UMA PROPOSTA DE MODELAMENTO DA PERCEPÇÃO DA ENTOAÇÃO DO PORTUGUÊS BRASILEIRO

Beatriz RAPOSO DE MEDEIROS*
Marcus Vinícius Moreira MARTINS**

- RESUMO: O objetivo deste artigo é apresentar considerações sobre um modelamento lógico-matemático desenvolvido para abordar o fenômeno da percepção da entoação para o Português Brasileiro. O modelamento foi feito com base no modelo de análise automática da entoação, desenvolvido por Ferreira Netto (2006, 2008, 2010), e utiliza os princípios desenvolvidos por Hart, Collier e Cohen (1990), no que tange ao fenômeno da conjugação entre a percepção e produção das curvas entoacionais. Além disso, aplicamos, em nosso modelo, os limiares de diferenciação tonal, estipulados por Consoni (2011), valores os quais nos asseguram a relativização em um estado neutro e estados relevantes para a percepção, por meio da entoação para palavras isoladas, bem como para frases de contexto. Nosso modelamento resume-se à criação de um sistema conjugado em que os valores de F_0 são processados de acordo com uma componente denominada tom médio (FERREIRA NETTO, 2008). Os limiares de diferenciação tonal operam como limites sistêmicos de lateralidade gerados a partir do tom médio e têm o papel de restringir as variações e explicar possíveis variações modalizadas. O modelamento demanda mais estudos de base para uma melhor funcionalidade, além de uma implementação computacional funcional a fim de se verificar a sua aplicabilidade.

- PALAVRAS-CHAVE: Fonética. Entoação. Modelamento. Fonologia.

Introdução

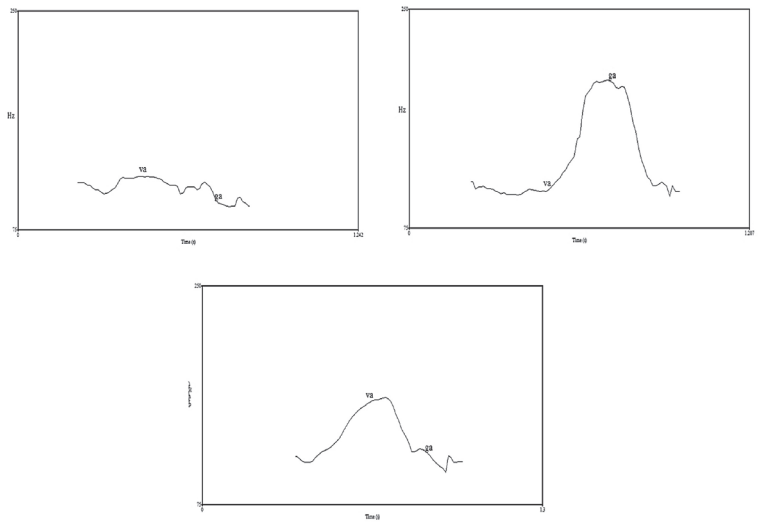
Desde a ascensão dos modelos autossegmentais em meados da década de 80, o estudo dos fenômenos ditos suprassegmentais tem ocupado espaço privilegiado dentro dos programas de pesquisa em Fonética e Fonologia (LADD, 1996; PIERREHUMBERT, 2000). Com os estudos da entoação não foi diferente e embora estes pertençam a uma área em ascensão no Brasil, já possuem discussões relevantes, principalmente no que concerne ao tratamento experimental do tema, como pontuam Maciel e Rothe-Neves (2006).

* USP – Universidade de São Paulo. Faculdade de Filosofia, Letras e Ciências Humanas. São Paulo – SP – Brasil. 05508-900 – biarm@usp.br

** USP – Universidade de São Paulo. Faculdade de Filosofia, Letras e Ciências Humanas. São Paulo – SP – Brasil. 05508-900 – marcusmartins@usp.br

O estudo da entoação tem se mostrado relativamente intricado, devido, principalmente, à diversidade de manifestações possíveis de curvas melódicas na fala. Essas curvas (ou contornos melódicos) constituem-se da produção e variação de frequência fundamental (F_0) de enunciados de fala, ao longo do tempo. A F_0 é, pois, um fenômeno acústico, grosso modo, resultante do fenômeno da vibração das pregas vocais, localizadas na laringe. A Figura 1 apresenta algumas das possíveis manifestações de F_0 para o item lexical /vaga/. Em Português Brasileiro, como se sabe, as escolhas melódicas não influenciam o significado do item lexical, como ocorre em Mandarim, por exemplo, em que estas variações de tom são elementos lexicalizados. Neste sentido, pensar em categorias para a entoação, como ocorre com o estudo dos segmentos, torna-se uma atividade complexa, uma vez que seria necessário imaginar um conjunto amplo de categorias, conforme as possíveis manifestações das curvas de F_0.

Figura 1 – Curvas de *pitch* para a palavra "vaga"



Fonte: Elaboração própria.

Destaca-se, também, a dificuldade em definir o termo "entoação" e restringi-lo a uma categoria única e restrita. O mesmo questionamento teórico foi visto e debatido por Vaissière (2004), que destaca que a dificuldade em se definir o termo está, também, em sua abrangência:

> *There is currently no universally accepted definition of intonation. The term may be strictly restricted to the perceived F_0 pattern, or include the*

> *perception of other prosodic parameters fulfilling the same functions: pauses, relative loudness, voice quality, duration, and segmental phenomena related to varying strengthening of the speech organs.* (VAISSIÈRE, 2004, p.238).

Segundo a autora, portanto, há duas possíveis abordagens para a entoação. Uma primeira, a qual podemos chamar de abordagem restrita, analisa apenas as variações de F_0 sem se ocupar de outras componentes, como o acento ou a duração, que são elementos geralmente considerados e estudados no âmbito da prosódia. A segunda seria uma abordagem ampla, comprometida em abarcar variados aspectos prosódicos, como os elementos rítmicos e, ainda, contemplar elementos como a qualidade de voz ou a intensidade do sinal de fala, sendo esses últimos tradicionalmente considerados paralinguísticos.

