

# Princípios básicos da teoria acústica de produção da fala<sup>1</sup>

Adriana S. Marusso  
Universidade Federal de Ouro Preto

## Abstract

This article presents the basic principles of the Acoustic Theory of Speech Production, also called *Linear Source-filter Theory of Speech Production* (FANT, 1960; KENT & READ, 1992; PICKETT, 1999). This description is justified since the theory is not available in Portuguese. This theory is useful in understanding articulatory-acoustic relationships. It may be applied to the analysis of a single language or in cross-linguistic studies. It also provides a foundation for many procedures for the acoustic analysis of speech.

## 1. INTRODUÇÃO

**E**ste artigo apresenta os princípios básicos da teoria acústica de produção da fala (FANT, 1960; KENT & READ, 1992; PICKETT, 1999). Tal descrição justifica-se, em primeiro lugar, por não contarmos com a publicação dessa teoria e seus princípios básicos em livros ou periódicos publicados em língua portuguesa. Em segundo lugar porque tal teoria continua sendo um recurso muito adequado para análise de uma única língua ou em estudos interlingüísticos, além de proporcionar uma fundamentação para muitos procedimentos de análise acústica da fala.

A onda acústica da fala é o principal meio pelo qual a mensagem do falante é comunicada ao ouvinte. A análise do sinal acústico da fala provê informações tanto sobre os processos de produção da fala, quanto dos processos pelos quais o ouvinte percebe a fala: produção e percepção.

A aplicação da análise acústica à fala requer a apreciação da inter-relação entre:

- 1- a teoria que sustenta a análise acústica da fala (a teoria acústica de produção da fala);
- 2- os instrumentos utilizados na análise;
- 3- as medidas resultantes de tais análises.

Neste artigo, apresentaremos a teoria acústica de produção da fala (FANT, 1960; KENT & READ, 1992; PICKETT, 1999). Os pontos 2 e 3 acima são metodológicos e dependem da aplicação da teoria. Para a aplicabilidade da teoria em questão recomendamos, dentre outros, DELATTRE, P., LIBERMAN, A., COOPER, F. & GERSTMAN,

L., 1952; GAY, T., 1978; FRY, D., 1979; LADEFOGED, P., 1993; LEHISTE, I. & PETERSON, G., 1961; LINDBLOM, B., 1963; LINDBLOM, B. & SUNDBER, J., 1971; MAEDA, S., 1991; MAGEN, H., 1989; OHMAN, S., 1966; KENT, DEMBOWSKI & LASS, 1996.

## 2. TEORIA ACÚSTICA DE PRODUÇÃO DA FALA

Esta teoria, comumente chamada de “teoria linear de fonte e filtro de produção da fala”, é muito útil para entender a relação entre aspectos articulatórios e acústicos. Também proporciona uma fundamentação teórica para vários procedimentos de análise acústica da fala. O princípio dessa teoria é que as características do sistema de produção da fala (o trato vocal) podem ser inferidas através da análise da saída acústica desse sistema. Essa teoria propõe que o sistema de produção da fala se divide em dois componentes primários: uma fonte, que provê a entrada acústica ao sistema; um filtro, que modula essa entrada.

Para facilitar a compreensão das principais características dessa teoria, apresentamos abaixo diagramas que representam alguns modelos de trato vocal. Tais diagramas estão baseados na proposta de Kent & Read (1992).

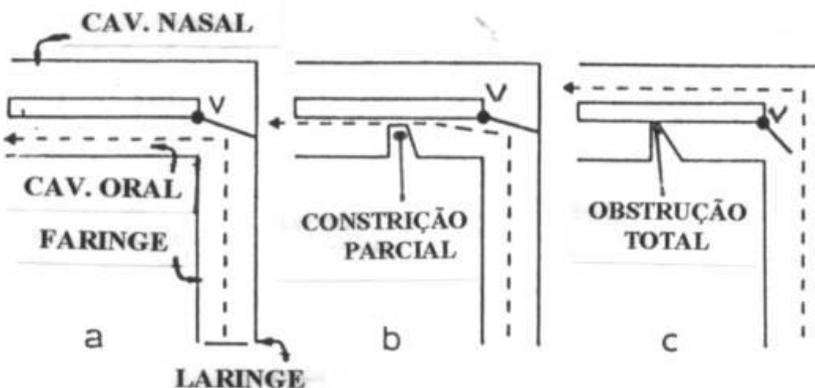


FIGURA 1: Modelos de trato vocal para diferentes classes de sons da fala: (a) vogais, (b) fricativas e (c) nasais.

Na figura 1, temos os modelos de trato vocal para diferentes classes de sons da fala: (a) vogais, (b) fricativas e (c) nasais. Em (a) não há constricção. Note-se a constricção parcial em (b) e a obstrução total em (c). Em (a) e (b), temos sons orais sendo que o véu palatino (**v**) está elevado e bloqueando a passagem de ar pela cavidade nasal; em (c), o véu palatino está abaixado, permitindo a saída do ar pelas narinas.

Observe-se o diagrama (a). As vogais são sons produzidos com vibração laríngea (nesse caso, o vozeamento é a fonte de energia) e um trato vocal relativamente aberto, configurado para produzir certos padrões de ressonância (de modo que o trato vocal inteiro funciona como um filtro ou como um sistema seletivo de transmissão de frequências). Um diagrama parecido com o de (a), porém com algumas modificações, explica as líquidas e as semivogais.

No diagrama (b), as fricativas são produzidas com um estreitamento em algum ponto do trato vocal. Ao passar por essa constricção, o ar gera um ruído turbulento. Esse ruído é então a fonte de energia para a produção dessa classe de sons. A fonte de ruído é filtrada pelo trato vocal, especialmente pela porção do trato vocal anterior à constricção. Esse modelo de diagrama sofre algumas modificações para explicar as oclusivas e as africadas. Ambas essas classes de sons envolvem uma breve oclusão do trato vocal. No caso das oclusivas, a obstrução é total como em (c), porém com o véu palatino elevado como em (a) e (b), obstruindo, desse modo, a passagem de ar pelo nariz. No caso das africadas, há um primeiro momento de obstrução total, como nas oclusivas, e um segundo momento de estreitamento, como nas fricativas (b). Tanto as oclusivas quanto as africadas são acusticamente complexas, pois resultam das rápidas transições articulatórias e acústicas inerentes a sua produção (*transients*). A fonte de energia tanto das oclusivas quanto das africadas resulta dos sucessivos aumentos no fluxo de ar que ocorrem ao soltar abruptamente a pressão de ar formada atrás da constricção.

