

UM ESTUDO ACÚSTICO DOS DITONGOS CENTRALIZANTES NA FALA CARIOCA

AN ACOUSTIC STUDY OF THE CENTERING DIPHTHONGS
IN THE RIO DE JANEIRO VARIETY OF BRAZILIAN PORTUGUESE

Pablo Arantes | [Lattes](#) | pabloarantes@gmail.com
Universidade Federal de São Carlos

Andrey Nikulin | [Lattes](#) | andre.n.guzman@gmail.com
Universidade de Brasília

Jessé da Silva Lima | [Lattes](#) | jesse.1548@gmail.com
Universidade de Brasília | CNPq

Aveliny Mantovan Lima | [Lattes](#) | avelinylima@gmail.com
Universidade de Brasília

Resumo: O objetivo do trabalho é fazer a caracterização acústica de ocorrências daquilo que chamamos de ditongos centralizantes (DC) na fala carioca. Definimos DC como vogais que, embora subjacentemente sejam monotongos, são percebidas como uma sequência de vogal seguida de uma semivogal com qualidade central. Ocorrências de DC e de monotongos realizados como tal foram buscadas em um *corpus* de entrevistas radiofônicas. As trajetórias formânticas e a duração dos dois tipos de sequências foram comparadas visualmente e quantitativamente por meio da transformada discreta do cosseno. Os resultados indicam que há diferenças marcantes entre os dois tipos de vogais. A trajetória dos DC apresenta movimentos de maior magnitude nas dimensões horizontal e vertical do espaço vocálico, que se iniciam na região mais periférica e terminam na região central, em alturas variadas. Os DC são também mais longos do que os monotongos, quer estejam em posição prosódica proeminente ou não. Os dados de contagem de ocorrências de DC sugerem que há uma preferência pela ocorrência desses segmentos em posição de proeminência prosódica e em momentos de disfluência. Apontamos que há a possibilidade de este viés distribucional evoluir para o estabelecimento de um fenômeno de distribuição complementar entre monotongos e DC. Sugerimos, por fim, com base na Fonologia Articulatória, que a emergência da articulação centralizante resulta da amplificação, em contextos que induzem o alongamento vocálico, da sincronia imperfeita entre os gestos de corpo de língua e glotal.

Palavras-chave: Português brasileiro; Fala carioca; Ditongo centralizante; Fonética acústica.

Abstract: In this study, we offer an acoustic characterization of the occurrences of so-called centering diphthongs (CD) in the Rio de Janeiro variety of Brazilian Portuguese. We define CD as underlying monophthongs that surface as sequences of a vowel followed by a central semivowel (inglide). CD and plain monophthongs were manually segmented from a radio interview corpus, and their formant trajectories and duration were subsequently compared both visually and quantitatively, using the technique of the discrete cosine transform. Our results show that there are striking differences between these two types of vowels. The articulatory gestures involved in CD production are of greater magnitude in horizontal and vertical dimensions of the vowel space; they originate in a peripheral region and are directed towards the central region at varying heights. The duration of the CD is also longer than that of plain monophthongs, regardless of whether the environment is prosodically prominent. A distributional analysis shows that there is a preference for the occurrence of CD in prosodically prominent environments, as well as in the context of disfluencies. We suggest that this distributional bias may contribute to the emergence of a complementary distribution between plain monophthongs and CD. Finally, we suggest, based on Articulatory Phonology, that centering glides in PB can emerge as a result of imperfect synchrony between the tongue body and glottal gestures that gets enhanced in lengthening-inducing environment.

Keywords: Brazilian Portuguese; Rio de Janeiro dialect; Centering diphthongs; Acoustic phonetics.

1. Introdução

Embora facilmente reconhecível pelos falantes do português brasileiro (PB), a variedade da língua utilizada na região metropolitana do Rio de Janeiro (daqui em diante *português carioca*) carece de estudos acústicos de algumas de suas características marcantes. Neste trabalho oferecemos uma descrição fonética preliminar do que vamos chamar de “ditongos centralizantes” (DC), fenômeno que consiste na realização ditongada opcional de algumas vogais, especialmente em posição tônica, que poderiam, preliminarmente, ser transcritas como segue: *amiga* [a' miɣ̞ɣ̞], *porta* ['pɔ̞ɣ̞tɔ̞], *até* [a' tɛ̞ɣ̞]¹. Não se enquadram nessa descrição, e, portanto, não serão tratados aqui, casos bastante comuns nessa e em outras variedades do PB de assimilação regressiva, como a iotização, em palavras como *vez* ['vej̞ʃ]), nos quais a vogal assimila o traço palatal da consoante seguinte.

¹ Ressaltamos que a escolha do [ɣ̞] como símbolo para a transcrição da realização centralizada dessa ditongação reflete apenas uma análise preliminar de oitiva e não há, nesse momento, um compromisso de correspondência com a exatidão dessa vogal no espaço vocálico.

No presente trabalho, com o objetivo de dar uma caracterização fonético-acústica para os DC, analisamos a trajetória das vogais /i, e, ε, a, ə, o, u/ de um *corpus* de fala espontânea, considerando tanto as realizações prototípicas (monotongos tônicos [i, e, ε, a, ə, o, u] e pós tônicos [ɪ, ɐ, ʊ]), como as realizações que apresentaram o fenômeno de DC. Não foram encontrados exemplos de DC para todas as vogais.

1.1. O sistema vocálico do PB

É consenso que o sistema vocálico do PB compreende 7 fonemas (/a, ə, o, u, ε, e, i²) (CÂMARA JR., 1970, p. 33; SEARA et al., 2011, p. 45), sendo que o *status* fonológico é geralmente negado às vogais nasais (CÂMARA JR., 1970, p. 49; CAGLIARI, 1977; LÓPEZ, 1979; WETZELS, 1991, 1997; MORAES; WETZELS, 1992). A estes fonemas correspondem os seguintes fones, encontrados na posição tônica: [a, ə, o, u, ε, e, ɪ, ẽ, ã, õ, ù, ẽ, ĩ].³

O PB permite a ocorrência de múltiplos segmentos vocálicos adjacentes, formando encontros vocálicos, que podem ser ora heterossilábicos (por exemplo, *caô* [ka.'o], *Caim* [ka.'ĩ]), ora tautossilábicos (por exemplo, *mau/mal* [maʊ], *muito* ['mũĩ.tu]). Embora os encontros vocálicos tautossilábicos sejam frequentemente chamados de *ditongos*, não adotaremos este uso no presente trabalho, reservando o termo “ditongo” somente às sequências vocálicas tautossilábicas que não permitem uma análise bissegmental. Como o fenômeno de DC cria segmentos de contorno passíveis de uma análise monosssegmental, neste trabalho optamos por usar o termo “ditongo” para nos referirmos a eles.

1.2. Estudos dos processos de ditongação e monotongação no PB

Muitas variedades regionais e socioletais do PB apresentam fenômenos de ditongação de monotongos ou monotongação de encontros vocálicos. Entre os casos mais bem estudados, principalmente dentro do paradigma variacionista, destacam-se:

- a realização opcional de [ẽ, ã] como [e, a], principalmente diante de [ɾ, ʃ, ʒ], e a realização opcional de [õ] como [o] em todos os ambientes (RECTOR, 1975; VEADO, 1983; MOTA, 1988; BISOL, 1994; CABREIRA, 1996; PAIVA, 1996, 2003; MOLLICA, 1998; ARAÚJO, 1999; LOPES, 2002; SILVA, 2004; PEREIRA, 2004; DUARTE; PAIVA, 2011; SEARA et al., 2011, p. 43-44; LEÃO, 2013; ARAGÃO, 2014);

² Não discutiremos aqui a possibilidade de atribuir o *status* fonêmico ao segmento [ɜ], encontrado em empréstimos recentes e nomes próprios como *hora do rush*, *Ruffles* ou *Pizza Hut* na fala de muitos falantes do PB.

³ [ẽ] é transcrito como [ɜ̃] por alguns autores.

- a inserção do *offglide* [ɪ] antes de /s/ em coda ou diante de fricativas pós-alveolares heterossilábicas (ALBANO, 2001; LEITE et al., 2003; COLLEY, 2009; ALMEIDA, 2009, p. 85, 88; ARAGÃO, 2014; SILVA, 2014; ROCHA et al., 2015);
- a realização de [õ] como [õõ] (ALMEIDA, 2009, p. 96);
- a realização de [ẽ] como [ẽĩ] (OUSHIRO; MENDES, 2014).

No entanto, os processos que envolvem a formação de ditongos decrescentes em que o movimento articulatório do complexo línguo-mandibular segue para uma posição mais baixa não foram suficientemente bem estudados. Além de um estudo sobre a realização opcional de [ẽ] como [ẽõ] por falantes da região metropolitana do Rio de Janeiro (ver COLLEY, 2011), encontramos apenas os estudos referentes à realização ditongada das vogais tônicas orais na variedade do PB falada em Porto Alegre (RS) (ver BATTISTI, 2013; BATTISTI; OLIVEIRA, 2014, 2016).

1.3. Objetivo e estrutura do trabalho

O objetivo buscado será descrever a realização ditongada de vogais da variante carioca do português brasileiro. Para isso, iremos contrapor vogais que se apresentam como DC e outras que são realizadas como monotongos convencionais.

O restante deste trabalho está estruturado da seguinte maneira: na seção 2, discutiremos a tipologia de processos de ditongação em diversas línguas do mundo com um enfoque na criação de ditongos centralizantes; a seção 3 abordará a metodologia utilizada neste trabalho, explicitando os materiais linguísticos, as variáveis definidas e o aparato instrumental empregado para efetuar as análises acústicas e estatísticas; na seção 4, discutiremos os resultados e o panorama que eles indicarem a partir dos dados obtidos; por fim, a seção 5 concluirá o estudo e apontará possibilidades de pesquisas futuras.