No presente trabalho, trataremos da entoação dentro da abordagem restrita, a qual consideramos mais pertinente ao tipo de análise por nós proposta. Assim, assumimos que a entoação é o conjunto de variações de F_0 no curso de uma fonação. Pretendemos, com esta definição, restringir nosso campo de abrangência e estabelecer um modelo lógico matemático simplificado, que explique a estrutura da entoação do Português Brasileiro, por meio de decomposições de séries temporais.

Questões em torno da entoação

Questões em torno da fonética e fonologia da entoação são muitas e na maioria das vezes esparsas, uma vez que não constituem um programa de pesquisa ou conjunto de estudos unívoco, ou são confundidas com outros aspectos, como o estudo do tom fonológico. De maneira geral, podemos dizer que há duas possíveis abordagens teórico-metodológicas para o estudo da entoação, sendo que ao invés de se oporem, elas se complementam, como propõe Pierrehumbert (2000). A raiz destas duas abordagens possíveis está na forma como se entende a entoação, ou seja, se devemos compreender a manifestação de F_0 em forma de níveis/sequência de tons. A primeira abordagem caracteriza a tradição americana fonêmica, a qual tem como principal nome Kenneth Lee Pike e seus posteriores desdobramentos guiados pela tese de doutoramento de Janet Pierrehumbert. A segunda abordagem tem, como principal autor, Dwight Bolinger e se filia de maneira mais direta à tradição britânica de estudos da linguagem. Além disso, difundiu-se de maneira mais categórica entre psicólogos e foneticistas, ligados aos estudos da percepção, sendo que muitos estudos guiados por esta tradição foram desenvolvidos pelos pesquisadores do IPO (Instiuut voor Perceptie Onderzoek, em holândes) ou Instituto para Pesquisa em Percepção (HART; COLLIER; COHEN, 1990).

A diferença entre as abordagens está na forma de se categorizar a entoação. Os estudos focados em níveis tonais têm como principal premissa a de que se pode imaginar camadas tonais distinguíveis entre si, como a oposição entre tons altos, médios e baixos, ao passo que o estudo das configurações tonais tende a abordar a questão das manifestações de F_0 ao longo de um intervalo de tempo. Vale salientar a diferença entre as duas abordagens: a primeira abordagem é mais comum em estudos fonológicos que buscam revelar distintividades no nível suprassegmental. Em outras palavras, tendem a discretizar um contínuo variável em categorias relativamente estáveis, pois ao se lidar com componentes como alturas há, necessariamente, de se pensar em camadas, no caso, entoacionais. A segunda abordagem encontra-se mais frequentemente em estudos psicológicos, por vislumbrar não a ocorrência de categorias dentro de um sinal variável, mas sim padrões de ocorrência, como no caso das inflexões causadas por emoções ou por aspectos pragmáticos.

Os autores mais favoráveis a uma abordagem por níveis adotaram uma fonologia da entoação (LADD, 1996; PIERREHUMBERT, 2000), que pressupõe a presença de categorias mínimas e regras de implementação fonética. Por esta razão, a abordagem por configurações tonais, aos poucos, foi sendo aceita por estudiosos focados no estudo da percepção (HART; COLLIER; COHEN, 1990). O tratamento dos dados depende da abordagem escolhida, isto é, a segunda abordagem, por questões metodológicas, passa a tratar seus dados com algum rigor estatístico (HART; COLLIER; COHEN, 1990; HART, 1981), enquanto a abordagem fonológica baseia suas análises em métodos impressionísticos auxiliados por algum instrumental básico (MAEDA, 1976). A seguir, faremos algumas considerações sobre estas questões, principalmente sobre os aspectos ligados à fonologia da entoação e os estudos de cunho perceptivo.

A questão da percepção da diferenciação nos estudos da entoação

A questão da mensuração de variação tonal tem suas raízes nos estudos de psicofísica e do sistema auditivo, os quais já eram debatidos no trabalho inicial de Helmholtz (1895). A questão dentro desses programas de pesquisas refere-se ao estudo das taxas de variação relativa, a partir de um ponto, para que a percepção seja modalizada. Em outras palavras, refere-se a que percentagem um parâmetro precisa variar, a partir de seu valor natural, para que ele seja interpretado como um novo parâmetro. Uma outra explicação ainda: trata-se de saber sua taxa de variação a partir do seu estado neutro. Estudos dessa natureza podem ter diversos propósitos, como a mensuração da concentração de sucralose em partes de água, necessária para se notar o sabor adocicado. Em acústica da fala, os estudos mais proeminentes foram os de Stevens, Volkman

e Newman (1937) e os que os seguiram, no intuito de se elaborar uma forma de mensurar e avaliar estas variações de F_0, com relação à intensidade, fossem elas locais ou globais.

A forma de mensurar estas mínimas variações tonais é por meio das *just noticeable differences* (jnd) (HART; COLLIER; COHEN, 1990; ROEDERER, 2002), que é a quantidade mínima de variação necessária para que ocorra uma variação notável na experiência sensória. Importa ressaltar que as jnd não se referem unicamente aos sons, mas são valores obtidos a partir das relações entre as magnitudes físicas do estímulo e as magnitudes percebidas pelo sujeito. Tais magnitudes são as bases da chamada "Lei de Weber", que relaciona duas grandezas, como demonstra a seguinte equação:

$$(1) \quad \frac{\Delta I}{I} = k$$

Onde, ΔI representa o limiar de diferenciação, ou a jnd, ao passo que I representa o valor do estímulo inicial e k é a constante que garante que a proporção entre os dois lados mantenha-se constante, independentemente do valor que I venha a receber.