Já no diagrama (c), os sons nasais são produzidos com o véu palatino abaixado, de modo que o som é radiado através da cavidade nasal. Se houver uma obstrução da cavidade oral, o som resultante

é uma consoante nasal. Se não houver tal obstrução, o som resultante é uma vogal nasalizada. As nasais, como as vogais, têm o vozeamento como fonte de energia. Entretanto, diferem das vogais porque o filtro da fonte de energia é determinado não apenas pela passagem oral, mas também pela passagem nasal.

A teoria acústica pode ser aplicada aos diversos sons da fala, entretanto, utilizaremos as vogais para exemplificar seu funcionamento. Para análises dos sons consonantais consultem-se, entre outros, BEHRENS, S. & BLUMSTEIN, S., 1988; BLUMSTEIN, S. & STEVENS, K., 1980; DART, S., 1991; FUJIMURA, O., 1962.

Para entender melhor como funciona o filtro, i.e., o trato vocal, imaginemos um tubo reto e uniforme, com um dos seus extremos parcialmente fechado por uma membrana elástica (que simula as cordas vocais) e com o outro extremo aberto (correspondente à abertura da boca). Eis o diagrama desse tubo (adaptado de KENT & READ, 1992):

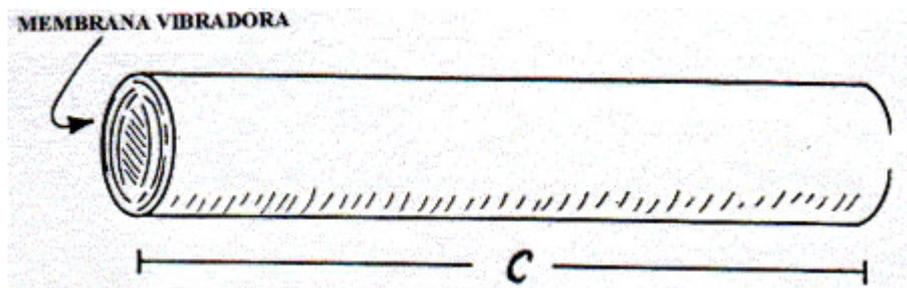


FIGURA 2: Modelo simples de produção de vogal: um tubo reto e uniforme com um dos extremos parcialmente fechado por uma membrana e com o outro extremo aberto.

A membrana vibra produzindo a fonte de energia acústica que viaja ao longo do tubo. O tubo é o ressonador. Esse tipo de tubo tem um número infinito de ressonâncias situadas em diferentes frequências. O valor das frequências de ressonância do tubo é determinado pela seguinte fórmula (KENT & READ, 1992:14):

$$F_n = (2n-1) c/4l$$

onde **c** é a velocidade do som (aproximadamente 35.000 cm/seg.) e **l** é o comprimento do tubo (do inglês *length*)

O significado da fórmula é: o tubo ressoa com amplitude<sup>2</sup> máxima, quando o comprimento da onda de um som é quatro vezes o comprimento do tubo. Sua aplicação resulta nas freqüências de ressonância do tubo. Na realidade, essas ressonâncias ocorrem em múltiplos e, por isso, a expressão (2n-1) é usada para gerar o conjunto de números ímpares. As ressonâncias ocorrem a  $1c/4l$ ,  $3c/4l$ ,  $5c/4l$ ,  $7c/4l$ , etc. Se assumirmos que o tubo tem um comprimento de 17,5cm., então a primeira ressonância terá uma freqüência calculada da seguinte forma:

$$\begin{aligned} F_1 &= ((2 \times 1) - 1) c/4l \\ &= (2-1) c/4l \\ &= 1 \times 35.000 \text{ cm/seg} / (4 \times 17,5\text{cm}) \\ &= \frac{35000\text{cm/seg}}{70 \text{ cm}} \\ &= 500 \text{ Hz} \end{aligned}$$

A segunda ressonância terá uma freqüência calculada da seguinte forma:

$$\begin{aligned} F_2 &= ((2 \times 2) - 1) c/4l \\ &= (4-1) c/4l \\ &= 3 \times 35.000 \text{ cm/seg} / (4 \times 17,5\text{cm}) \\ &= \frac{105000 \text{ cm/seg}}{70 \text{ cm}} \\ &= 1500 \text{ Hz} \end{aligned}$$

As ressonâncias mais altas são calculadas através das diferentes soluções de (2n-1), assim obtemos 2500, 3500, 4500Hz, etc. As freqüências ocorrem a intervalos de 1000Hz a partir da primeira freqüência (F1) em 500Hz.

Para tornar esse exemplo relevante para a produção da fala, temos que levar em consideração os seguintes pontos:

- 1) a média do trato vocal de um homem adulto é de, aproximadamente, 17,5cm. – da glote até os lábios;
- 2) o trato vocal tem, aproximadamente, as mesmas freqüências de ressonância que o tubo aqui exemplificado.

Isto é, um tubo tão simples como o que foi descrito é suficiente para produzir uma vogal. Em seguida, vemos a configuração do trato vocal ao produzir essa vogal.

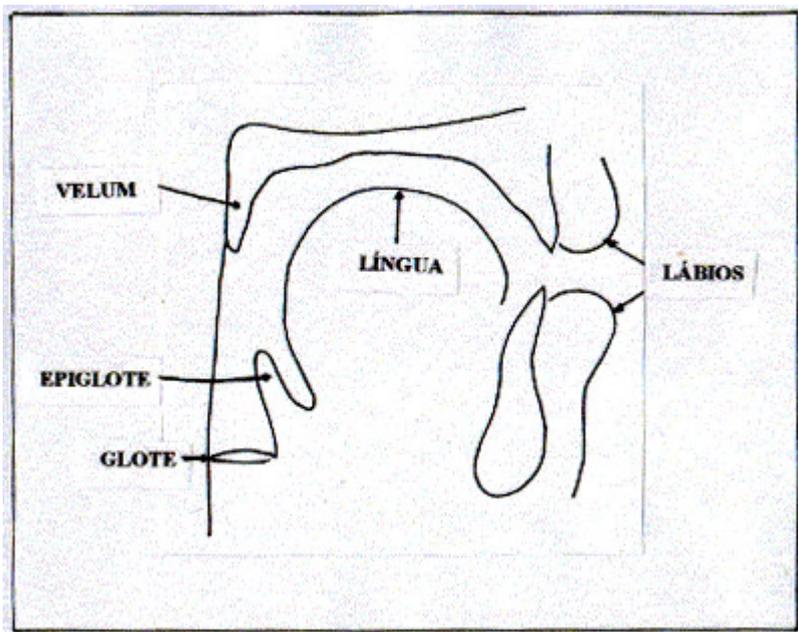


FIGURA 3: Configuração do trato vocal que corresponde ao formato do tubo em FIG. 2. A área transversal é essencialmente a mesma da glote até os lábios.