2. Ditongação diacrônica: o movimento em direção ao centro

Para um melhor entendimento do processo a ser analisado neste trabalho, é importante levar em consideração que um incremento de duração de uma vogal pode favorecer a criação de segmentos de contorno. De fato, Battisti (2013) já observou que o fenômeno de *ingliding* na fala porto-alegrense é registrado nas vogais com realização alongada. No entanto, Battisti e Oliveira (2014, 2016) argumentam que o *ingliding* pode ocorrer em vo-

gais tanto alongadas quanto não alongadas, sendo que o alongamento facilita a percepção de oitiva do *ingliding*. Segundo os autores, na variedade porto-alegrense do PB o *ingliding* nas vogais não alongadas pode ser detectado apenas via análise acústica.

Contudo, as evidências diacrônicas de outras línguas mostram que os ditongos decrescentes, fonêmicos ou não, frequentemente emergem como um resultado de um processo diacrônico de ditongação (*vowel breaking*) de monotongos alongados (tipicamente se trata de vogais médias). Esse desenvolvimento, quando não causado por processos de *umlaut*, tende a seguir um dos dois caminhos possíveis.

A primeira possibilidade envolve a retenção, ainda que parcial, da região articulatória e do arredondamento dos segmentos originais, conforme ilustrado na Tabela 1 abaixo.

Tabela 1. Criação de ditongos decrescentes não-centralizantes a partir de monotongos

língua	alvo	fonte	referência
finlandês	iɛ̄, uɔ̄, yø̄ ⁴	*e:, *o:, *ø:	ITKONEN, 1983, p. 374; HAKULINEN et al., 2004, p. 55-56
iacute	iɛ̄, uɔ̄, yø̄ ⁵	*e:, *o:, *ø:	TENIŠEV; DYBO, 2006, p. 224-225
latgalês	iɛ̄, uɔ̄	*e:, *o:	NAU, 2011, p. 9; SCHMALSTIEG, 1961
elfdaliano	iɛ̄, uɔ̄, yœ̄ ⁶	*ē, *o:, *ø:	SAPIR, 2005
iucaguir da tundra, iucaguir de Kolymá	iɛ̄, uɔ̄, yø̄ ⁷	*e:, *o:, *ø:	MUDRAK, 2010, p. 167

Fonte: Elaborada pelos autores.

Na história de outras línguas, as vogais médias (geralmente longas) se transformaram em ditongos centralizantes. Nos ditongos desse tipo observa-se um movimento articulatorio para a região central, geralmente acompanhado pela perda do arredondamento, conforme ilustrado na Tabela 2 abaixo.

⁴ Em algumas variedades regionais do finlandês o movimento articulatorio na produção destes três ditongos chegou a fazer um percurso ainda maior, atingindo o grau de abertura mais baixo: [iɛ̄ ~ iǣ / iɔ̄, uɔ̄ ~ uǣ, yø̄] (HAKULINEN et al., 2004, p. 55-56). Contudo, trata-se de um desenvolvimento recente, cujo ponto de partida não foram as vogais médias longas e sim os ditongos decrescentes com um offglide médio.

⁵ Este processo foi acompanhado pela formação do ditongo [iɔ̄] a partir de sequências heterossilábicas.

⁶ Segundo Sapir (2005, p. 20), /uo/ pode ser realizado como [yœ̄] em algumas variedades. Os três ditongos podem ainda ser nasalizados ou alongados.

⁷ Nessas duas línguas existe ainda um ditongo [iɔ̄] de origem incerta.

Tabela 2. Criação de ditongos decrescentes centralizantes a partir de monotongos

língua	alvo	fonte	referência
lituano	i̯, u̯	*e:, *o:	BALODE; HOLVOET, 2001b, p. 46; SCHMALSTIEG, 1961
frísio ocidental	i̯, i̯, u̯, o̯ ⁸	*e:, *ɛ:, *o:, *ɔ:	TIERSMA, 1983, p. 60
luxemburguês	i̯, u̯	*e/*ɛ, *a/*o	GILLES; TROUVAIN, 2013

Fonte: Elaborada pelos autores.

Há ainda uma terceira possibilidade, na qual as duas estratégias de ditongação co-atuam. Nas poucas línguas que apresentam esse tipo de desenvolvimento diacrônico, algumas vogais desenvolvem um *offglide* na mesma região articulatória, enquanto outras criam um *offglide* centralizante. Isso está exemplificado na Tabela 3.

Tabela 3. Criação de ditongos centralizantes e não-centralizantes a partir de monotongos diferentes

língua	alvo	fonte	referência
letão	i̯, u̯	*e:, *o:	BALODE; HOLVOET, 2001a, 11; SCHMALSTIEG, 1961
africâner	i̯, u̯, i̯ ⁹	*e:, *o:, *ø:	LASS, 1987

Fonte: Elaborada pelos autores.

Neste trabalho, procuraremos situar o português carioca nesta classificação, demonstrando que o processo de ditongação nessa variedade do PB cria ditongos centralizantes (como no lituano, no frísio ocidental e no luxemburguês), possivelmente a partir dos alofones alongados por fatores prosódicos ainda não estudados. O processo, ainda em andamento e, portanto, ainda não incorporado aos contrastes fonêmicos da variedade em questão, teria atingido as vogais médias-baixas, médias-altas e altas, mas não a vogal baixa /a/.

⁸ O frísio possui, ainda, os ditongos [y̯, ø̯], de baixa ocorrência. Estes provavelmente provêm de *ɔ:, *œ:, embora este fato não seja mencionado na fonte consultada.

⁹ Lass (1987) registra um alto grau de variação fonética para esses fonemas: [i̯ ~ i̯ ~ i̯], [u̯ ~ u̯ ~ u̯], [ø̯ ~ ø̯].

3. Metodologia

Nesta seção apresentaremos o material linguístico utilizado na presente pesquisa, incluindo os critérios de recorte adotados (3.1), assim como o aparato tecnológico e estatístico (3.2) empregado para a medição, quantificação e análise dos dados selecionados.

3.1. Material de fala

O *corpus* utilizado neste trabalho foi constituído por gravações de fala espontânea, em sua maioria colhidas de *podcasts* gravados pela rádio CBN (disponíveis no site www.cbn.globoradio.globo.com). Essa prática foi adotada de forma bem-sucedida em trabalhos como Kohler (1995) e Ostendorf et al. (2001), tendo a vantagem de permitir a recolha rápida e barata de exemplos de fala bastante diversos em gravações de duração relativamente curta.

Os áudios recolhidos geralmente têm como temática a profissão do sujeito convidado, como, por exemplo, a carreira jornalística, a prática médica ou a produção de energia renovável, além das vivências pessoais de cada indivíduo. Os trechos em que o entrevistador participa da conversa foram descartados para que os possíveis achados não pudessem ser atribuídos ao fonoestilo jornalístico.

As amostras de fala do *corpus* provêm de um total de 13 falantes, sendo oito mulheres e cinco homens, todos naturais da região metropolitana do Rio de Janeiro, de condição social elevada. Em termos de escolaridade, todos apresentam ensino médio ou superior completo. A duração combinada de todas as amostras do *corpus* totaliza 15m55.

3.1.1. Critérios para a seleção dos dados

Como critério para decidir quais dados de fato seriam analisados neste trabalho, optamos por excluir variáveis que poderiam interferir no fenômeno de DC. A fim de evitar a necessidade de uma análise multifatorial, consideramos apenas as vogais que em seu contexto posterior não apresentavam consoantes coronais (ou seja, cuja articulação não envolvia ponta de língua). Com esta delimitação, procuramos minimizar a potencial interação com um fenômeno de ditongação distinto, descrito por autores como Albano (2001, p. 86-91), Leite et al. (2003), Colley (2009), Almeida (2009, p. 85, 88), Aragão (2014), Silva (2014), Rocha et al. (2015), entre outros. Albano (2001) analisa as realizações ditongadas de palavras com o arquifonema /S/ em coda final, como por exemplo “três” e “arroz”, que ela denomina *iotização*, e postula a inserção de [j] diante de /s/.

Segundo a autora, embasada na fonologia articulatória de Browman e Goldstein (1989), a iotização seria motivada não apenas por questões prosódicas, mas também por razões articulatórias:

A lentificação da transição implica uma antecipação da trajetória variável do trato local de constrição da ponta de língua (LCPL) em direção ao alvo. Ao mover-se lentamente, a ponta da língua leva consigo o corpo da língua como articulador passivo. Esse, então, desencadeia a iotização ao passar pelo palato duro numa velocidade apenas um pouco superior à de uma semivogal. (ALBANO, 2001, p. 89)

Dessa forma, escolhemos analisar as vogais seguidas por consoantes bilabiais, labiodentais, velares (incluindo as realizações velares¹⁰ dos róticos do PB), além de considerar também aquelas que precederam pausas. Ademais, decidimos evitar contextos fônicos que pudessem atrapalhar a medição acústica dos formantes, como a nasalidade, em razão das dificuldades relacionadas à identificação automatizada de F_1 e F_2 em trechos afetados por essa característica acústico-articulatória (STEVENS, 1998, p. 303-322).