Outra questão em torno do tema da percepção refere-se à existência ou não de categorias internalizadas para entoação. A priori podemos deduzir que, diferentemente do nível segmental, não haveria unidades categoriais mínimas, as quais seriam interpretadas em um nível representacional. Esse pressuposto guiou grande parte dos estudos entoacionais que estabeleceram como categorias mínimas os níveis acima expostos. Entretanto, a questão se põe quando observada a implementação fonética. A vibração das pregas vogais teria uma componente fonética duplamente funcional, ao se modular variações de F_0 em posições locais, como no caso das consoantes sonoras e das vogais, além de outras aplicadas no nível frasal, ou seja, capaz de diferenciar orações. Para Troubetzkoy (1976), a questão da dupla função da variação de F_0 era ponto pacífico, uma vez que os elementos por ele denominados "prosodemas" seriam ligados ao nível frasal. Para outros, como Maeda (1976), a implementação fonética da entoação seguiria uma tendência a valores cada vez mais baixos. Essa tendência é considerada uma componente nomeada por Maeda de *baseline*, ou valores de referência, de modo que as implementações locais estariam associadas a este valor de referência.

Hart, Collier e Cohen (1990) consideram que mais do que essa *baseline*, que assegura uma declinação paulatina dos valores de F_0, há outro componente, os movimentos de *pitch* (*pitch movements*), que, na interpretação dos autores, seriam "movimentos tonais" relativamente estáveis na língua e, assim, passíveis de serem considerados unidades de descrição da entoação. Em nosso entendimento,

esta abordagem teria como pressuposto básico a conjugação entre percepção e produção, uma vez que o módulo de processamento fonético estaria diretamente ligado ao módulo de processamento tonal. Esta junção de módulos seria capaz de produzir tons diferenciáveis no curso da fonação, conforme o desejo do falante. Contudo, é necessário ressaltar que estes tons respeitariam, de algum modo, certos limites, como o limite máximo de variação tonal e alguns limites fisiológicos, como a taxa de vibração das cordas vocais.

Além destes aspectos, há de se levantar aqueles ligados a propriedades próprias da percepção. A percepção não é uma propriedade humana una, pelo contrário, grande parte de sua manifestação está ligada a componentes tão complexos quanto ela própria, como é o caso da atenção e da memória, por exemplo (VERNON, 1974). A psicologia da Gestalt proposta por Wertheimer (1938) foi uma maneira de se tratar a forma como a percepção opera. De acordo com esta corrente do pensamento, o mundo estaria disposto em um *continuum*, de modo que os humanos dividem-no ou arranjam-no de forma que ele se torne inteligível. O autor ainda destaca duas maneiras de se compreender a percepção: uma, pela relação de proximidade, e outra, pela relação de similaridade. A primeira diz respeito ao fato de que dois elementos que estejam próximos um do outro sejam percebidos mais facilmente como um conjunto só, ao passo que objetos dispostos a uma distância maior sejam percebidos como distintos uns dos outros. A relação de similaridade afirma que dois objetos similares entre si serão percebidos como equivalentes, ao passo que a dessemelhança entre eles faz com que sejam percebidos como distintos.

Diante da ideia de elementos no interior de um sistema de comparação, postos em regiões mais próximas ou distantes, parece-nos que a questão da relativização dos tons pode se tornar mais clara se imaginarmos que há um tom de referência, de modo que os tons ao longo do enunciado a ele se assemelhem ou dele se diferenciem, dentro das limitações das jnd. A melhor alternativa dentro deste quadro é a de *perceptual magnetic effect*, desenvolvida por Patricia Kuhl (KUHL, 1991; KUHL; IVERSON, 1995; KUHL et al., 2006). De acordo com essa proposta, haveria uma forma prototípica fixa (no caso, os fones), não abstrata, que atuaria como parâmetro de comparação inicial para todas as demais formas que venham a ser percebidas. Nosso modelo leva em conta que há dois valores referenciais, os chamados limiares de diferenciação tonal (ou limites laterais), os quais seriam valores de referência para a variação de F_0, de modo que os demais valores estariam entre esses limites laterais ou para além deles. Os valores entre esses limiares seriam interpretados como invariáveis, ao passo que os que extrapolarem seriam notados como diferentes pelo ouvinte. Assim, uma construção com uma focalização, por meio da entoação, teria de ultrapassar esses limiares para que fosse notada como tal.

Séries temporais

Seguindo esta linha de raciocínio, pode-se definir a entoação como um fenômeno linguístico dotado de um mecanismo complexo, o qual envolve tanto a percepção dos sons, quanto o processamento da produção, devido ao fato de não haver referências/categorias internalizadas capazes de fornecer um *feedback* instantâneo dado um estímulo sonoro. A entoação também se caracteriza por ser uma manifestação suprassegmental, isto é, não estabelece uma relação unívoca entre segmentos e variações de F_0, embora seja possível, a nosso ver, controlá-las fisiologicamente, gerando frequências moduladas, em uma interação do fonético com o fonológico (XU; WANG, 1997; XU, 2005). Estas modulações de frequência, em última análise, seriam a "tessitura entoacional", isto é, a região de variação da frequência fundamental (F_0). Seguindo alguns destes pressupostos, Ferreira Netto (2008) compreende a entoação como sendo uma série temporal, na qual os elementos linguísticos que a compõem têm uma ligação temporal intrínseca.

Séries temporais são um dos diversos tipos de processos estocásticos, isto é, processos cujas variáveis aleatórias $(X: \Omega \rightarrow \mathrm{R})$ são indexadas por elementos t pertencentes a um intervalo temporal. Exemplos de processos estocásticos são as variações de inflação ao longo de uma década ou ano, ou mesmo as variações de temperatura em uma dada localidade. Um dos exemplos mais clássicos na área refere-se à evolução do preço médio do trigo do período que vai de 1500 a 1869 e é a base para diversos estudos até os dias de hoje. Neste sentido, processos estocásticos podem ser definidos como em Morettin (1999, p.27):

> **Definição 1** Seja T um conjunto arbitrário. Um processo estocástico é uma família $X(t), t \in \mathrm{T}$, tal que $t \in T$, $\mathrm{X}(t)$ é uma variável aleatória.