Observe-se que, assim como na figura 2, o trato vocal da figura 3 é uniforme, sendo a área transversal essencialmente a mesma da glote até os lábios. A vogal produzida com essa configuração do trato

vocal é o *schwa*. Essa configuração do trato vocal é também freqüentemente usada nas hesitações da fala.

Se mudarmos o comprimento do tubo, mudarão também as freqüências das ressonâncias. Se o comprimento do tubo for o dobro, i.e., 35cm., então as freqüências de ressonância serão proporcionalmente mais baixas: 250, 750, 1250Hz, etc. Por outro lado, se o comprimento do tubo for 8,75cm., então as freqüências serão 1000, 3000, 5000Hz, etc. Esses resultados explicam porque, em um instrumento musical como o órgão, os tubos mais longos produzem os tons mais baixos, e os tubos mais curtos produzem os tons mais altos. Do mesmo modo, na fala, podem-se explicar as diferenças nas freqüências de uma criança para um adulto. Uma criança tem um trato vocal que é, aproximadamente, a metade daquele de um adulto. Por isso, crianças têm freqüências mais altas do que os adultos. Podemos concluir que o comprimento do trato vocal do falante determinará a altura das freqüências de ressonância: quanto mais comprido for o tubo, mais baixas serão as freqüências e menor será a separação entre as freqüências. Por outro lado, quanto mais curto for o trato vocal, mais altas e mais separadas serão as freqüências.

Conforme foi dito, os resultados até aqui mostrados pertencem a uma única vogal: uma vogal média central (i.e., *schwa*), que é, transversalmente, a mesma ao longo do trato vocal. Mas quais são as freqüências para as outras vogais? A resposta pode ser determinada experimentalmente, descobrindo-se as freqüências de ressonância para várias formas de tubo do mesmo comprimento. Sendo o tubo reto (FIG. 2) ou curvo (FIG. 3), os valores das freqüências não são afetados significativamente. Portanto, uma vez que é mais fácil desenhar tubos retos, mostraremos os formatos correspondentes a outras vogais através de modelos de tubos desse formato, conforme vemos na figura a seguir.

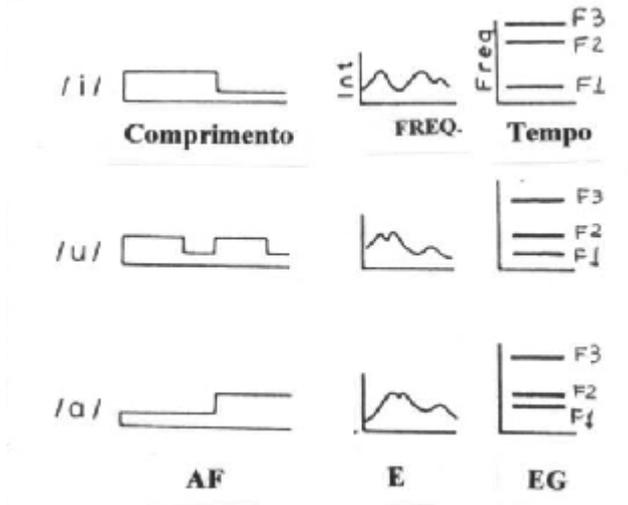


FIGURA 4: Idealização da área de função (AF), do espectro (E) e do espectrograma (EG) para as vogais /i, u, a/. Em AF, o lado fechado representa a glote e o aberto, os lábios. Os formantes estão representados em E como picos e, em EG, como bandas horizontais (figura adaptada de KENT & READ, 1992).

Na figura 4, cada diagrama da coluna da esquerda (AF) corresponde à forma do trato vocal de uma vogal. O primeiro desenho corresponde à vogal /i/, o segundo, a /u/, e o terceiro corresponde a /a/. Na fig. 4, os diagramas que estão na coluna do meio ilustram o espectro<sup>3</sup> correspondente a essas vogais. Os picos espectrais são as frequências de ressonância dos tubos. Compare-se o tubo que representa a vogal /i/ com aquele que representa a vogal /a/. O estreitamento da parte direita do tubo que representa a vogal /i/ mostra a constrição palatal para /i/, enquanto que o estreitamento da parte esquerda do tubo que representa a vogal /a/ mostra a constrição faríngea para /a/. Do mesmo modo, todas as outras vogais poderiam ser modeladas através de modificações no tubo.

Então, qual seria a relação entre um ressonador (tal como um tubo) e uma fonte de energia (tal como uma membrana elástica que vibra)? Em grande parte, a fonte de energia e o ressonador são independentes, exceto por condições especiais. Esse fato é muito

importante e explica porque um falante pode produzir uma vogal [i] com altura melódica<sup>4</sup> baixa ou alta sem perder a distinção fonética da vogal. A altura melódica da vogal é determinada quase totalmente pelas freqüências de vibração das cordas vocais. Quanto mais baixa for a taxa de vibração, mais baixa a altura melódica. Então, um registro de voz baixo tem uma freqüência de vibração das cordas vocais mais baixa do que a de soprano. Mas a freqüência de vibração das cordas vocais não afeta as propriedades do ressonador. As freqüências de ressonância de um tubo ressonador são determinadas quase que totalmente pelos seguintes fatores: o comprimento do tubo e sua área transversal como função do seu comprimento. A mudança da freqüência da fonte de energia não muda as freqüências de ressonância do tubo que recebe essa energia.

Os princípios que foram introduzidos até agora podem ser resumidos na figura 5.