3.1.2. Rotulagem dos dados

Para um melhor manejo do *corpus*, as vogais e os encontros vocálicos selecionados para a análise foram etiquetados, sendo que os rótulos incluem, obrigatoriamente, a informação sobre as seguintes variáveis:

- sujeito de fala (número de identificação);
- sexo do sujeito de fala (m = feminino, h = masculino);
- tipo de sequência segmental (mt = monotongo, ca = ditongo carioca, dt = encontro vocálico tautossilábico);
- ambiente posterior (ps = pausa, R = rótico velar/pós-velar em coda, rr = rótico velar/pós-velar em ataque, bb = bilabial, lb = labiodental);
- presença da proeminência frasal (s = presente, n = ausente, sh = hesitação);¹¹
- representação das vogais segmentadas (a = /a/ tônica, A = /a/ pós-tônica, O = /ɔ/ tônica, o = /o/ tônica, u = /u/ tônica, U = /u/ pós-tônica, E = /ɛ/ tônica, e = /e/ tônica, i = /i/ tônica, I = /i/ pós-tônica).

¹⁰ De acordo com a análise de oitiva de nossos dados, no português carioca o rótico em coda é realizado predominantemente com uma fricção característica de fricativas pós-velares e não velares (*contra* CALLOU et al., 1998). No entanto, uma análise mais detalhada foge do escopo do presente trabalho.

¹¹ Optamos por distinguir as instâncias de proeminência que ocorrem em trechos fluentes (rótulo s) das ocorrências de hesitação e das marcas discursivas (rótulo sh), associadas com a fala disfluente.

É necessário ressaltar que, dada a inexistência de um critério acústico objetivo para a identificação de um DC (uma aproximação desta definição é um dos objetivos deste trabalho), a classificação dos vocoides selecionados para análise em monotongos ou DC foi realizada de oitiva. Dois dos autores do trabalho fizeram a classificação de forma independente. Nos casos em que houve discordância, os outros dois autores foram consultados a fim de chegar-se a um consenso por meio da audição conjunta. Uma vez que havia apenas dois juízes fazendo a classificação independente e os casos de discordância não passaram de 10%, julgamos que não era o caso de usar um teste estatístico como kappa para quantificar as discrepâncias. O mesmo procedimento foi aplicado em relação à proeminência frasal. O critério estabelecido para classificar um vocoide como DC foi a identificação de uma realização na qual houvesse mudança de qualidade perceptível ao longo de sua articulação. Quanto à proeminência frasal, foram rotuladas como proeminentes (rótulo *s*) as vogais tônicas das palavras que se destacavam dentro de um trecho entre pausas. As vogais que ocorriam em trechos disfluentes (por exemplo, aquelas que faziam parte de marcadores discursivos ou de hesitações) foram rotuladas à parte (rótulo *sh*).

Na Tabela 4, apresentamos o número de ocorrências de cada vogal nos dados selecionados para a análise.

Tabela 4. Número de instâncias analisadas de cada vogal

vogal	posição	DC	monotongo
/i/	tônica	11	22
	pós-tônica	1	2
/e/	tônica	23	6
/ɛ/	tônica	8	17
/u/	tônica	1	6
	pós-tônica	0	4
/o/	tônica	2	11
/ɔ/	tônica	17	7
/a/	tônica	0	21
	pós-tônica	0	1

Fonte: Elaborada pelos autores.

A ausência total das ocorrências de DC na vogal /a/ já foi atestada por Battisti e Oliveira (2014, p. 47; 2016, p. 16). Uma possível explicação desta lacuna distribucional

é que, conforme mostraremos na seção 4, o grau de abertura dos DC aumenta dinamicamente no *offglide*, o que seria impossibilitado no caso da vogal /a/, já que a realização padrão dessa vogal no PB envolve a maior abertura dentre as vogais da língua.

3.2. Análise acústica

Nesta seção detalharemos os procedimentos adotados para a análise acústica dos dados, discutindo a análise da duração (3.2.1), dos valores de formantes (3.2.2), os procedimentos de normalização (3.2.3) e a análise estatística (3.2.4).

3.2.1. Análise de duração

A duração acústica dos segmentos foi determinada a partir dos instantes de início e fim dos segmentos, que foram identificados manualmente e marcados em arquivos de metadados do *Praat* (BOERSMA, WEENINK, 2017) para posterior extração automatizada por meio de um *script* do mesmo programa. A combinação de informações visuais presentes no oscilograma e no espectrograma do sinal acústico foram usados para identificar os instantes de início e fim dos vocoides. O instante de início está associado ao aumento súbito na amplitude do oscilograma e no aparecimento de energia nas frequências altas do espectrograma, indicando a presença clara dos formantes superiores (F_2 e maiores). O instante de fim foi associado à diminuição rápida da amplitude do oscilograma e no esmaecimento da presença dos formantes superiores no espectrograma. A duração bruta dos segmentos passou pelo procedimento de normalização por *z-score* para tornar possível a comparação conjunta de todas as qualidades vocálicas sem a interferência da duração intrínseca. A técnica usada para fazer a normalização da duração é o *z-score* entendido, assim como apresentada detalhadamente em Barbosa (2006). Para a aplicação desse procedimento, recorre-se a uma tabela, apresentada no apêndice B.1 de Barbosa (2006), com valores de média e desvio-padrão para os fones do PB que são usados como referência. Os valores foram calculados a partir de um *corpus* de fala lida, no qual cada segmento do PB apareceu em todos os contextos fônicos admitidos pela fonotaxe da língua. A normalização da duração é feita segundo a fórmula

$$z = \frac{dur - \mu}{\sigma},$$

onde z é o valor normalizado, dur é a duração a ser normalizada, expressa em unidade de tempo, de um segmento e μ e σ são, respectivamente, os valores da média e do desvio-padrão extraídos da tabela de referência para uma vogal do mesmo tipo daquela a ser normalizada.¹² Assim, se o dado a ser normalizado é a duração de uma vogal [i], os valores μ e σ usados na fórmula serão os de uma vogal [i]. No caso das vogais classificadas perceptualmente como DC, os valores de referência usados para fazer a normalização foram os do monotongo correspondente. Esse procedimento foi adotado pelas seguintes razões: (a) os DC não têm, em PB, valor fonológico distinto dos monotongos; (b) comparar os casos classificados como DC com a referência dos monotongos permitirá saber se os DC são caracterizados por um alongamento sistemático relativamente às vogais que não receberam essa classificação; (c) a tabela de referência disponível para os sons do PB lista os ditongos com valor fonológico, que são sempre *offglides*, terminados nas semivogais [i] e [u] e não há razão a priori nem fonética nem fonológica para supor que se possa pôr em equivalência os chamados ditongos verdadeiros do PB com as vogais classificadas aqui como DC. O procedimento de normalização da duração bruta foi automatizado por meio de um *script* para o ambiente de computação estatística R.

A mesma etiquetagem discutida na seção 3.1.2 foi aproveitada para a análise de duração dos segmentos classificados como monotongos ou DC. O interesse no parâmetro de duração está na comparação entre as durações de monotongos e DC e entre os segmentos marcados como proeminentes ou não proeminentes frasalmente, além da interação entre essas duas variáveis.

¹² Um dos pareceristas anônimos sugere que um valor de z -score muito alto poderia indicar que a ocorrência do alongamento indicado por esse valor poderia ser considerada improvável caso se interpretasse o valor de z como uma probabilidade. Embora a transformação de dados brutos em z -scores tenha origem na estatística, não nos parece que Barbosa (2006) sugira que os valores de duração normalizados através dessa transformação devam ser interpretados como probabilidades, mesmo porque, para isso, média e desvio-padrão de pausa silenciosa deveriam entrar para o cálculo de um z -score estatístico. Esses valores dados pela normalização são um índice descritivo de alongamento ou encurtamento de uma determinada unidade silábica relativamente a descritores de tendência central e dispersão típicos dos fones que a compõem. Exemplificando para o caso de fones, consideramos que um fone com valor z de 1,5 sofreu um alongamento menor do que outro com valor z de 3, mas não derivamos daí que o primeiro alongamento seja estatisticamente mais provável do que o segundo. Interpretar os valores de z -score como probabilidades no contexto da normalização de duração assim como proposta por Barbosa (2006) introduziria também a questão da distribuição de probabilidade a ser usada como referência para julgar um valor empírico de z . O uso de valores z para fazer inferências supõe uma distribuição normal, suposição que não é necessariamente válida. Barbosa (2006) e Rosen (2005) mostram que durações segmentais de vogais e consoantes são melhor modeladas por distribuições não-normais, como lognormal e gama.

3.2.2. Extração dos valores de formantes

Para os formantes, o intervalo de análise foi o mesmo definido para a duração. O *script* FormantPro (XU, GAO, 2018) foi usado para extrair a trajetória normalizada temporalmente dos valores de F_1 , F_2 e F_3 das vogais do *corpus*. Especificamos a coleta de 20 amostras igualmente separadas no tempo a serem extraídas em cada trecho. Para as falantes femininas, utilizamos os valores padrão sugeridos para os parâmetros da técnica LPC, usada pelo *script* para a extração dos formantes com frequência máxima de 5,5 kHz e um máximo de 5 formantes. Para os falantes masculinos, fixamos a frequência máxima em 5 kHz e um máximo de 5 formantes. Verificamos posteriormente que, para algumas das ocorrências, especialmente de /i e u/, essas especificações produziram extrações errôneas. Nesses casos, aplicamos o procedimento de variar separadamente os valores dos dois parâmetros relevantes de extração até que fosse possível conseguir uma extração aceitável. Nenhum dado foi descartado por ter sido impossível conseguir uma extração adequada.