É importante ressaltar que este conjunto T é normalmente considerado como parte dos inteiros $\mathbb{Z}$, ou ainda dos reais $\mathbb{R}$. As vantagens de se compreender qualquer sistema de acumulação como uma série temporal são de se poder analisar sua variação ao longo do tempo passado, bem como projetar ou dimensionar sua variação esperada dado um intervalo de tempo posterior ao momento atual. Por esta razão, séries temporais são usuais em Economia, na área de Engenharia de Produção, bem como em áreas como a Climatologia. Em Ciências da Linguagem, o mesmo princípio pode ser assumido para as áreas de Fonética e Fonologia, uma vez que elas pressupõem uma ocorrência temporal de elementos, haja vista a aplicação de Cadeias de Markov (um tipo de processo estocástico) na área de processamento de *speech to text*, em que se é possível prever qual será a próximo elemento na cadeia, por exemplo, em Português Brasileiro é provável que após um [t ʃ] ocorra uma vogal [i] e não um [a] (JURAFSKY; MARTIN, 2009).

Nesse sentido, propomos o estabelecimento de um sistema linear que dê conta de um *input* sonoro retornando um *output* interpretável seja pela máquina, seja pelo homem. De antemão, assumimos que a interpretação destes dados fornecidos pelo *output* por uma máquina demandaria algum tipo de inteligência artificial ou rede neural especializada. O esquema apresentado na Figura 2 é uma maneira de se representar nossa exposição até o momento. Nas seções a seguir, serão expostos os detalhes e as equações que definem o sistema linear, apresentado na Figura 2.

Figura 2 – Esquema de análise da entoação do Português Brasileiro

Fonte: Adaptado de Chatfield (2004).

O modelo matemático: tempos e frequências

A entoação, pelo que foi exposto até agora, pode ser definida como as variações de F_0 em um intervalo de tempo, em que os elementos possuem uma relação temporal intrínseca, ou seja, como em uma série temporal (EHLERS, 2009; CHATFIELD, 2004; GREGSON, 1983). Nossa proposta é a de se entender a entoação como sendo uma série temporal de tempo discreto e frequência contínua. Nessa abordagem, teremos uma função $f(t)$ da frequência fundamental e iremos analisá-la em instantes de tempo equidistantes, como apresenta a equação (2).

(2) $f_t = f(t\Delta t), t = 0,1,2,3...$

Os valores de t podem ser tomados como janelas temporais. Boemio et al. (2005) determinam que a velocidade de resolução do córtex auditivo esquerdo (tido como o de melhor resolução temporal) é de cerca de 25 a 30 milissegundos, ao passo que a resolução temporal do córtex auditivo direito é em torno de 250 a 300 milissegundos. Ou seja, é necessário um *spam* temporal de ao menos 25 milissegundos para que o som seja processado pelo cérebro. Desta maneira, podemos assumir que mesmo o processamento cerebral dos sons é de tempo discreto, o que corrobora nossa opção por entender a entoação por este meio, uma vez que haverá sempre 25 milissegundos de atraso na análise neuronal dos sons, isto é, o processamento dos sons tem um atraso intrínseco. A entoação, portanto,

pode ser compreendida como a organização dos elementos entoacionais em uma série temporal θ, de modo que os elementos que a compõem são as medições de frequência Z ($z_1, z_2 \ldots z_{n-1}, z_n$, com $n - 1 > 0$, para $z_n = F_0(t)$), tomadas nos instantes t ($t_1, t_2 \ldots t_{n-1}, t_n$, onde $n\text{-}1 > 0$).

A equação (3) é a forma matemática do que foi acima descrito e retorna o conjunto de variações de frequência (z), dado um intervalo temporal que vá de t a t_n (de 0 ao infinito), estabelecendo uma relação de dependência entre os termos z e t:

(3) $\theta = z(t) = (z_1(t_1), z_2(t_2), \ldots, z_n(t_n))$

Onde, θ é o símbolo para série temporal e $z(t)$ é igual ao instante z em função do tempo (t). Em termos matemáticos, a equação (3) demonstra que conjunto das observações z estaria contido no eixo y de um plano cartesiano, ao passo que os instantes t estariam contidos no eixo x. Entender as curvas de F_0 como série temporal permite-nos um melhor refinamento analítico, uma vez que poderíamos eliminar, de nossa análise, a acelaração angular ω e passar a operar com um coeficiente angular α, de modo a tornar possível a comparação entre alturas em função do tempo. Neste caso não se analisariam ondas, mas sim funções lineares, em que os instantes seriam intrinsicamentes dependentes, isto é, o instante z_2 estaria ligado e dependeria do instante z_1. Em suma, analisaríamos retas, ao invés de se analisar curvas de entoação.

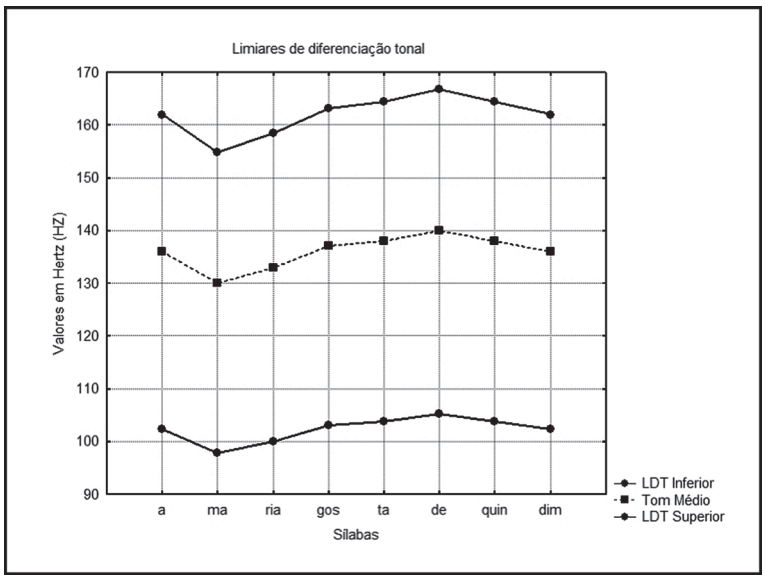
A vantagem em obter tais retas é a de trabalhar com um sistema autorreferencial, em que as informações numéricas estão simplificadas e as comparações podem ser feitas entre os próprios elementos que a constituem, sem que dependam de um sistema de padrões de curvas estabelecidos *a priori*, a serem comparados. Esta opção por se trabalhar – pelo menos por ora – com um sistema autorreferencial deve-se ao fato de que a entoação da fala é altamente variável, levando-nos a analisar com cautela as relações entre variação de frequência fundamental e outros componentes linguísticos como componentes sintáticos ou, até mesmo, como componentes pragmáticos.

