está numa situação intermediária, pois o valor de sua derivada é marginalmente diferente da nulidade. Esta situação poderia favorecer uma aproximação da categoria PP, de comportamento intermediariamente ascendente, com a categoria cujo comportamento é inequivocamente ascendente (média PT = 2,57 *z-scores/V-V*, $t(23) = 6,6$, $p < 10^{-5}$). Esta evidência, embora plausível, não é forte o suficiente para explicar sozinha a confusão entre as duas categorias, pois baseia-se num resultado estatístico marginal. Deve-se buscar nos demais parâmetros outras evidências que corroborem a possibilidade de confusão pela semelhança entre os padrões acústicos das duas categorias.

3.1.4 Frequência Fundamental (f_0)

Se a f_0 fosse o parâmetro mais importante empregado pelos ouvintes para reconhecer as combinações experimentais, os resultados deveriam apontar PP como a categoria reconhecida com maior acuidade, pois é a única para a qual existe contraste muito confiável no parâmetro f_0 seguida de PT, combinação para a qual existe diferença marginal entre as duas posições. Em último lugar viria TA, sem contrastes significativos. Tais hipóteses não se confirmam nos resultados realmente observados. É imprudente imaginar que a pouca confusão observada entre as categorias PT e TA e o alto grau de acertos associados a elas se devem a uma diferença marginal entre as posições 1 e 2 em PT e a não diferença entre elas no caso de TA. Além do mais, ficaria difícil explicar o alto grau de confusão verificado para PP se a combinação que mais gera confusão é PT que é aquela que guarda com PP o maior contraste em termos de f_0 . Outro argumento em favor da pouca relevância do parâmetro f_0 para o reconhecimento das categorias vem dos resultados conseguidos para os dados de f_0 transformados em semitons (re 100Hz). O contraste existente entre as posições 1 e 2 nas duas combinações, aproximadamente 2 ST em PP e -2 ST em PT, são potencialmente relevantes do ponto de vista psico-acústico, se considerarmos o limiar menos conservador de 1,5 ST relatado por Nootboom (1999). A diferença, por mais relevante que possa ser do ponto de vista psico-acústico parece não sê-lo para a tarefa lingüística em questão.

3.1.5 A intensidade

Os dados referentes ao parâmetro intensidade não mostram haver contraste algum entre as posições 1 e 2 em cada categoria. Nem mesmo em TA, onde se esperaria que a vogal postônica apresentasse valores comparativamente mais baixos de intensidade relativamente à vogal tônica que a antecede. Somente as posições envolvendo as postônicas (menos a postônica não-final de TA) são diferentes das demais posições. Isto parece ser um indício para uma separação, em termos do parâmetro intensidade, que agrupe juntas pretônicas e tônicas com a possibilidade de alguma gradiência e separe as postônicas.

4. Discussão

O resultado das análises dos parâmetros acústicos encontrados no *corpus* que compõe o experimento ajuda a elucidar de que maneira e em que extensão os

parâmetros acústicos exercem influência sobre o reconhecimento de palavras a partir de fragmentos contendo pistas a respeito da posição do acento lexical.

A hipótese subjacente ao teste é que, para realizar o reconhecimento da palavra “máquina”, por exemplo, a partir do trecho [‘makI] (categoria TA), da palavra “maquina” a partir de [ma’ki] (categoria PT) e da palavra “maquinar” a partir de [maki] (categoria PP), o ouvinte deve de alguma maneira fazer um julgamento de proeminência baseado nas possíveis diferenças acústicas entre as sílabas presentes no fragmento apresentado. Uma segunda hipótese assumida pelo experimento é a da existência, de fato, de diferenças acústicas entre as sílabas que tradicionalmente são classificadas como pretônicas, tônicas e postônicas de acordo com seu valor acentual. Seria esta possibilidade assumida pela última hipótese, segundo as teorias de percepção de fala tradicionais, o que permitiria o reconhecimento das palavras na corrente da fala. Em outras palavras, se há valores característicos dos parâmetros acústicos para cada categoria acentual, basta ao ouvinte perceber estes valores na corrente da fala para reconhecer quais são as palavras presentes ali. Sendo verdadeira esta imagem sobre o que é a percepção da fala, dever-se-ia sempre encontrar na produção distinções claras que permitissem a discriminação e reconhecimento da categoria acentual das sílabas produzidas pelos falantes.

A comparação entre os resultados de acerto/erro no reconhecimento e os resultados acerca da caracterização dos parâmetros acústicos das palavras usadas no experimento mostram que uma teoria da percepção da fala baseada nas hipóteses explicitadas acima teria muita dificuldade em explicar o comportamento dos sujeitos na tarefa de reconhecimento.

Os parâmetros mais tradicionais (duração de vogais e sílabas, f_0 e intensidade), acertos nos casos das categorias TA e PT. Entretanto, falham de maneira patente em explicar o comportamento dos sujeitos em relação à categoria PP. Somente uma medida virtualmente inexistente na literatura tradicional, a derivada da duração das unidades V-V, cujo significado e relevância para a percepção seria dificilmente explicado por esta mesma literatura, pode fornecer uma explicação (mesmo assim fraca) para o padrão não-aleatório dos erros no reconhecimento de PP.