3.2.3. Procedimento de normalização

Como o *corpus* é formado por amostras de fala de múltiplos falantes, tanto do sexo feminino quanto masculino, foi necessário recorrer a uma técnica de normalização dos valores dos formantes para que fosse possível analisar conjuntamente todos os dados. Em função das características do *corpus*, não foi possível obter para todos os falantes exemplos de todas as vogais do inventário da língua. Por conta disso não foi possível aplicar as técnicas de normalização do tipo *vowel-extrinsic* mais conhecidas, como as de Lobanov (1971) e Nearey (1978), por exemplo. Por conta dessa limitação, recorreremos a uma técnica de tipo *vowel-intrinsic*, que pode ser aplicada a partir de informações obtidas para cada ocorrência de vogal no *corpus*. Recorreremos a um procedimento proposto por Syrdal e Gopal (1986), que consiste em representar a dimensão do avanço da língua no eixo anterior-posterior e a dimensão da altura por meio de diferenças entre valores de formantes convertidos da escala física Hertz para a escala perceptual Bark. Seguindo sugestão de Thomas (2011, p. 163), que modifica a proposta original de Syrdal e Gopal, a dimensão do avanço é dada pela diferença $Z_3 - Z_2$ e a da altura por $Z_3 - Z_1$, em que Z_i é o valor do i -ésimo formante convertido da escala Hz para a escala Bark. Para a conversão, empregamos a fórmula

$$Z_i = \frac{26,81}{1 + \frac{1960}{F_i} - 0,53},$$

apresentada em Thomas (2011, p. 58). O ambiente de computação estatística R (R CORE TEAM, 2017) foi usado para a conversão dos valores e demais operações necessárias para a análise das trajetórias formânticas.

A extração de 20 amostras ao longo de cada trecho analisado permite que avaliemos a evolução da trajetória formântica no espaço articulatório ao longo do tempo. No caso dos monotongos, esperamos que a trajetória seja basicamente estacionária, com pouco grau de movimentação. Nos casos identificados na análise de oitiva como ocorrências de ditongos centralizantes, esperamos encontrar evidência de um grau comparativamente maior de movimentação. A hipótese que levantamos é que, nesses casos, a trajetória tem início na região articulatória próxima àquela ocupada pelo monotongo de mesma qualidade vocálica e gradualmente se move em direção à região central do espaço vocálico.

A fim de testar essa hipótese, geramos a trajetória formântica média para os monotongos e para os casos de DC separadamente (vide Tabela 4 para as quantidades de ocorrências de cada qualidade vocálica no nosso *corpus*). Para evitar os efeitos mais pronunciados da coarticulação com os segmentos circundantes, desprezamos os dois primeiros e os dois últimos valores da série de vinte gerados pela aplicação do *script* FormantPro.

3.2.4. Análise quantitativa das trajetórias

A observação do traçado das trajetórias dos monotongos e DC no espaço vocálico, apresentados na seção 4, permite uma análise visual e qualitativa das possíveis diferenças entre ambos os tipos. Para permitir uma quantificação das diferenças entre os dois tipos que complementasse a análise qualitativa, recorreremos à técnica da transformada discreta do cosseno (*discrete cosine transform*, DCT, em inglês), da maneira sugerida por Watson e Harrington (1999). Conforme explicam os autores, a técnica permite modelar a trajetória temporal de um formante reduzindo-a a coeficientes da DCT.¹³ Esses coeficientes têm valores que são proporcionais à presença de diferentes componentes derivados de uma função-base, nesse caso a função cosseno. O primeiro coeficiente representa uma linha reta e seu valor é proporcional à média dos valores da trajetória. O segundo coeficiente quantifica a presença de um componente que é um meio ciclo da função cosseno. No presente contexto, ele pode ser entendido com uma medida da direção e da magnitude da

¹³ Um parecerista observa que existem outras técnicas que poderiam ser utilizadas com a mesma finalidade (isto é, a decomposição da trajetória formântica em componentes elementares), tais como os polinômios ortogonais (LEVITI; NEUMAN, 1991; GRABE et al., 2007) ou a análise funcional de dados (ASTON et al., 2010), entre outras. A escolha da DCT para o presente estudo não foi precedida por uma comparação sistemática de todas as técnicas disponíveis, mas pelo fato dela ser apropriada para os propósitos do trabalho.

inclinação da trajetória do formante. Assim, se uma hipotética trajetória de F_1 de um monotongo é quase estacionária, o primeiro coeficiente terá um valor proporcional ao valor médio de F_1 e o segundo coeficiente terá um valor próximo a zero. Em comparação, se a trajetória tem uma inclinação pronunciada, que se poderia esperar no caso de um ditongo como [aĩ] em português brasileiro, o segundo coeficiente terá um valor distante de zero, que pode ser positivo ou negativo, dependendo da direção do movimento. Outros coeficientes quantificam a presença de componentes mais complexos da função-base, como um ciclo completo do cosseno, por exemplo. Seguindo Watson e Harrington (1999), limitamos a análise aos dois primeiros coeficientes de cada formante, uma vez que o estudo dos autores mostrou que eles são suficientes para caracterizar a dinâmica da trajetória formântica.¹⁴ Em nosso caso, aplicamos a DCT aos valores das dimensões avanço e altura das trajetórias médias tanto dos monotongos quanto dos DC, gerando no processo dois valores de coeficiente para cada dimensão. Usamos a função *dct* da biblioteca *emuR* desenvolvida para o ambiente de computação estatística R para realizar essa análise.

3.3. Análise estatística

Para os testes de hipótese envolvendo a variável dependente duração, o teste de Fligner-Killeen (FLIGNER; KILLEEN, 1976) foi aplicado para verificar o pressuposto de homocedasticidade antes da aplicação do teste de comparação de médias. Testes paramétricos foram usados nos casos em que a amostra é homocedástica: teste-*t* no caso de variáveis independentes com apenas dois níveis e análise de variância quando há mais do que dois. Testes não-paramétricos foram usados quando o requisito da homocedasticidade não se verificou: teste de Wilcoxon em lugar do teste-*t* e teste de Kruskal-Wallis em lugar da análise de variância. Quando, nos casos em que a variável independente apresenta mais do que dois níveis, foi preciso fazer comparações múltiplas, recorreu-se a testes-*t* ou Wilcoxon pareados e os valores de *p* foram corrigidos usando a técnica de Benjamini e Hochberg (1995), que controla a quantidade de hipóteses nulas erroneamente descartadas, a chamada *false discovery rate*. Adotamos um nível de 5% para a rejeição da hipótese nula em todos os testes. Usamos o ambiente de computação estatística R em sua versão mais recente no momento da escrita do trabalho para realizar as análises estatísticas, gerar os gráficos e sumários de estatística descritiva.

¹⁴ Em nossos dados, os dois primeiros coeficientes de cada formante foram suficientes para distinguir as trajetórias formânticas dos dois tipos de segmentos analisados, tornando dispensável a extração dos coeficientes superiores ao segundo para os nossos fins.

4. Resultados e discussão

As Figuras 1 e 2 mostram, respectivamente, a trajetória formântica média das vogais cuja realização foi classificada auditivamente como DC (Figura 1) e monotongos (Figura 2). A comparação das duas figuras mostra que há grandes diferenças visuais entre as trajetórias das vogais dos dois tipos. As realizações monotongadas não apresentam trajetórias estritamente estáveis (se isso acontecesse, todos os pontos estariam sobrepostos e não haveria propriamente uma trajetória), porém os movimentos têm amplitude menor quando comparados aos observados nas versões ditongadas. Mais adiante quantificaremos essas diferenças usando a técnica da transformada discreta do cosseno.

Figura 1. Trajetória formântica das instâncias classificadas como DC no espaço definido pelas dimensões do avanço e da altura. O triângulo indica o ponto de início e o quadrado o ponto final de cada trajetória. Os símbolos que representam as vogais estão localizados no ponto mediano da trajetória das versões monotongadas para servir como referência. O símbolo “x” indica o centro do polígono vocálico.