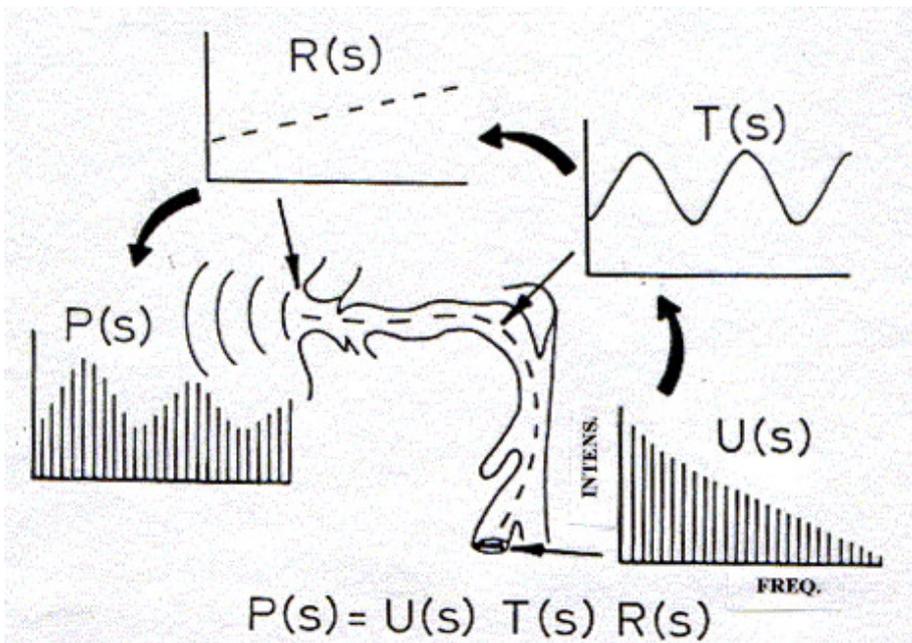


FIGURA 5: Representação diagramática do conceito de fonte e filtro para vogais.

A figura 5 está baseada na proposta de KENT & READ, 1992. O espectro da fonte laríngea,  $U(s)$ , é filtrado pela função de transferência do trato vocal,  $T(s)$ , e a radiação característica,  $R(s)$ , para dar como resultado o espectro de saída,  $P(s)$ . Matematicamente,  $P(s) = U(s) \cdot T(s) \cdot R(s)$ , onde  $s$  = frequência.

Essa teoria, aplicada à produção de vogais, evidencia que a energia de saída é o produto da energia da fonte e do ressonador (ou filtro). Essa teoria é também conhecida como *teoria linear de fonte e filtro* por estar baseada em um modelo matemático linear. É necessário entender a energia da fonte na forma de um espectro. As cordas vocais, ao vibrarem, produzem um espectro, similar a  $U(s)$  na figura 5, no qual a energia decresce a frequências distintas determinadas pela taxa de vibração das cordas vocais. Observe-se que em  $U(s)$ , a intensidade está representada no eixo vertical (ordenada) e a frequência, no eixo horizontal (abscissa). O resultado é chamado de **espectro linear**, ou seja, é um espectro no qual a distribuição da energia toma forma de linhas. O comprimento relativo de cada linha representa a amplitude relativa de cada componente. O espectro da energia do vozeamento pode ser idealizado como um espectro linear no qual as linhas individuais caem a múltiplos integrais da frequência vibratória fundamental<sup>5</sup>. Por exemplo, a **frequência fundamental** média de um homem adulto é de 120Hz, e a energia deste espectro da fonte cairá a frequências de 120, 240, 360, 480Hz e assim sucessivamente. Entretanto, um homem pode produzir frequências de vibração mais altas ou mais baixas do que esse valor médio. Por exemplo, se a frequência fundamental de um homem subir para 300Hz, então a energia do espectro fonte cairá a 300, 600, 900, 1200Hz e assim sucessivamente.

Já a frequência fundamental típica de uma mulher é de, aproximadamente, 225Hz, mas também é possível obter outros valores. Em crianças, o valor típico é de 300Hz, mas também não é o único. Essas mudanças na frequência vibratória de um determinado falante são mudanças que afetam apenas a fonte, sem conseqüências para o ressonador ou filtro. Do mesmo modo, a amplitude da

vibração das cordas vocais pode mudar. Um falante pode falar em um volume baixo ou alto. Tais mudanças apenas afetam o ressonador, determinando o nível de energia que o mesmo receberá. A independência relativa entre fonte e filtro possibilita que produzamos fala inteligível com diferentes fontes de energia, incluindo a altura melódica alta ou baixa, o sussurro, o falar grave ou outras variações fonatórias.

Conforme vimos na figura 5, é característica do espectro laríngeo que a energia, em seus componentes harmônicos,<sup>6</sup> decline à medida que a frequência aumenta. Esse declínio da energia dos harmônicos mais altos mostra que a maior energia da fala vozeada está nas frequências mais baixas. A taxa do declínio da energia é de 12 dB por oitava ou uma queda de energia de 12 dB cada vez que a frequência dobra. Pode-se dizer que o espectro laríngeo é um espectro linear no qual a energia dos harmônicos cai com a frequência a uma taxa de 12 dB/oitava.

As ressonâncias do filtro são tecnicamente chamadas de **formantes**. Um formante é um modo natural de vibração (i.e., ressonância) do trato vocal. Teoricamente, há um número infinito de formantes porque o trato vocal, assim como todos os tubos ressonadores, possui um número infinito de frequências de ressonância. Porém, como a energia laríngea que ativa as ressonâncias ocorre principalmente a frequências inferiores a 5 kHz, apenas os três ou quatro primeiros formantes são suficientes para a análise acústica de segmentos fonéticos e para a identificação perceptiva dos sons da fala. Os formantes não produzem energia, apenas modificam a energia produzida pela fonte. São identificados por números – F1, F2, F3, e F4 – numerados sucessivamente partindo do formante com frequência mais baixa (FIG. 6). Cada formante pode ser descrito por duas características: a **frequência central** de ressonância (conhecida como frequência do formante) e a **largura da banda** (do inglês **bandwidth**). A frequência central refere-se às frequências de maior amplitude, e a largura da banda é a variação nos valores de frequência com amplitude maior do que 3 dB no centro de maior amplitude.

Juntos, os formantes constituem a **função de transferência** do trato vocal. A função de transferência é a relação entrada-saída e é também uma forma de descrever o processo de filtragem. Cada formante é associado a um pico na função de transferência (FIG. 5) e, portanto, é potencialmente associado a um pico no espectro de saída (ou radiação do espectro).

Por último, a **radiação característica** se refere ao efeito de filtragem que surge quando um som sai da boca e é irradiado no espaço exterior. O som irradiado se espalha em todas as direções ao sair da boca. Esse tipo de radiação característica funciona como um filtro que reduz mais a energia nas frequências baixas do que nas altas. Assumimos que o som de saída aumenta em frequência a uma taxa de 6dB/oitava. Como essa é uma característica constante, é comumente combinada com a queda de 12 dB/oitava do espectro laríngeo para dar como resultado 6 dB/oitava.