Para melhor entender o que queremos dizer com "análise de retas", pensemos em um carro. Este carro está em uma velocidade variável dentro de uma pista circular, de diâmetro e raio indefinidos. O deslocamento linear do carro ao longo desta pista será sempre igual a x, o mesmo vale para o deslocamento angular, o qual será sempre 2π. Teremos, portanto, duas velocidades: uma linear e outra angular e outras duas acelerações. Ao fim da corrida, interessará a velocidade linear e não a angular, pois é ela quem dirá quão veloz foi nosso carro. Nossa proposta segue o mesmo princípio: em vez de analisarmos as variações angulares de uma curva, analisaremos o *step-by-step* da construção dela. Esta ideia, aparentemente díspar, dado que temos padrões já pré-estabelecidos para sentenças tipo em

várias línguas, auxilia-nos a tratar fenômenos de grande variação, como é o caso da entoação da fala.

A proposta deste modelo é a de que haveria intervalos laterais para as alturas, definidos com base em propriedades da percepção humana. Hart, Collier e Cohen (1990) nomeiam estes intervalos laterais de **limiares de diferenciação tonal.**[1] Hart (1981), com base em modelos experimentais, estipulou que estes valores seriam de 3 semitons para o holandês, isto é, para que se note qualquer diferença tonal no decurso de uma fonação no holandês é necessário que o F_0 varie ao menos 3 semitons para cima ou para baixo. Consoni et al. (2009), Consoni e Ferreira Netto (2008) e Consoni (2011), ao replicarem os experimentos de Hart, encontraram os valores de +3st e −4st para o Português Brasileiro, estabelecendo, assim, uma faixa de banda de 7st, como demonstrado no Gráfico 1.

Gráfico 1 – Esquema de análise da entoação do Português Brasileiro



Fonte: Elaboração própria.

Para se estipular esses limites laterais, Ferreira Netto (2006, 2008) propõe uma componente cognitivo-funcional chamada **tom médio** (τ), melhor definida na subseção a seguir. No Gráfico 1 pode-se ver os limiares de diferenciação tonal inferior e superior (as linhas sólidas) gerados a partir de um tom médio (a linha tracejada) obtido na análise da oração "A Maria gosta de quindim". Os valores de

[1] Aqui, por vezes, referiremo-nos a estes limiares pela sigla LDT (Limiares de Diferenciação Tonal).

frequência, apresentados no gráfico, em cada sílaba foram estabelecidos pela soma das frequências válidas (maiores que 50Hz e menores que 350Hz) encontradas a cada 25ms dividido pela duração total da sílaba dividida por 25ms, obtendo-se assim uma média. A equação (4) representa como foi feito o cálculo.

(4) $$F_m = \frac{\sum_{i=1}^{n} F_0}{\Delta_t/0{,}025}$$

Onde, Fm é a frequência média (em Hz) de uma sílaba, $\Sigma_{i=1}^{n} F_0$ é a soma de todas as frequências válidas na sílaba e Δ_t é a duração total da sílaba.

O tom médio

De acordo com o modelo de Ferreira Netto (2008), o falante, durante o processo inicial de fonação, estabeleceria um valor de referência, de modo que durante o curso de fala ele teria a tendência a se esforçar para manter os valores de frequência em torno desta referência. Desta maneira, o tom médio seria uma *baseline* de referência para se estipular o intervalo lateral de a para $b(a \rightarrow b)$, os quais seriam os limiares de diferenciação tonal, contidos no eixo y da série temporal θ, como representado no Gráfico 1 pelas linhas sólidas. A equação (6) define a forma matemática do tom médio, ao passo que o sistema (7) define como seriam estabelecidos os valores de a e b. Importante ressaltar que a equação (6) é uma média aritmética instante a instante, isto é, para t_2, por exemplo, temos $t_2 = \frac{t_1+t_2}{2}$, em que a média em t_2 é a soma das médias dos instantes anteriores, no caso t_1 e t_2. Assim:

Para a série temporal θ:

(5) $$z(t) = (z_1(t_1), z_2(t_2), \ldots, z_n(t_n))$$

Temos o tom médio (τ) definido pelas médias da série temporal θ:

(6) $$\tau = \sum_{i=1}^{n} z_n(t_n)$$

Onde, τ é o tom médio e $\sum_{i=1}^{n} z_n(t_n)$ é a média instante a instante da série temporal. A partir dela podemos pressupor o seguinte sistema, no qual a e b são os extremos do intervalo de frequência, os limiares de diferenciação tonal:[2]

(7) $$p/(a,b) = \begin{cases} a = \tau * 1{,}191 \\ b = \tau - (\tau * 0{,}248) \end{cases}$$

[2] O valor 1,191 refere-se a 3 semitons e o valor 0,248 refere-se a 4 semitons.

Onde, a e b são os limiares de diferenciação tonal superior e inferior expressos em termos de semitons.