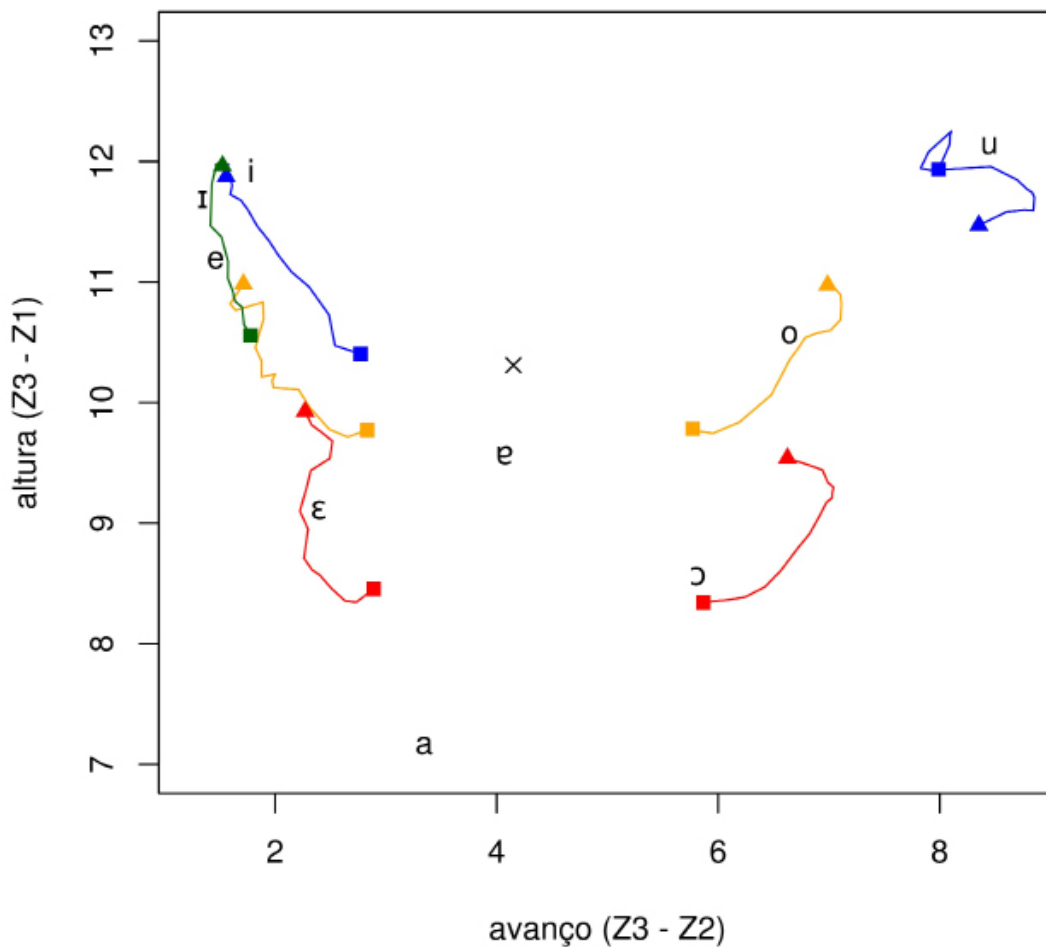
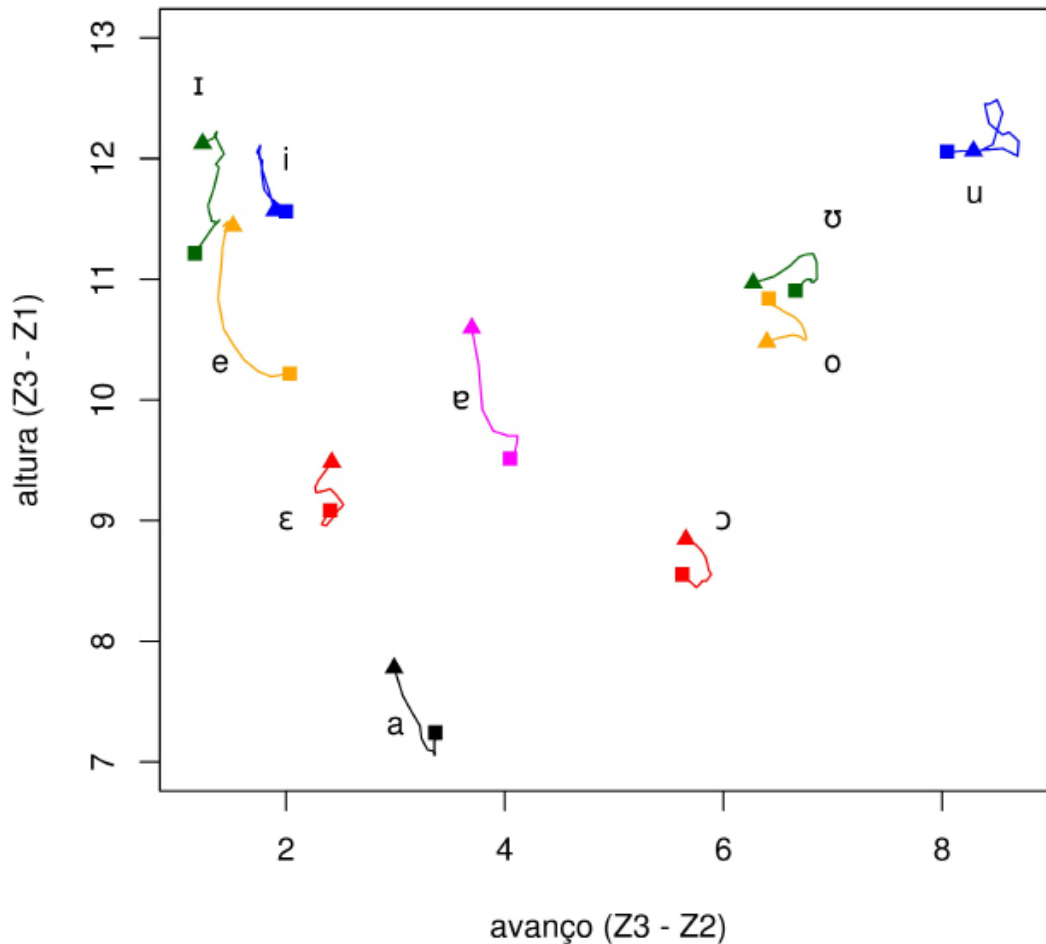


Figura 2: Trajetória formântica das instâncias classificadas como monotongos no espaço definido pelas dimensões do avanço e da altura. O triângulo indica o ponto de início e o quadrado o ponto final de cada trajetória.



A observação, na Figura 1, das trajetórias das vogais classificadas como realizações ditongadas sugere que a transcrição dessas instâncias como ditongos centralizantes é correta, isto é, essas vogais parecem ter um segundo alvo articulatorio na região central do espaço vocálico (ver seção 2).¹⁵ A figura mostra que o formato da trajetória de todas as vogais é compatível com a percepção de um ditongo, dado o movimento de grande magnitude nas duas dimensões, avanço e altura. Em todos os casos, com exceção do /u/, observa-se algum grau de abaixamento e centralização entre o início e o fim da trajetória.¹⁶

¹⁵ Pode-se debater a escolha do símbolo fonético mais apropriado (inclusive se ele deve ser o mesmo em todos os casos), mas o fato de as trajetórias convergirem para o centro do espaço vocálico está fora de dúvida.

¹⁶ Como apontado na seção 3.2, a extração dos formantes de algumas instâncias da vogal /u/ apresentou dificuldades, em função da proximidade entre F_1 e F_2 .

Podemos observar na Figura 1 que o ponto final da trajetória das vogais anteriores /ε, e, i/ na dimensão do avanço é praticamente o mesmo. Entre as vogais posteriores, o ponto final da trajetória das vogais /ɔ, o/ também é praticamente o mesmo na dimensão do avanço, indicando graus de centralização parecidos para essas vogais. Nota-se, também, que as trajetórias não convergem para um mesmo ponto na região central, elas terminam em patamares espaçados que mais ou menos espelham a distância que separa as diferentes vogais no início das trajetórias. Essa é uma diferença marcante entre os presentes resultados e aqueles reportados por Battisti e Oliveira (2014, p. 16) para a variedade de Porto Alegre, que sugerem uma convergência da trajetória de todas as realizações centralizantes para uma região próxima à da vogal representada por [ɐ], muito embora a trajetória representada seja esquemática, não gerada a partir de valores formânticos reais.

4.1. Análise quantitativa das trajetórias formânticas

A Tabela 5 mostra os valores dos dois primeiros coeficientes da transformada discreta do cosseno estimados a partir da trajetória da dimensão da altura (equivalente ao primeiro formante) e do avanço (equivalente ao segundo formante) dos dois tipos de sequência (realização ditongada e monotongada) de sete vogais.

O exame comparativo dos valores do primeiro coeficiente dos dois tipos de vogais mostra que grande parte das diferenças entre os valores do primeiro coeficiente é menor do que 5%. Isso indica que, de forma geral, como seria possível esperar, os dois tipos de vogais ocupam lugares semelhantes no espaço vocálico. A análise do segundo coeficiente, que quantifica a presença de inclinação na trajetória dos dois primeiros formantes, mostra que os DC apresentam valores mais elevados quando comparados às vogais com realização monotongada. Os coeficientes das vogais do primeiro tipo são, em média, 6,3 vezes maiores na dimensão do avanço e 3,7 vezes maiores na dimensão da altura, indicando maior movimentação na dimensão do avanço no caso das realizações com DC. Esses resultados permitem dizer que o grau de movimentação da trajetória é o principal fator que diferencia os dois tipos de realização. As versões monotongadas das vogais [e i] apresentam valores para o segundo coeficiente que não são tão menores do que os valores das versões ditongadas. Um segundo exame auditivo, feito após a análise quantitativa, confirmou a classificação dessas ocorrências como realizações monotongadas e um exame visual da trajetória dos formantes sugere que o movimento na dimensão da altura que essas vogais apresentam é resultado do efeito coarticulatório com a consoante seguinte,

que frequentemente¹⁷ tem um ponto de articulação velar (ou, mais precisamente, pós-velar; ver nota 10).

Tabela 5. Valor do 1º e 2º coeficientes da DCT das trajetórias formânticas das vogais realizadas como ditongos centralizantes ou monotongos

vogal	tipo	dimensão	coeficientes	
			1º	2º
ɛ	ditongo centralizante	avanço	3,44	0,14
		altura	12,82	0,78
	monotongo	avanço	3,37	-0,03
		altura	12,95	0,17
e	ditongo centralizante	avanço	2,90	-0,44
		altura	14,58	0,55
	monotongo	avanço	2,18	-0,16
		altura	15,49	0,70
i	ditongo centralizante	avanço	2,74	-0,51
		altura	16,09	0,68
	monotongo	avanço	2,56	-0,04
		altura	16,79	0,10
I	ditongo centralizante	avanço	2,22	-0,13
		altura	16,03	0,73
	monotongo	avanço	1,88	0,02
		altura	16,63	0,44
ɔ	ditongo centralizante	avanço	9,46	0,38
		altura	12,70	0,58
	monotongo	avanço	9,36	0,06
		altura	15,01	-0,16
O	ditongo centralizante	avanço	9,43	0,59
		altura	14,76	0,61
	monotongo	avanço	9,36	0,06
		altura	15,01	-0,16
u	ditongo centralizante	avanço	11,86	0,42
		altura	16,77	-0,31
	monotongo	avanço	11,75	0,19
		altura	17,25	0,00

¹⁷ Nos nossos dados, esta situação se deu em 8 ocorrências (6 no caso de /e/ e 2 no caso de /i/ pós-tônico).