O resultado mais conspícuo revelado pelo padrão de acertos/erros no reconhecimento é aquele que mostra que os falantes, uma vez apresentados à categoria PP, não fazem diferença sistemática entre ela e a categoria PT. Este fato indica que, em certas circunstâncias, uma sílaba pretônica passa por ser uma tônica. No experimento reportado neste trabalho, a circunstância que parece bloquear essa confusão é a combinação em que entra em jogo a sílaba postônica, i. e., a categoria TA. Esta é a categoria com menor incidência de erro de identificação (13,3%). Isto é um indício de que, ao julgar uma sílaba qualquer como tônica, o fato dela ocorrer imediatamente antes de uma sílaba átona, fornece respaldo para sua identificação como tal. Outra circunstância que garantiu a pronta identificação da tônica foram os casos em que a duração aumentou de forma considerável entre a primeira e a segunda posições. No caso do padrão PP, em que não havia contraste com uma postônica, a pequena derivada da duração serviu como indício para sua classificação como tônica.

Em termos gerais, os resultados corroboram uma visão das relações e hierarquias acentuais como aquela expressa por Massini-Cagliari (1999) Albano (2001) e Barbosa (2004). Albano propõe uma distinção entre os padrões de acento lexical marcada pela especificação, na pauta gestual que compõe cada item lexical, de gestos

reduzidos em termos de sua magnitude espacial (que teria sua expressão acústica marcada por vogais de menor intensidade e por um padrão formântico característico) em contraposição aos gestos plenos que caracterizam as vogais plenas que antecedem as possíveis vogais reduzidas (em PB, o padrão é haver até duas). Neste caso, os parâmetros acústicos duração e f_0 estariam a serviço da organização prosódica do enunciado todo (que por sua vez interage com níveis superiores da estrutura lingüística, como a sintaxe, semântica, estrutura do discurso etc.) e não responsáveis por caracterizar uma hierarquia acentual adstrita ao item lexical. No caso do presente experimento, a duração serviu como pista para a tonicidade da sílaba em função da coincidência da culminância do acento frasal com a vogal tônica. Por fim, os dados parecem constituir evidência preliminar para a hipótese de uma atividade de simulação da ação subjacente à percepção da fala. Se o que foi dito no parágrafo anterior for válido, pode-se presumir que, quando os falantes ouvem uma seqüência de sílabas, esperam que ocorram em certos pontos da fala aumentos mais ou menos pronunciados na duração (precedidos de uma zona gradiente de desaceleração) e, logo após este aumento, esperam ocorrer uma redução das durações e também da intensidade e do padrão formântico da vogal que vem em seguida.

Aparentemente, o que os sujeitos do experimento estão fazendo ao julgar a categoria PP é imaginar que a pequena derivada positiva de duração existente no fragmento (que de fato faz parte da zona gradiente anterior à desaceleração máxima, já que contém sílabas pretônicas) é a sinalização de uma culminância. Neste ponto, sem contar mais com nenhum material fônico para julgar, os mecanismos envolvidos na percepção fazem uma simulação e antecipam a ocorrência de uma redução de duração e uma redução vocálica para os próximos instantes. Sem uma confirmação sobre a pertinência ou não da antecipação, o julgamento que é possível fazer sobre o fragmento é que ele deveria conter uma sílaba tônica, uma vez que o que deveria vir em seguida era uma vogal reduzida. Esta é uma interpretação bastante livre dos resultados e requer muitos outros experimentos para ser corroborada de forma mais substancial. Há de se considerar a imensa gama de possibilidades de interpretações que não foram esgotadas no mesmo.

Notas

1 Os dados aqui reportados fazem parte da dissertação de mestrado, em andamento, de Fernanda Consoni, e foram discutidos em conjunto com os demais co-autores em trabalho inter-institucional realizado no IEL-UNICAMP

2 Os resultados aqui apresentados são preliminares e não estão relatados na íntegra, como dito em nota anterior, uma vez que a pesquisa ainda está em andamento.

3 Unidades delimitadas entre dois onsets vocálicos consecutivos. Ver Barbosa e Madureira (1999) para resultados que indicam ser as unidades V-V a unidade de programação rítmica cuja manipulação é responsável, em português brasileiro, pela marcação da acentuação no nível do enunciado.

5. Referências

ALBANO, Eleonora Cavalcante. *O gesto e suas bordas: esboço de fonologia*.

- acústico-articulatória do português brasileiro*. 1ª ed. Mercado de Letras: Campinas, 2001.
- BARBOSA, Plínio A. Elementos para uma tipologia do ritmo (lingüístico) da fala. *InCognito - Cadernos Românicos em Ciências Cognitivas*, 1 (2), 31-58.
- BARBOSA, Plínio A. e MADUREIRA, Sandra (1999). Toward a hierarchical model of rhythm production: evidence from phrase stress domains in Brazilian Portuguese. *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, volume 1, pages 297-300.
- BERTHOZ, Alain. *The Brain sense of movement*. Harvard University Press: Cambridge, 2000.
- BOERSMA, Paul e WEENINK, David (2005). Praat: doing phonetics by computer (Version 4.3.02) [Programa de computador]. Baixado em Fevereiro 27, 2005 de <http://www.praat.org/>
- MASSINI-CAGLIARI, Gladis. *Acento e Ritmo*. Editora Contexto: São Paulo, 1992.
- MORAES, J.A. de (1986) “Acentuação lexical e acentuação frasal em português: um estudo acústico perceptivo”. Comunicação apresentada no II Encontro Nacional de Fonética e Fonologia, Brasília.
- NOOTEBOOM, Sieb. Prosody of speech: melody and rhythm. *In: HARDCASTLE, W e LAVER, J. Handbook of Phonetic Sciences*. Blackwell: London, 1997.