O modelo de Ferreira Netto (2006, 2008) supõe uma terceira componente, além do intervalo lateral e do tom médio, a saber: a finalização. De acordo com Peres, Consoni e Ferreira Netto (2011), o fecho de um ato de fala seria notado no instante em que a frequência fundamental chegasse ao valor de –7st de distância do tom médio estabelecido. Seguindo este princípio seria necessária a inclusão de um terceiro elemento no sistema (7), como demonstrado abaixo.

$$(8)\quad p/(a \to b, c) = \begin{cases} a = \tau * 1{,}191 & \text{estabelece o LDT superior.} \\ b = \tau - (\tau * 0{,}248) & \text{estabelece o LDT inferior.} \\ c = \tau - (\tau * 0{,}35) & \text{estabelece o valor de finalização.} \end{cases}$$

Onde, c é o valor de finalização expresso em termos de semitons.

Por fim, podemos incluir neste sistema, com vistas a futura computação e representação gráfica, mais dois elementos, definindo o sistema $\varepsilon(t)$,[3] para todo $F_0 > 0$:

$$(9)$$
$$= \epsilon(t) \begin{cases} a = \tau * 1,19 & \text{estabelece o LDT superior.} \\ F_0 = \lvert z(t) & \text{variação da frequência fundamental.} \\ \tau(t) & \text{tom médio.} \\ b = \tau - (\tau * 0,22) & \text{estabelece LDT inferior.} \\ c = \tau - (\tau * 0,35) & \text{estabelece o valor de finalização.} \end{cases}$$

O sistema de equações (9) é a forma final do sistema linear apresentado na Figura 2. Sua principal característica é a de decompor um sinal de F_0 em três diferentes componentes: os LDTs, a finalização e o tom médio. A principal componente do sistema é a equação (6), a qual define as demais. Trata-se, em parte, da decomposição de uma série temporal. O elemento $F_0 = \mathrm{z}(t)$ garante que a variação de F_0 será levada em consideração durante a decomposição, ao passo que transforma a variação ω de F_0 em uma variação α, isto é, garante que a variação ondulatória seja desconsiderada, levando em conta pontos de variação da frequência fundamental. Trata-se de uma das vantagens de nosso modelo, no que se refere à comparação de curvas, como feito em Martins e Ferreira Netto (2010, 2011), em que se procurou detectar qual a melhor forma de

[3] Usaremos a letra grega ε para indicar "Entoação".

se medir variações de frequência fundamental. Ressaltamos que os valores aqui apresentados e a computação do modelo podem ser verificados com a rotina ExProsodia, desenvolvida por (FERREIRA NETTO, 2010). A seguir apresentaremos uma análise usando o modelo aqui descrito.

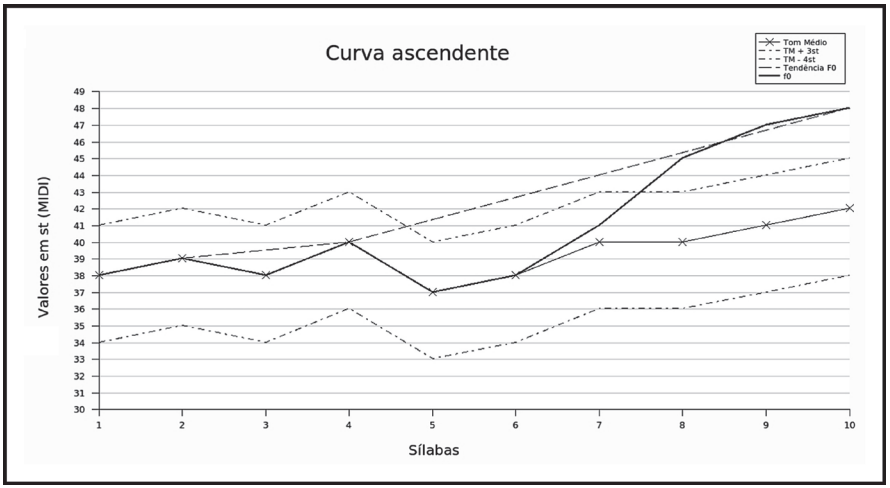
Exemplos de análises

Empregando o sistema linear acima descrito, propomos uma análise entoacional do enunciado "O Domenico entrar aqui... como assim?", retirado de uma cena de filme. Nossa escolha foi motivada pelo fato de o enunciado-alvo ser um manifestação ensaiada, isto é, não espontânea.[4] Em uma fala ensaiada ou preparada, que é o caso de filmes de ficção, as modalidades de sentenças devem ser cumpridas conforme um texto previamente dado. Neste caso, uma pergunta precisa soar canonicamente como tal. Apoiamo-nos, portanto, nesta característica da fala interpretativa, para assegurarmos que a intenção do falante, bem como sua realização do enunciado são as de uma pergunta e de nenhuma outra modalidade de sentença. Isso nos permite analisar algumas componentes, como o necessário aumento de F_0 ao final da sentença, o que determina que se trata de uma sentença interrogativa. Dentro do modelo ToBI podemos dizer que se trata de uma H-H%, ou seja, um tom alto, seguido de um tom mais alto ao fim.

O Gráfico 2 apresenta uma análise para a oração acima apresentada, utilizando-se do algoritmo acima descrito. Note-se que a linha central marcada com "x" refere-se ao tom médio. É notável que quase não ocorram variações sistemáticas do valor do tom médio, indicando que ele possui uma relativa estabilidade. Observa-se também que a curva de F_0, toda em preto, somente ultrapassará as margens –3 e +4, ou seja, os LDTs, ao fim da oração, quando há um aumento paulatino da frequência fundamental com a intenção de se marcar que se trata de uma interrogativa. Os valores de frequência, no caso analisado, aparecem no eixo das coordenadas e foram tomadas as médias de frequência de cada sílaba, sendo estes valores em seguida convertidos para semitons.