4.2. Análise da duração das vogais

A Tabela 6 mostra a duração normalizada média das realizações vocálicas classificadas como ditongação carioca e monotongação em posição proeminente e não proeminente no nível frasal. Os resultados estatísticos são apresentados a seguir.

Tabela 6. Duração normalizada (e desvio-padrão) das vogais, expressa em z-scores, em função do tipo de realização do segmento (segunda coluna) e do cruzamento de seu grau de proeminência e do tipo de realização (quinta coluna). A coluna “N” indica o número de instâncias de cada tipo analisadas.

tipo	duração	proeminência	N	duração
ditongo centralizante	3,6 (3,8)	não	17	2,1 (3,7)
		sim	21	4,8 (3,4)
monotongo	0,1 (3,3)	não	65	-0,7 (2,4)
		sim	21	2,7 (4,2)

Os dados da segunda coluna da Tabela 6 mostram que a média de duração normalizada das vogais percebidas como ditongos centralizantes é maior do que a dos monotongos. A realização centralizante apresenta também um nível de variabilidade um pouco mais elevado em comparação à realização monotongada. O efeito de alongamento provocado pelo movimento articulatório centralizante é significativo do ponto de vista estatístico tanto no caso da média [$\chi^2(1) = 31 p < 0,001$] quanto da variância [$\chi^2(1) = 4 p < 0,05$]. O fato da vogal estar ou não em posição de proeminência de nível frasal pode ter um papel na duração da vogal, independentemente de sua realização monotongada ou ditongada por conta do efeito de alongamento em posição de proeminência frasal. No presente *corpus*, as vogais em posição de proeminência têm média de duração normalizada de 3,8 (DP 3,9) e as que não estão em posição proeminente têm média de -0,1 (DP 2,9). Os números mostram que as vogais em posição proeminente sofrem mais alongamento e são também mais variáveis. Esse efeito é significativo do ponto de vista estatístico para as médias [$\chi^2(1) = 31 p < 0,001$] e para a variância [$\chi^2(1) = 14 p < 0,001$]. Os resultados indicam que as duas variáveis independentes, tipo de realização das vogais e proeminência, têm efeito sobre a duração dos vocoides, quer sua realização seja monotongada ou centralizante. É importante, agora, verificar como se dá a interação entre elas, isto é, se o alongamento devido ao tipo de realização e a proeminência da vogal se compõem ou se anulam. A quinta coluna da Tabela 6 sugere que os efeitos se reforçam, pois, entre as vogais em posição proeminente frasalmente, aquelas realizadas como DC são mais longas do que os monotongos (DC

4,8 vs. monot. 2,7; $p < 0,04$). O maior alongamento das vogais realizadas como DC em comparação com os monotongos também se mantém nas posições não proeminentes frasalmente (DC 2,1 vs. monot. -0,7; $p < 0,001$). Crucialmente, o maior alongamento dos DC em comparação aos monotongos não contradiz o efeito de maior alongamento devido à proeminência prosódica, que se mantém quando são considerados os dois tipos de realização das vogais. Entre os DC, a média dos proeminentes é 4,8 e a dos não proeminentes 2,1 ($p < 0,01$). Entre os monotongos, a média dos proeminentes é 2,7 e a dos não proeminentes é -0,7 ($p < 0,001$).

A Tabela 7 mostra o número de ocorrências, a duração acústica e a duração normalizada média das sete vogais para as quais há ocorrências dos dois tipos de sequências (realizações monotongadas e ditongadas). O exame visual mostra que, independentemente da qualidade da vogal, aquelas classificadas como DC tem duração mais longa do que a versão percebida como monotongo.

Tabela 7. Duração bruta (em milissegundos) e normalizada (z-scores) média das diferentes vogais em função do tipo de realização

		ε	e	i	ɪ	ɔ	o	u
ditongo centralizante	N	8	6	7	1	13	2	1
	dur (ms)	194	180	147	260	153	156	167
	dur (z)	4,8	3,7	3,2	14,6	2,5	2,2	4,6
monotongo	N	14	8	19	1	6	10	4
	dur (ms)	119	148	87	145	79	109	96
	dur (z)	0,4	2	0	6,4	-2,2	-0,1	-0,5

Como dissemos na seção 3.1.2, tomamos a decisão de analisar separadamente as vogais que ocorreram em trechos de fala fluente e disfluente. No *corpus*, as vogais a serem analisadas que ocorreram em trechos disfluente correspondem a ocorrências de hesitação, condição em que as vogais frequentemente sofreram alongamento, e de marcadores conversacionais, especialmente a partícula “né”. Segundo a observação de um parecerista, é possível questionar se a normalização da duração das vogais nos trechos de fala disfluente poderia ser feita usando valores de referência que foram obtidos a partir de fala fluente, pois se trata de regimes de fala diferentes. Em razão de nosso desconhecimento de fontes publicadas nas quais se encontrem valores de referência da duração de fones do PB na fala disfluente, usamos os da fala fluente. Para que fiquem claros os possíveis vieses que a normalização pode introduzir neste caso, reportaremos os resultados da estatística

descritiva e dos testes de hipótese tanto para os valores normalizados quanto brutos.

Na amostra analisada neste estudo, há 22 ocorrências de vogais em trechos de disfluência. Destas, 14 foram classificadas como realizações de DC e sua duração bruta média é de 333 ms (DP 209) e normalizada média de 13,3 z-scores (DP 12,3), e oito como realizações monotongadas, com duração bruta média de 315 ms (DP 195) e normalizada média de 11 z-scores (DP 9,9). A comparação das médias indica que o alongamento das vogais em trechos disfluentes é maior e mais variável do que nos trechos fluentes. O teste de homogeneidade de variância indica que o grau de variabilidade dos DC e dos monotongos não é significativamente diferente, quer os dados de duração sejam brutos [$\chi^2(1) = 0,03$ ns] ou normalizados [$\chi^2(1) = 0,04$ ns]. Testes-*t* de duas amostras indicam que a média de duração dos DC e dos monotongos não é estatisticamente diferente quer os dados considerados sejam brutos [$t(20) = 0,02$ ns] ou normalizados [$t(20) = 0,4$ ns].

Uma observação adicional que podemos fazer sobre os dados diz respeito ao fato de que a distribuição de ocorrências de DC e de monotongos não é uniforme nos diferentes contextos. Começando pelos trechos de fala disfluente, nos quais as vogais tendem a sofrer alongamentos consideráveis, das 22 ocorrências quase 65% das vogais foram percebidas como DC. Nos trechos de fala fluente, 66% das 124 ocorrências correspondem a vogais em posições não proeminentes no nível frasal. Nesse conjunto, apenas 26% das vogais foram percebidas como DC. Nas posições percebidas como proeminentes ao nível frasal, grupo que perfaz os 33% restantes do total da amostra, 50% das vogais foram percebidas como DC. Esses dados podem ser interpretados como indicação de que os contextos que estão associados a um alongamento maior da vogal (sílabas proeminentes no nível frasal e trechos de disfluência) são mais propícios para a realização de vogais percebidas como DC.

5. Considerações finais

Neste trabalho, procuramos contribuir para a caracterização acústica do que chamamos de ditongação centralizante na fala carioca, que é um fenômeno característico de uma das variedades mais reconhecíveis do PB e que, até o presente, foi alvo de um número insuficiente de pesquisas fonéticas e fonológicas. Uma novidade introduzida pelo trabalho é a análise da trajetória formântica das sequências de interesse por meio da técnica da normalização temporal da trajetória formântica. A técnica da normalização temporal permite tanto uma análise qualitativa (visual) quanto quantitativa da trajetória dos dois formantes no espaço vocálico, resultando em uma análise mais detalhada do fenômeno.

Os trabalhos anteriores sobre o tema do *ingliding* no PB (BATTISTI; OLIVEIRA, 2014, 2016) analisaram, além da duração, apenas a variação de F_2 , medida em dois pontos: no início e no fim da vogal.

Os dados analisados aqui demonstram que instâncias vocálicas percebidas como DC podem, de fato, ser descritas como a realização de uma vogal com uma finalização articulada na região central do espaço articulatório com variação na altura. As evidências para essa afirmação vêm da observação comparativa das trajetórias de DC e monotongos no espaço vocálico definido pelas dimensões de avanço e altura por um processo de normalização dos formantes e da análise dos coeficientes gerados pela aplicação da transformada discreta do cosseno. Essas duas análises, em conjunto, permitem afirmar que, nos casos de vogais percebidas com DC, as trajetórias apresentam mais movimentação do que aquelas percebidas como monotongos. A maior duração das vogais percebidas como DC também pode ser interpretada como corroboração para sua caracterização como um *inglide*, já que o movimento da língua em direção ao centro do espaço vocálico toma um tempo adicional.

Uma segunda questão importante a respeito do fenômeno dos DC é o seu estatuto linguístico. Não há dúvida de que, no presente momento da variedade carioca, os falantes não fazem uso distintivo dos DC, isto é, a oposição monotongo-DC não é produtiva, ou seja, a comutação de um segmento por outro não gera contraste entre itens da língua. A questão, portanto, é se estamos diante de uma variação livre típica ou se é possível apontar evidências para uma distribuição complementar. Battisti e Oliveira (2016), no seu estudo sobre a variedade falada em Porto Alegre, sugerem que, naquela região metropolitana, o aparecimento de um ou outro alofone é interpretado como índice de estilo, revestido de significados sociais próprios daquela comunidade falante. Não encontramos na literatura nenhuma pesquisa análoga à dos autores mencionados voltada para a variedade carioca. Fica a cargo de trabalhos futuros, que façam uso de metodologia semelhante à empregada por Battisti e Oliveira (op. cit.), levantar hipóteses semelhantes no caso da variedade carioca. Os resultados, por outro lado, também não podem concluir de forma definitiva a respeito do caráter complementar da alofonia entre monotongos e DC. Há indícios, apresentados na seção 4.2, que apontam, malgrado a limitação imposta pelo tamanho relativamente reduzido do *corpus*, para uma preferência dos DC por contextos nos quais se pode esperar o alongamento de segmentos e sílabas, como a posição de proeminência ao nível da frase e trechos de disfluência. Essa preferência, que não é categórica no presente momento, pode vir a ser reinterpretada pelos falantes, ou por um conjunto deles, em algum

momento no futuro, como uma relação necessária e constituir a base para uma regra de distribuição complementar entre alofones monotongados e DC. Os dados apresentados na seção 2 ilustram como outras línguas fizeram o processo de incorporação de ditongos centralizantes em sua fonologia a partir de vogais longas ou alongadas.