Conforme vimos na FIG. 5, a teoria de fonte e filtro se resume na seguinte equação matemática:

$$P(f) = U(f) T(f) R(f)$$

$P(f)$  é a pressão espectral do som irradiado.  $P$  representa a pressão e  $(f)$  indica a função da frequência. Como o ouvido humano responde a variações de pressão atmosférica, é útil descrever o sinal de saída da fala como uma onda da pressão do som (no domínio temporal) ou como um espectro da pressão do som (no domínio das frequências). Os três termos da parte direita da equação referem-se, respectivamente, ao espectro da fonte laríngea, à função de transferência do trato vocal e à radiação característica. O termo  $U$  se refere à velocidade do volume e é usado porque as cordas vocais atuam como uma fonte de pulsos de ar. Transformando a equação em palavras, temos que a onda de pressão do som irradiado é o produto do espectro laríngeo, da função de transferência do trato vocal e da radiação característica.

Assumimos que os termos  $U(f)$  e  $R(f)$  são constantes nas diferentes vogais. Isto é, as diferentes vogais são descritas como variações da função de transferência,  $T(f)$ , e do espectro radiado,  $P(f)$ . Como  $T(f)$  consiste nos formantes das vogais, é através dos diferentes padrões dos formantes que são descritas as diferentes vogais. Os formantes  $F_1$ ,  $F_2$  e  $F_3$  são suficientes para caracterizar, acusticamente, qualquer vogal. Os demais formantes, i.e.,  $F_4$ ,  $F_5$ , etc., permanecem bastante constantes em termos de altura de frequência independentemente das mudanças articulatórias (PICKETT, 1999), por isso são geralmente desconsiderados nas análises acústicas que têm por objetivo caracterizar foneticamente os sons da fala. Contudo, esses formantes mais altos fornecem informação quanto à qualidade da voz e, portanto, são estudados quando se trata de identificação de falantes, por exemplo, na Fonética Forense.

Segundo KENT & READ (1992), formante é sinônimo de ressonância do trato vocal. O formante é comumente associado a um pico no espectro acústico, mas não necessariamente. Um dos objetivos da análise acústica é estimar a estrutura dos formantes de um segmento sonoro. Em seguida, temos um espectrograma das vogais anteriores do inglês, mais precisamente, das palavras *beed*, *bid*, *head*, *bad*. Na escala vertical, estão representados os valores de frequência em Hz: as frequências mais baixas estão representadas na parte mais baixa do espectrograma. Na escala horizontal, está representado o tempo, em milésimos de segundo (msec.), aumentando de esquerda para direita a partir de um ponto inicial arbitrário (o primeiro som da palavra *beed* está no extremo esquerdo). O grau de negrura no espectrograma mostra a amplitude a diferentes frequências no sinal acústico num determinado ponto na dimensão temporal. Observa-se que, no caso das vogais, essa maior concentração de energia se dá em forma de bandas estreitas (os formantes) na parte inferior do espectrograma. Podem-se observar os primeiros quatro formantes de cada vogal. O  $F_1$  é o mais baixo, i.e., o que possui menor frequência. Os outros formantes –  $F_2$ ,  $F_3$  e  $F_4$  – vão aumentando no valor de frequência de ressonância.

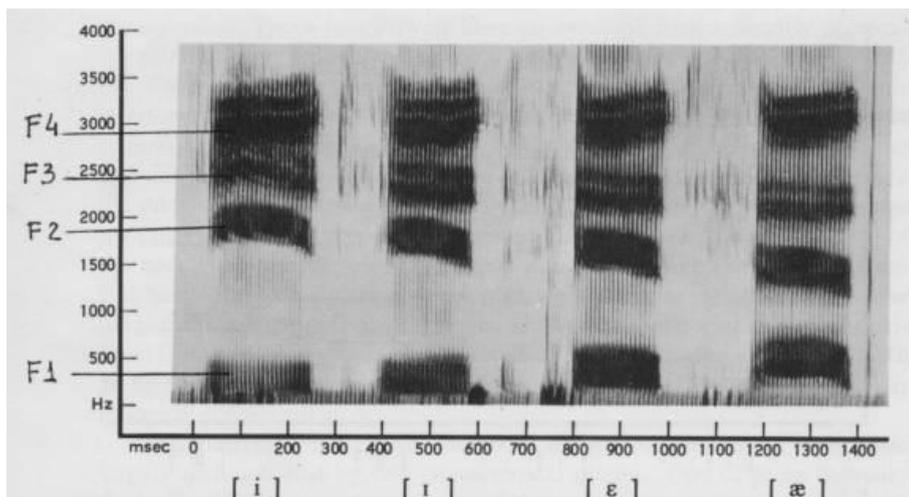


FIGURA 6: Espectrograma das vogais anteriores do inglês produzidas nas palavras *beed*, *hid*, *head*, *had*.

#### Frequência de F1:

A frequência do F1 abaixa com a constrição próxima à metade anterior do trato, i.e., a cavidade oral. Isso ocorre, por exemplo, quando os lábios se fecham para formar /u/ ou quando a língua se eleva perto dos alvéolos para /i/. A frequência é aumentada pela constrição na parte posterior-inferior do trato, i.e., a faringe, como ocorre para /ɑ/. Observe-se a FIG. 6 e comparem-se os valores de F1 para /i/, que é uma vogal alta, e para /æ/, que é uma vogal baixa.

#### Frequência de F2:

A frequência do F2 abaixa pela constrição nos lábios ou na parte mais posterior da cavidade oral bem acima da faringe. É o caso do que ocorre para /u/. A frequência do F2 sobe pela constrição na parte anterior da cavidade oral, atrás dos lábios e dentes. Isso ocorre para /i/ (cf. na figura 6, F2 de /i/ a, aproximadamente, 2000Hz).

#### Frequência de F3:

A frequência do F3 abaixa pela constrição nos lábios, na parte média da cavidade oral ou na parte superior da faringe. É o que

ocorre na produção de /r/ ou de /ʒ/. A frequência do F3 sobe pela constrição perto da cavidade oral anterior, como para /j/, ou perto da área da cavidade oral posterior – parte superior da faringe como para /ɑ/.