[4] Uma das maneiras de garantir, de certa forma, a entoação do enunciado como típica de uma interrogativa seria por meio do julgamento por um grupo de sujeitos ouvintes. No entanto, consideramos que retirar uma frase de filme seria um modo próximo da situação da "fala de laboratório", em que o que há é uma fala preparada e, portanto, "julgada" a priori. Em defesa da fala de laboratório temos Xu (2010), que trata de diversos mitos relativos a esta fala. Um deles diz respeito ao suposto empobrecimento da prosódia, o qual Xu (2010) rebate, dizendo que são necessários métodos adequados para se elicitar determinados padrões prosódicos. Certamente, assumimos que a escolha do enunciado interrogativo, tal como foi feita, contorna a questão da verificação do padrão entoacional da sentença e entendemos que o julgamento deve ser feito, futuramente, quando da sequência do desenvolvimento do modelo ora apresentado.

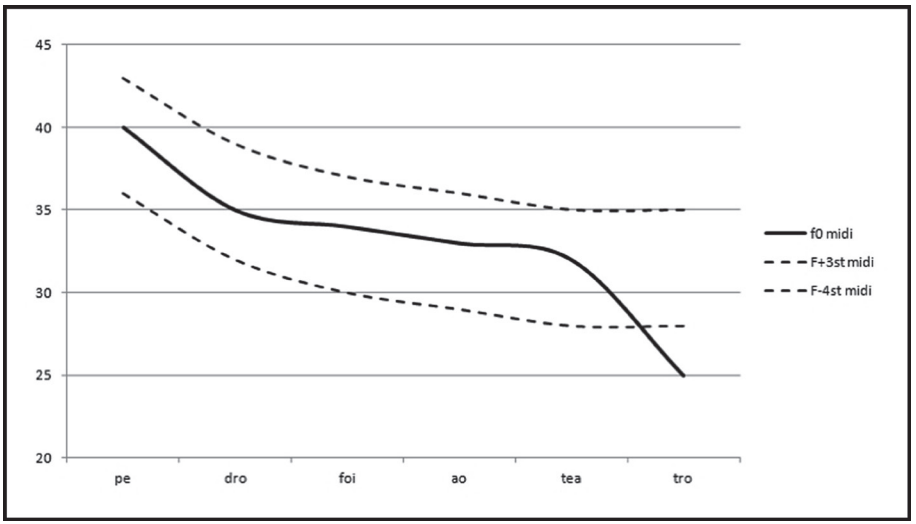
Gráfico 2 – Exemplo de análise se usando o tom médio para a oração "O Domenico entrar aqui... como assim?"



Fonte: Elaboração própria.

A título de comparação, podemos notar que no Gráfico 3 há um decrescimento paulatino de F_0 até que se atinja um valor ainda mais baixo, a chamada declinação frasal.

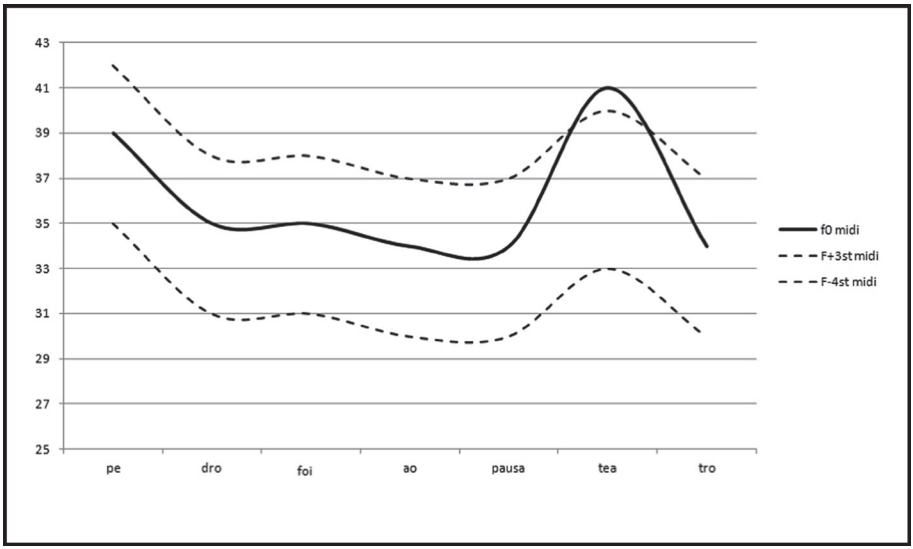
Gráfico 3 – Análise para oração "PEdro foi ao teatro"



Fonte: Baseado em Cagliari (1983).

Além disso, pode-se notar uma elevação de F_0 na sílaba (pe), indicando uma possível focalização do item (pe. drU), ao passo que o Gráfico 4 tem como item focalizado a palavra (te. a.trU), notando uma elevação de F_0 na sílaba (te. a).

Gráfico 4 – Análise para oração "Pedro foi ao TEatro"



Fonte: Baseado em Cagliari (1983).

Conclusão

O objetivo principal do presente trabalho foi o de fazer algumas considerações em torno de um possível modelamento da percepção da entoação, considerando que este fenômeno da fala requer limiares de altura previsíveis e mensuráveis. Não levantamos pois – e isto foi proposital – uma questão linguística propriamente dita, como questões de alinhamento tonal ou curvas entoacionais, por exemplo. O sistema linear apresentado como produto final representa antes um algoritmo que pode conjugar percepção e produção da entoção da fala.

Podemos, no entanto, vislumbrar algumas aplicações, como a detecção de foco em uma oração, por exemplo. Se a subtração de um dos limiares (a, b) a um instante $Z(t)$ for maior ou igual ou menor ou igual a zero, podemos dizer que, naquele ponto, ocorre uma focalização, como no exemplo apresentado no Gráfico 4. As equações (10) e (11) são as representações matemáticas do que foi dito:

(10) $Z(t) - a \geq 0 \cong E^{+}$

(11) $Z(t) - b \leq 0 \geq Z(t) - c \cong E^{-}$

Onde, E^{+} e E^{-} são respectivamente foco positivo, isto é, com um acréscimo de F_0, e foco negativo, com um decrescimento de F_0, sem que se atinja o valor de finalização c. Além disso, podemos imaginar outros conjuntos de equações, usando cálculos estatísticos, como a covariância de uma série temporal, aplicada para a interpretação de emoções. A covariância, neste caso, poderia relacionar, de um lado a emoção veiculada e de outro a variação de F_0 ao longo do tempo. A variação pode ser comparada à variação de uma fala neutra e/ou ao tom médio desta última.