De todo modo, nos parece importante esboçar uma explicação para o fato de que dois contextos nos quais vogais sofrem com frequência alongamentos (proeminência prosódica ao nível da frase e hesitações/marcadores discursivos) possam gerar uma variação fonética: um alongamento simples, sem implicações para a articulação da vogal, e um alongamento que vem acompanhado de mudanças articulatórias, como a realização da articulação centralizante documentada no presente trabalho, que é inusual em PB¹⁸ e não é usada nessa língua com valor distintivo no nível lexical.

Para o esboço de explicação, assumimos o modelo da Fonologia Articulatória (BROWMAN; GOLDSTEIN, 1986, 1989) com as modificações propostas por Tatham (1995), que incluem um mecanismo de supervisão do processo de produção da fala, que simplifica alguns aspectos das pautas gestuais propostas no âmbito do modelo ao dar conta de detalhes de implementação fonética que o autor classifica como não automáticos e que podem ser explorados de maneira sistemática pelas línguas (para alguns exemplos, vide TATHAM, 1979).

No caso que analisamos aqui, o processo de supervisão deveria garantir que aquilo que foi especificado na pauta gestual como uma vogal simples, um monotongo, seja assim realizado. Para isso, a pauta deve especificar para a variável do trato ‘corpo de língua’ (TB, de *tongue body* em inglês) um lugar e um grau de constrição apropriados para atingir o timbre vocálico intencionado pelo falante. Além disso, deve haver coincidência entre o gesto de TB e o gesto glotal (GLO), com especificação *fechada*, responsável pelo vozeamento. O mecanismo de supervisão fica responsável por garantir que, durante o processo de transformação dessa pauta gestual em especificações de trajetória (operação que é feita pelo componente de dinâmica da tarefa), os gestos de TB e GLO sejam ativados em simultaneidade e durem o tempo determinado pela interação entre a pauta gestual do item lexical e o modelo de duração prosódica (para essa relação, ver BARBOSA, 2006).

Nossa hipótese para explicar o surgimento da articulação centralizante é que essa sequência fônica emerge quando o gesto TB é desativado fora de sincronia relativamente ao gesto GLO por conta de uma falha momentânea no mecanismo de supervisão. A

¹⁸ Contudo, em uma perspectiva tipológica, o fenômeno de formação de ditongos centralizantes constitui uma mudança fonética natural (alguns precedentes foram examinados na seção 2 referentes a línguas que não o PB).

desativação de TB resulta em instruções motoras para o retorno do corpo de língua para a posição articulatória neutra (na qual o dorso da língua está próxima ao centro do espaço vocálico) enquanto o gesto GLO, ainda ativado e em posição *fechada*, produz voz. Essa conformação gera o efeito acústico de *inglide*, observado nas trajetórias da Figura 1. Quanto mais alongada a vogal estiver, mais longa será a trajetória em direção ao centro gerada pela assincronia na desativação dos dois gestos. Voltando aos dados apresentados ao final da seção 4.2, vemos que a maior prevalência relativa da articulação centralizante acontece no contexto de fala disfluente, que é aquele associado aos maiores índices de alongamento, seguido do contexto em que a vogal está em posição proeminente no nível frasal, este também associado a níveis mais elevados de alongamento.

Se aceitarmos a hipótese delineada nos parágrafos acima, devemos concluir que a articulação centralizante que documentamos neste trabalho na verdade não configura um ditongo no sentido de ser uma combinação de vogal e semivogal. O que teríamos, nesse caso, é a especificação de uma vogal simples que é realizada com um *inglide* por conta da desativação do gesto de corpo de língua fora de sincronia com o gesto glotal.

Abreviações

CBN	Central Brasileira de Notícias	ms	milissegundo
DC	ditongos centralizantes	<i>N</i>	quantidade
DCT	transformada discreta do cosseno	PB	português brasileiro
DP	desvio padrão	RS	Rio Grande do Sul
F _{<i>i</i>}	<i>i</i> -ésimo formante	TB	corpo de língua
GLO	glotal (<i>gesto</i>)		
LCPL	local de constrição da ponta de língua		

Referências

- ALBANO, E. C. *O gesto e suas bordas: esboço de fonologia acústico-articulatória do português brasileiro*. Campinas: Mercado de Letras, 2001.
- ALMEIDA, M. M. S. *Vogais do falar ribeirinho cuiabano*. 2009. 141 f. Tese (Livre-Docência em Filologia e Língua Portuguesa) — Departamento de Letras Clássicas e Vernáculas da Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2009.
- ARAGÃO, M. do S. S. de. Ditongação e monotongação nas capitais brasileiras. In: HORA, D.; PEDROSA, J. L. R.; LUCENA, R. M. de (Org.). *Anais da ALFAL 2014: XVII Congresso Internacional Asociación de Lingüística y Filología de América Latina (ALFAL)*. João Pessoa: Ideia, 2014. p. 2089-2101.

- ARAÚJO, M. F. R. de. A monotongação do ditongo decrescente [ẽ] no português caxiense. *Revista Estudos Lingüísticos*, Belo Horizonte, v. 8, n. 2, p. 23-51, jul./dez. 1999.
- ASTON, J. A. D.; CHIOU, J. M.; EVANS, J. P. Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Oxford, v. 59, n. 2, p. 297-317, 2010.
- BALODE, L.; HOLVOET, A. The Latvian language and its dialects. In: DAHL, Ö.; KOPTJEVSKAJA-TAMM, M. (Ed.). *Circum-Baltic Languages. Volume I: Past and present*. Amsterdam; Philadelphia: John Benjamins, 2001a. p. 3-40.
- _____. The Lithuanian language and its dialects. In: DAHL, Ö.; KOPTJEVSKAJA-TAMM, M. (Ed.). *Circum-Baltic Languages. Volume I: Past and present*. Amsterdam; Philadelphia: John Benjamins, 2001b. p. 41-79.
- BARBOSA, P. A. *Incursões em torno do ritmo da fala*. Campinas: Pontes, 2006.
- BATTISTI, E. Realizações variáveis de vogais tônicas em Porto Alegre (RS): Ditongação ou ingliding? *Fragmentum*, Santa Maria, v. 39, p. 60-78, 2013.
- BATTISTI, E.; OLIVEIRA, S. G. de. Alongamento e *ingliding* de vogais em sílabas tônicas no português falado em Porto Alegre (RS). *Revista (Con)Textos Lingüísticos*, Vitória, v. 8, n. 11, p. 39-56, 2014.
- _____. Significados sociais do *ingliding* de vogais tônicas no português falado em Porto Alegre (RS). *Todas as Letras*, São Paulo, v. 18, n. 2, p. 14-29, maio/ago. 2016.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, Oxford, v. 57, n. 1, p. 289-300, 1995.
- BISOL, L. Ditongos derivados. *D.E.L.T.A.*, São Paulo, v. 10, p. 123-140, 1994.
- BOERSMA, P.; WEENINK. *Praat: doing phonetics by computer* [Programa de computador]. Versão 6.0.33. Disponível em: <http://www.praat.org/>. Acesso em: 26 set. 2017.
- BROWMAN, C. P.; GOLDSTEIN, L. Towards an articulatory phonology. In: C., Ewan; ANDERSON, J. (Org.). *Phonology Yearbook 3*. Cambridge: Cambridge University Press, 1986, p. 219-252.
- _____. Articulatory gestures as phonological units. *Phonology*, Cambridge, v. 6, 201-251, 1989.
- CABREIRA, S. H. A monotongação dos ditongos orais decrescentes em Curitiba, Florianópolis e Porto Alegre. 115 f. Dissertação (Mestrado em Linguística) — Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1996.
- CAGLIARI, L. C. *An experimental study of nasality with particular reference to Brazilian Portuguese*. 320 f. Tese (PhD) — University of Edinburgh, Edimburgo, 1977.
- CALLOU, D.; MORAES, J.; LEITE, Y. Apagamento do R final no dialeto carioca: um estudo em tempo aparente e em tempo real. *D.E.L.T.A.*, vol. 14, 61-72, 1998.