Resumindo, os formantes abaixam:

- 1- pelo alongamento do tubo do trato vocal;
- 2- pela constrição labial.

As duas coisas ocorrem em português na produção da vogal posterior alta [u], que é convencionalmente arredondada em português e, portanto, têm as frequências de formantes mais baixas das vogais.

Os primeiros trabalhos sobre acústica das vogais associam a frequência do F1 com a altura da vogal (i.e., o grau relativo de constrição entre a língua e o palato na cavidade oral) e a frequência do F2 com o lugar anterior-posterior da constrição na cavidade oral. A regra geral para F1 é que sua frequência é inversamente proporcional à altura da língua. Então, vogais baixas como /ɑ/ têm F1 alto; já vogais como /i/ têm o F1 baixo. Então, o que chamamos de vogais baixas são abertas no trato anterior, mas formam uma constrição posterior (faríngea), enquanto aquelas que chamamos de vogais altas são formadas por uma constrição mais anterior. A regra geral para F2 é que sua frequência abaixa na medida em que o lugar da constrição se move de uma região anterior para uma posterior dentro da cavidade oral.

Essas regras mostram algumas posições características dos formantes para vogais ou segmentos parecidos com vogais que podem ser produzidos de diferentes formas. A partir dos valores de F1 e de F2 é possível construir um gráfico com as vogais de uma língua. Em seguida temos um gráfico com os valores dos formantes para as vogais tônicas orais do português brasileiro (PB)<sup>7</sup>.

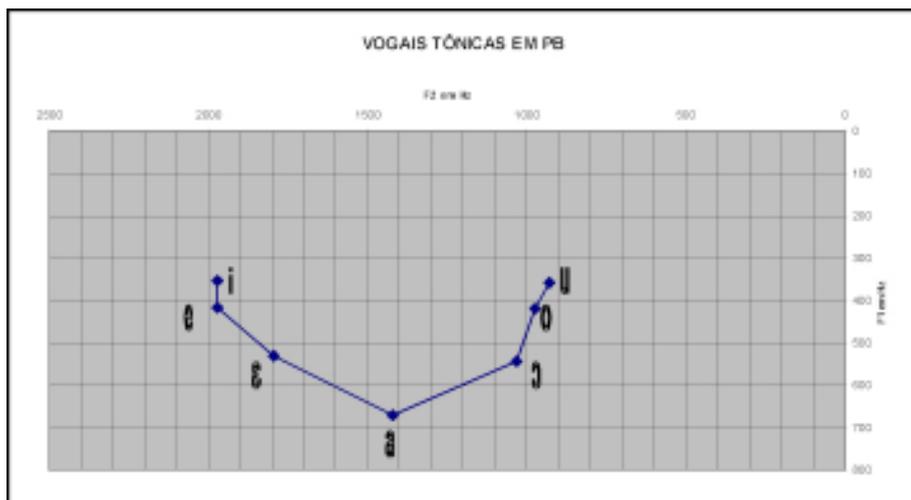


GRÁFICO 1: Valores dos formantes para as vogais tônicas orais do português brasileiro.

O gráfico 1 exemplifica a aplicação das regras gerais para F1 e F2 que foram descritas. Observa-se que a vogal alta, anterior, não-arredondada /i/, tem um F1 baixo (353 Hz) e um F2 alto (1970 Hz). Já a vogal alta, posterior, arredondada /u/ possui um F1 baixo (358 Hz) e um F2 também baixo (928 Hz) devido ao arredondamento dos lábios e por ser uma vogal posterior. Já a vogal baixa, central, não-arredondada /a/, tem um F1 alto (670 Hz) e um F2 médio (1420 Hz).

### 3. A APLICAÇÃO DA TEORIA ACÚSTICA E SEUS DESAFIOS

A teoria descrita na seção anterior pode ser utilizada na análise acústica das línguas naturais. A avaliação acústica segmental pode contribuir para identificar particularidades de um segmento e corroborar as análises articulatórias e auditivas. Embora existam alguns trabalhos de análise segmental do português (cf. ABAURRE, M. & CAGLIARI, L., 1986; CALLOU, D., DE MORAES, J. & LEITE, Y., 1996; SILVA, E., 1994; SILVA, M., 1989; VIEGAS, M., 1987; YACOVENCO, M., 1993; MARUSSO, A., 2003) é ainda necessário a ampliação dessa área de pesquisa.

A análise acústica é um recurso útil, por exemplo, para caracterizar acusticamente as vogais tônicas e átonas de uma língua ou também pode permitir uma diferenciação mais apurada de segmentos relativamente parecidos em duas línguas diferentes. Dessa forma se torna uma ferramenta valiosa para os estudos de aquisição de segunda língua, trazendo conseqüências positivas para o ensino de línguas estrangeiras (cf. BAPTISTA, 2000).

A análise acústica é também de grande valia para o desenvolvimento da ciência computacional e das telecomunicações. Nessa área, o desafio é saber quais elementos acústicos são realmente relevantes para finalidades específicas, por exemplo, de fala sintética, de reconhecimento automático de fala, etc. A definição dos aspectos acústicos e articulatórios da língua e do seu sistema fonológico permite o aperfeiçoamento dos equipamentos. Por outro lado, na Fonética Forense, o aprofundamento dos estudos sobre a caracterização das particularidades da fala individual permite a utilização de evidências da fala para a identificação de criminosos.

Embora a análise acústica tenha evoluído bastante desde seu surgimento, há ainda muitos pontos a serem resolvidos. Kent, Dembowski & Lass (1996) apresentam, entre outros, os seguintes: a normalização do trato vocal do falante, ou seja, o ajuste dos dados acústicos de modo que se cancelem as diferenças associadas com o tamanho do trato vocal permanece como desafio. Embora já tenham sido propostos alguns algoritmos, nenhum deles consegue superar a capacidade do ser humano de reconhecer sons produzidos por diferentes falantes.

A descrição acústica de diferentes tipos de voz, inclusive das que apresentam patologias, requer o uso de métodos acústicos para derivar descrições quantitativas das vozes normais e patológicas. O sucesso nesse sentido tem sido limitado. Esse mesmo problema está presente da descrição acústica de diferentes padrões de fala, incluindo tanto dialetos quanto problemas de fala.

A integração da informação segmental e supra-segmental requer que as características acústicas desses dois aspectos da fala

sejam combinadas num mesmo sistema. As pesquisas já identificaram alguns princípios básicos, mas há necessidade de mais estudos nessa área.