- CÂMARA JR., J. M. *Estrutura da Língua Portuguesa*. Petrópolis: Vozes, 1970.
- COLLEY, M. *Diphthongization in Brazilian Portuguese*. 2009. 151 f. Tese (PhD) — Rice University, Houston, 2009.
- _____. Variação fonética nas vogais nasais e nasalizadas no português carioca. In: *Anais do 7º Congresso Internacional da Associação Brasileira de Linguística*. Curitiba: Editora da UFPR, 2011.
- DUARTE, M. E.; PAIVA, M. C. A variação linguística e o papel dos fatores linguísticos. *Revista da Abralin / Associação Brasileira de Linguística*, Curitiba, v. 1 (eletrônico), n. 1 (especial, 1ª parte 2011), p. 91-120, 2011.
- FLIGNER, M. A.; KILLEEN, T. J. Distribution-free two-sample tests for scale. *Journal of the American Statistical Association*, Alexandria, v. 71, n. 353, p. 210-213, 1976.
- GILLES, P.; TROUVAIN, J. Luxembourgish. *Journal of the International Phonetic Association*, Cambridge, v. 43, n. 1, p. 67-74, 2013.
- GRABE, E.; KOCHANSKI, G.; COLEMAN, J. Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech*, Berlin, v. 50, n. 3, 281-310, 2007.
- HAKULINEN, A.; VILKUNA, M.; KORHONEN, R.; KOIVISTO, V.; HEINONEN, T. R.; ALHO, I. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura, 2004.
- ITKONEN, T. Välikatsaus suomen kielen juuriin. *Virittäjä*, Helsinki, v. 87, n. 3, p. 349-386, 1983.
- KOHLER, K. J. ToBIG and PROLAB: Two prosodic transcription systems for German compared. *Report at Workshop on Prosodic Labelling*. ICPhS Stockholm. 1995.
- LASS, R. Intradiphthongal Dependencies. In: ANDERSON, J.; DURAND, J. (Ed.). *Explorations in Dependency Phonology*. Dordrecht: Foris Publications Holland, 1987. p. 109-131.
- LEÃO, T. M. R. *Falares regionais: panorama da redução de ditongos decrescentes no Brasil*. 2013. 33 f. Trabalho de conclusão de curso (Licenciatura em Letras) — Departamento de Linguística, Português e Línguas Clássicas, Universidade de Brasília, Brasília, 2013.
- LEITE, Y.; CALLOU, D.; MORAES, J. Processos em curso no português do Brasil: a ditongação. In: HORA, D.; COLLISCHONN, G. (Org.). *Teoria linguística: fonologia e outros temas*. João Pessoa: Universitária, 2003. p. 232-250.
- LEVITT, H.; NEUMAN, A. C. Evaluation of orthogonal polynomial compression. *The Journal of the Acoustical Society of America*, Melville, v. 90, n. 1, 241-252, jul. 1991.
- LOBANOV, B. M. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, Melville, v. 49, n. 2B, 606-608, fev. 1971.
- LOPES, R. A realização variável dos ditongos [ow] e [ej] no português falado em Altamira/PA. 97 f. Dissertação (Pós-Graduação em Letras/Mestrado em Linguística) — Centro de Letras e Artes, Universidade Federal do Pará, Belém, 2002.

LÓPEZ, B. S. *The sound pattern of Brazilian Portuguese (Cariocan dialect)*. 265 f. Tese (Doutorado em Linguística) - University of California in Los Angeles, Los Angeles, 1979.

MOLLICA, M. C. *A influência da fala na alfabetização*. Rio de Janeiro: Tempo brasileiro, 1998.

MORAES, J. A.; WETZELS, L. Sobre a duração dos segmentos vocálicos nasais e nasalizados em português. Um exercício de fonologia experimental. *Cadernos de Estudos Linguísticos*, Campinas, n. 23, p. 153-166, jul/dez. 1992.

MOTA, J. Variação entre /ei/ e /e/ em Sergipe. In: FERREIRA, C. et al. (Ed.). *Diversidade do português no Brasil: estudos de dialetologia rural e outros*. Salvador: Centro Editorial e Didático da UFBA, 1988. p. 143-148.

MUDRAK, O. (Resenha.) Irina Nikolaeva. A Historical Dictionary of Yukaghir. 2006. *Journal of Language Relationship*, Moscou, n. 3, p. 164-173, 2010.

NAU, N. *A short grammar of Latgalian*. München: Lincom Europa, 2011. (Languages of the World/Materials 482.)

NEAREY, T. M. *Phonetic Feature Systems for Vowels*. Bloomington: Indiana University Linguistics Club, 1978.

OSTENDORF, M.; PRICE, P. J., SHATTUCK-HUFNAGEL, S. The Boston University Radio News Corpus. *Boston University Technical Report* No. ECS-95-001. 2001.

OUSHIRO, L.; MENDES, R. B. Sali[ẽ]cia social e mudança linguística: a ditongação de /e/ nasal no português paulistano. *Revista do GEL*, São Paulo, v. 11, n. 32, p. 9-46, 2014.

PAIVA, M. C. Supressão das semivogais nos ditongos decrescentes. In: OLIVEIRA E SILVA, G.; SCHERRE, M. M. P. (Org.). *Padrões sociolinguísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro, 1996. p. 217-238.

_____. O percurso da monotongação de [ey]: observações no tempo real. In: PAIVA, M. da C. de; DUARTE, M. E. L. (Org.). *Mudança linguística em tempo real*. Rio de Janeiro: Editora Contracapa, 2003. p. 31-46.

PEREIRA, G. *A monotongação dos ditongos <ej>, <ow> e <aj> no português falado em Tubarão (SC): Estudo de casos*. 134 f. Dissertação (Mestrado em Ciência de Linguagem) — Universidade do Sul de Santa Catarina, Tubarão, 2004.

RECTOR, M. *A linguagem da juventude: um estudo geo-sociolingüístico*. Petrópolis: Editora Vozes, 1975.

ROCHA, M. S.; SILVA, A. I. dos A.; NEVES, F. J. L. das. Uma análise sobre a ditongação das vogais tônicas finais seguidas de /s/. *EFDeportes.com*, Buenos Aires, v. 20, n. 208, sep. 2015.

ROSEN, K. M. Analysis of speech segment duration with the lognormal distribution: A basis for unification and comparison. *Journal of Phonetics*, Los Angeles, v. 33, n. 4, 411-426, 2005.

- SAPIR, Y. Elfdalian, the Vernacular of Övdaln. In: NYSTRÖM, G. (Ed.). *Fuost konferensn um övdalskų / Första konferensen om älvdalska*. Uppsala: Uppsala Universitet, 2005.
- SCHMALSTIEG, W. R. Primitive east Baltic *-uo-, *-ie- and the 2nd Sg. ending. *Lingua*, v. 10, p. 369-374, 1961.
- SEARA, I. C.; NUNES, V. G.; LAZZAROTTO-VOLCÃO, C. *Fonética e Fonologia do Português Brasileiro*. 2. ed. Florianópolis: LLV/CCE/UFSC, 2011.
- SILVA, A. dos R. Variação fonética em capitais brasileiras: a ditongação diante de /s/ e as realizações fonéticas do /s/ em coda. In: HORA, D. da; PEDROSA, J. L. R.; LUCENA, R. M. de (Org.). *Anais da ALFAL 2014: XVII Congresso Internacional Asociación de Lingüística y Filología de América Latina (ALFAL)*. João Pessoa: Ideia, 2014. p. 4813-4829.
- SILVA, F. de S. O processo de monotongação em João Pessoa. In: HORA, D. da (Org.). *Estudos sociolinguísticos: perfil de uma comunidade*. João Pessoa: CNPq/ILAPPEC/VALPB, 2004. p. 29-44.
- STEVENS, K. N. *Acoustic phonetics*. Cambridge: MIT Press, 1998.
- SYRDAL, A. K.; GOPAL, H. S. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, v. 79, p. 1086-1100, 1986.
- TATHAM, M. Some Problems in Phonetic Theory. In: HOLLIEN, H.; HOLLIEN, P. (Org.). *Current Issues in the Phonetic Sciences*. Amsterdam: John Benjamins, 1979, p. 93-106.
- _____. Dynamic articulatory phonology and the supervision of speech production. In: ELENIUS, K.; BRANDERUD, P. (Org.). *Proceedings of the XIIIth International Congress of Phonetic Sciences (ICPhS 95)*. Stockholm: [s.n.], 1995, v. 1, p. 58-61.
- TENIŠEV, Ä. R.; DYBO, A. V. (Ed.). *Sravnitel'no-istoričeskaja grammatika tjurkskix jazykov. Prattjurskij jazyk-osnova. Kartina mira prattjurskogo ètnosa po dannym jazyka*. Moscou: Nauka, 2006.
- THOMAS, E. R. *Sociophonetics: an introduction*. London: Palgrave Macmillan, 2011.
- TIERSMA, P. The Nature of Phonological Representation: Evidence from Breaking in Frisian. *Journal of Linguistics*, Cambridge, v. 19, n. 1, p. 59-78, março 1983.
- XU, Y. 2007-2015. FormantPro.praat. Disponível em: <http://www.homepages.ucl.ac.uk/~uclyyix/FormantPro/>.
- XU, Y.; GAO, H. FormantPro as a tool for speech analysis and segmentation. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 26, n. 4, p. 1435-1454, 2018.
- VEADO, Rosa M. A redução do ditongo - uma variável sociolinguística. *Ensaios de Lingüística*, Belo Horizonte, n. 9, p. 209-229, 1983.

WATSON, C. I.; HARRINGTON, J. Acoustic evidence for dynamic formant trajectories in Australian English Vowels. *The Journal of the Acoustical Society of America*, Melville, v. 106, p. 458-468, 1999.

WETZELS, W. L. The lexical representation of nasality in Brazilian Portuguese. *Probus*, v. 9, n. 2, p. 203-232, 1997.

WETZELS, W. L. Contrastive and allophonic properties of Brazilian Portuguese. In: WANNER, D.; KIBBEE, D. A. (Ed.). *New Analyses in Romance Linguistics: Selected papers from the Linguistic Symposium on Romance Languages XVIII*, Urbana-Champaign, April 7-9, 1988. Amsterdam: John Benjamins, 1991. p. 77-100. (Current Issues in Linguistic Theory 69.)



Data de submissão: 13/03/2018

Data de aceite: 20/11/2018