#### 4. LEITURAS E SITES RECOMENDADOS

Este artigo é apenas uma introdução à teoria acústica, portanto, as referências bibliográficas apresentadas são insuficientes para o leitor interessado em um estudo mais amplo e profundo dessa teoria. Para tanto, em seguida, apresentamos um detalhamento maior de algumas das obras citadas e de outras que recomendamos. A ordem não é de importância senão cronológica, começando por aquela que, de certa forma, deu origem à teoria.

FANT, G. *Acoustic Theory of Speech Production*. Singular Publishing, San Diego, 1960.

Este clássico é um tratado técnico sobre a análise acústica de produção da fala. É a primeira exposição completa da teoria de fonte-filtro e suas aplicações. Abrange todas as maiores classes de sons e de traços, com ênfase especial na teoria de traços distintivos. Um aspecto muito importante é a apresentação de gráficos que mostram como as frequências dos formantes do trato vocal são afetadas por dois fatores principais da constrictão: o ponto da constrictão e o grau da mesma, e a combinação de ambos. Usando esses gráficos, é possível encontrar a frequência de formantes para muitos tipos de configuração do trato vocal e examinar as mudanças decorrentes de diferentes pontos e graus de constrictão. A teoria também explica os efeitos de graus de nasalidade e de articulações de líquidas e de sons retroflexos.

FLANAGAN, J.L. *Speech Analysis Synthesis and Perception*. 2<sup>nd</sup>. Edition. Springer-Verlag, New York, 1972.

Esta obra é também um clássico que trata, de forma completa, de todos os aspectos e ramificações da fala desde o ponto de vista da acústica e da engenharia.

LINDBLOM, B. and ÖHMAN, S. (Eds.) *Frontiers of Speech Communication Research*. Academic Press, London, 1979.

É uma coleção de importantes artigos de pesquisas realizadas nos anos 70 que abordam uma ampla gama de tópicos sobre acústica, fisiologia e percepção da fala. É indicado para alunos avançados.

BOOTHROYD, A. *Speech Acoustics and Speech Perception*. Allyn and Bacon, Boston, 1986.

Este livro substitui o clássico Fry (1979) como um breve livro introdutório à análise acústica. Apresenta, de forma clara e profunda, todos os princípios acústicos básicos assim como também tópicos sobre fala e audição. É indicado para estudantes iniciantes.

KENT, R. and READ, C. *The Acoustic Analysis of Speech*. Singular Publishing, San Diego, 1992.

Este pequeno livro, em apenas oito capítulos, aborda, de forma eficiente, todos os princípios básicos da acústica da fala e dos fatores relacionados num nível introdutório.

BORDEN, G., HARRIS, K. and RAPHAEL, L. *Speech Science Primer*. 3<sup>rd</sup>. edition. Williams and Wilkins, Baltimore, 1994.

Este livro é bastante indicado para cursos introdutórios à ciência da fala. O tratamento é simples, mas completo, com exemplos e ilustrações. Apresenta os princípios acústicos básicos, produção de sons vocálicos, articulação de vogais e consoantes e percepção da fala. Inclui também a explicação de importantes teorias como a teoria de fonte-filtro de produção da fala e a teoria motora de percepção. Traça a história da ciência da fala, através das biografias dos pioneiros, e também da própria evolução da fala. O capítulo sobre ferramentas de pesquisa descreve, de forma excelente, os métodos de pesquisa, incluindo muitos exemplos com dados analisados, o que é de grande valia em um livro introdutório.

DENES, P. and PINSON, E. *The Speech Chain. The Physics and biology of Spoken Language*. 2<sup>nd</sup>. Edition. W. H. Freeman, New York, 1995.

Esta é a edição revisada e ampliada do clássico sobre acústica da fala, publicado vinte e cinco anos antes. Além dos capítulos sobre física, anatomia e fisiologia, esta edição inclui capítulos sobre análise, síntese e reconhecimento automático de fala via computadores.

LASS, N. (Ed.). *Principles of Experimental Phonetics*. Mosby, St. Louis, 1996.

É uma excelente coleção de discussões de nível avançado, feitas por autoridades em cada área, cobrindo métodos de pesquisa, acústica da fala, produção e percepção da fala.

STEVENS, K. *Acoustic Phonetics*. MIT Press, Cambridge, MA, 1998.

Nos Estados Unidos, Kenneth Stevens é um dos líderes do estudo acústico da fala. Seu livro apresenta uma descrição completa dos princípios acústicos da fala e da percepção da fala com ênfase no componente fonológico das línguas. Inclui também a teoria quântica, desenvolvida pelo próprio Stevens.

Em seguida, temos alguns *sites* de interesse para pesquisa especialmente em Fonética Acústica.

[www.lpl.univ-aix.fr](http://www.lpl.univ-aix.fr): *site* do *Laboratoire Parole et Langage* e do *Centre National de la Recherche Scientifique*. Tem *links* para publicações, pesquisas em andamento, grupos de pesquisa, divulgação de eventos, etc.

[www.ling.ed.ac.uk](http://www.ling.ed.ac.uk): *site* do Departamento de Lingüística Teórica e Aplicada da Universidade de Edimburgo, com *links* para *Centre for Speech Technology Research* e para *Language Evolution & Computation Research Unit*. Este último com softwares disponíveis para *download*. Ambos incluem *links* para publicações, pesquisas em andamento, grupos de pesquisa, divulgação de eventos, etc.

[www.arts.gla.ac.uk/IPA/ipa.html](http://www.arts.gla.ac.uk/IPA/ipa.html): *site* da Associação Internacional de Fonética com informações sobre o histórico da Associação, artigos,

membros do Comitê e ainda provas de proficiência fonética, tabelas com símbolos e fontes de símbolos disponíveis para *download*.

[www.phon.ucl.ac.uk](http://www.phon.ucl.ac.uk): *site* da Universidade de Londres com *links* para publicações, pesquisas em andamento, grupos de pesquisa, divulgação de eventos, etc. As publicações estão organizadas por áreas específicas: percepção e audição da fala, reconhecimento e síntese de fala.

[www.asa.aip.org](http://www.asa.aip.org): *site* da *Acoustical Society of America*, associação responsável pela publicação do *Journal of the Acoustical Society of America*. Inclui informação sobre como se associar, publicações e registros sonoros tanto de seres humanos quanto de animais, foguetes, etc. Disponibiliza, ainda, um folheto sobre como construir ambientes acusticamente adequados.

[www.praat.org](http://www.praat.org): *site* de Paul Boersma e David Weenink, ambos do Instituto de Ciências Fonéticas da Universidade de Amsterdam e criadores do programa de análise acústica PRAAT©. Tal programa, disponível para *download* gratuito, é muito sofisticado e completo. Com ele é possível estudar aspectos segmentais e supra-segmentais da fala, manipular, rotular e segmentar a fala, trabalhar com fala sintética, realizar gráficos e, ainda, programar funções específicas.

## 5. CONCLUSÃO

Este artigo apresentou os princípios básicos da teoria acústica de produção de fala. As vogais serviram como exemplo para explicar sua aplicação, embora a teoria possa ser aplicada aos diversos sons da fala. Em seguida, foram apontados os méritos e os desafios à teoria, sinalizando, de algum modo, rumos para futuras pesquisas. Por último, foram indicadas obras que permitem um aprofundamento do tema em questão, e *sites* úteis tanto pelo acesso a publicações recentes e pesquisas em andamento, quanto pelo livre acesso a software específico.

## NOTAS

<sup>1</sup> Este artigo é parte do trabalho desenvolvido na Tese de Doutorado intitulada *Redução Vocálica: Estudo de caso no português brasileiro e no inglês britânico*, FALE/UFMG, fevereiro/2003.

<sup>2</sup> Amplitude é a magnitude de deslocamento de uma onda sonora na dimensão temporal. Até certo ponto, a amplitude de um som determina o volume no qual o som é percebido.

<sup>3</sup> Espectro é a representação gráfica de um som que mostra a amplitude plotada em função da frequência.

<sup>4</sup> Altura melódica corresponde ao termo *pitch* do inglês.

<sup>5</sup> A frequência fundamental é a frequência mais baixa, i.e., a primeira senóide, de uma onda complexa periódica.

<sup>6</sup> Cada linha vertical é um harmônico da frequência fundamental.

<sup>7</sup> Este gráfico foi construído baseado na média geral para as vogais tônicas orais do português brasileiro obtidas por CALLOU, D., DE MORAES, J. & LEITE, Y. (1996).

## REFERÊNCIAS BIBLIOGRÁFICAS

ABAURRE, M. & CAGLIARI, L. Investigação instrumental das relações de padrões rítmicos e processos fonológicos no português brasileiro. *Cadernos de Estudos Lingüísticos*, 10, Campinas, 1986. p. 39-57.

BAPTISTA, B. *The acquisition of English vowels by Brazilian-Portuguese speakers*. ARES 6, Florianópolis: UFSC, 2000.

BEHRENS, S. & BLUMSTEIN, S. Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics*, 16, 1988. p. 295-298.

BLUMSTEIN, S. & STEVENS, K. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66, 1979. p. 1001-1017.

CALLOU, D., DE MORAES, J. & LEITE, Y. O vocalismo do português do Brasil. In: *Letras de Hoje*, nº 104, 1996. p. 27-40.

DART, S. Articulatory and acoustic properties of apical and laminal articulations. *Working Papers in Phonetics*, 79, Linguistics Department, Univ. of California, Los Angeles, 1991.

DELATTRE, P., LIBERMAN, A., COOPER, F. & GERSTMAN, L. An experimental study of the acoustic determinants of vowel colour: observations on one and two-formant vowels synthesized from spectrographic patterns. In: *Word* 8, 1952. p. 195-210.

FANT, G. On the predictability of formant levels and spectrum envelopes from formant frequencies. In: Halle, M, Lunt, H. & Maclean, H. (Eds.) *For Roman Jakobson*. The Hague: Mouton, 1956.

\_\_\_\_\_. *Acoustic Theory of Speech Production*. The Hague: Mouton, 1960.

FRY, D. *The physics of speech*. Cambridge Textbooks in Linguistics, 1979.

FUJIMURA, O. Analysis of nasal consonants. *Journal of the Acoustical Society of America*, 34, 1962. p. 1865-1875.

GAY, T. Effect of speaking rate on vowel formant movements. In: *Journal of the Acoustical Society of America* 63, 1978. p. 223-230.

KENT, R., DEMBOWSKI, J. & LASS, N. The Acoustic Characteristics of American English. In: Lass, N. (ed) *Principles of Experimental Phonetics*. Mosby, St. Louis, 1996.

KENT, R. & READ, C. *The Acoustic Analysis of Speech*. San Diego, California: Singular Publishing Group, Inc., 1992.

LADEFOGED, P. *A Course in Phonetics*. 3<sup>rd</sup>. ed. Harcourt Brace Jovanovich, Inc. (1.ed. 1975), 1993.

LEHISTE, I. & PETERSON, G. Transitions, glides and diphthongs. In: *Journal of the Acoustical Society of America* 33, 1961. p. 268-277.

LINDBLOM, B. Spectrographic study of vowel reduction. In: *Journal of the Acoustical Society of America* 35 (11): 1963. p. 1773-1781.

LINDBLOM, B. & SUNDBER, J. Acoustical consequences of lip, tongue, jaw and larynx movement. In: *Journal of the Acoustical Society of America* 50, 1971. p. 1166-1179.

MAEDA, S. On articulatory and acoustic variabilities. In: *Journal of Phonetics* 19, 1991. p. 321-331.

MAGEN, H. *An acoustic study of vowel-to-vowel coarticulation in English*. PhD Thesis, Yale, 1989.

MARUSSO, A. *Redução Vocálica: Estudo de caso no português brasileiro e no inglês britânico*. 2003. 454f. Tese (Doutorado em Estudos Lingüísticos) – FALE/UFMG.

OHMAN, S. Coarticulation in VCV utterances: spectrographic measurements. In: *Journal of the Acoustical Society of America* 39, 1966. p.151-168.

PICKETT, J. *The acoustics of speech communication*. Boston: Allyn & Bacon, 1999.

SILVA, E. *Variação dialetal: as pretônicas no dialeto fluminense*. Comunicação apresentada no IX Encontro Anual da ANPOLL, 1994.

SILVA, M. *As pretônicas no falar baiano*. 1989. Tese (Doutorado) – UFRJ, Rio de Janeiro.

VIEGAS, M. *Alçamento das vogais pretônicas*. 1987. Dissertação (Mestrado em Língua Portuguesa) – FALE/UFMG, Belo Horizonte.

YACOVENCO, M. *As vogais médias pretônicas na fala culta carioca*. 1993. Dissertação (Mestrado) – UFRJ, Rio de Janeiro.