Obviamente, a matematização baseada no fenômeno da percepção da entoação demanda estudos mais categóricos de aspectos propriamente linguísticos e fisiológicos da produção da entoação. Nossa intenção foi a de desenvolver um sistema que decompusesse um sinal de F_0 em componentes menores e passíveis de serem analisadas independentemente, embora essa análise demande ainda alguma interferência humana, no que se refere ao julgamento. Ou seja, sabemos que características são detectadas pela interpretação humana impressionística, o que torna difícil objetivar absolutamente os dados de fala que envolvem emoção e/ou atitude.

Para estudos futuros, que deem continuidade à proposta apresentada, pretendemos desenvolver o modelo de decomposição aqui apresentado, utilizando-se de ferramentas matemáticas mais elaboradas, as quais forneçam uma análise componencial conjugada a mecanismos de detecção de propriedades linguísticas mais acurados.

RAPOSO DE MEDEIROS, B.; MARTINS, M. V. M. A modeling proposal for the intonation perception of Brazilian Portuguese speech. *Alfa,* São Paulo, v.58, n.1, p.195-213, 2014.

- *ABSTRACT: The main purpose of this paper is to present a logical-mathematical modeling to the phenomenon of intonation perception in Brazilian Portuguese speech. This modeling is based on the standard automatic analysis of intonation developed by Ferreira Netto (2006, 2008, 2010); and it employs the principles developed by Hart, Collier and Cohen (1990), regarding the phenomenon of conjugation between perception and production of intonational curves. Furthermore, we apply thresholds of tonal differentiation, which are prescribed by Consoni (2011) to sing le words and to contextualized sentences. Threshold values ensure the perception of relativization regarding the intonation. Our modeling basically aims to create a conjugated system in which the values of F0 are processed according to a component called the mid-tone (FERREIRA NETTO, 2008). The thresholds of tonal differentiation operate as limits to the systematical laterality generated by the mid-tone; and they act as a restriction*

to variations. The modeling requires more studies for its better functionality, as well as a functional computer implementation in order to verify its applicability.

▪ *KEYWORDS: Phonetics. Intonation. Modeling. Phonology.*

REFERÊNCIAS

BOEMIO, A. et al. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, Londres, v.8, n.3, p.389–395, 2005.

CAGLIARI, L. C. *Elementos de fonética do português brasileiro*. São Paulo: Paulistana, 1983.

CHATFIELD, C. *The analysis of Time Series*. 6.ed. New York: Chapmann & Hall, 2004.

CONSONI, F. *Aspectos da percepção da proeminência tonal em português brasileiro*. 2011. 129f. Tese (Doutorado em Linguística) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011.

CONSONI, F.; FERREIRA NETTO, W. A percepção de variação em semitons ascendentes em palavras isoladas no português brasileiro. In: CONGRESSO NACIONAL DE FONÉTICA E FONOLOGIA, 10.; CONGRESSO INTERNACIONAL DE FONÉTICA E FONOLOGIA, 4., 2008, Niterói. *Anais...* Niterói: UFF, 2008. Disponível em: < http://www.academia.edu/4909708/A_percepcao_de_variacao_em_semitons_ascendentes_para_falantes_de_Portugues_Brasileiro_em_palavras_isoladas>. Acesso em: 28 jun. 2009.

CONSONI, F. et al**. The brazilian speaker sensibility of perception in rising semitones variation for isolated words**. 2009. Trabalho apresentado ao 40. Poznan Linguistic Meeting, Poznan.

EHLERS, R. *Análise de séries temporais:* Relatório Técnico. São Paulo: USP, 2009. Disponível em: <http://www.icmc.usp.br/ ehlers/stemp/stemp.pdf.>. Acesso em: 07 maio 2007.

FERREIRA NETTO, W. ExProsodia. processo nº08992-2. *Revista da Propriedade Industrial*, Brasília, v.2038, p.167, 2010.

_____. Decomposição da entoação frasal em componentes estruturadoras e em componentes semântico-funcionais. In: CONGRESSO NACIONAL DE FONÉTICA E FONOLOGIA, 10.; CONGRESSO INTERNACIONAL DE FONÉTICA E FONOLOGIA, 4., 2008, Niterói. *Anais...* Niterói: UFF, 2008. Disponível em: <http://www.academia.edu/4909737/DECOMPOSICAO_DA_ENTOACAO_FRASAL_EM_COMPONENTES_ESTRUTURAIS_E_SEMANTICO-FUNCIONAIS_UM_TESTE_COM_ANALISE_DA_VARIACAO_DE_GENERO>. Acesso em: 28 jun. 2009.

_____. *Variação de frequência e constituição da prosódia da língua portuguesa*. 2006. 89f. Tese (Livre-docência) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2006.

GREGSON, R. A. M. *Time Series in Psychology*. 1.ed. Hillsdale: Lawrence Erlbaum Associates, 1983.

HART, J. t'. Differential sensitivity to pitch distance, particularly in speech. *Journal of Acoustical Society of America*, New York, v.3, n.69, p.811–821, 1981.

HART, J. t'; COLLIER, R.; COHEN, A. *A perceptual study of intonation*. Cambridge: Cambridge University Press, 1990.

HELMHOLTZ, H. L. F. *On the sensations of tone as a physiological basis for the theory of music*. 1.ed. London: Longmans, Green, 1985.

JURAFSKY, D.; MARTIN, J. H. *Speech and language processing:* an introduction to natural language processing, computational linguistics, and speech recognition. 1.ed. New Jersey: Prentice Hall, 2009.

KUHL, P. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics,* Austin, v.50, n.2, p.93–107, 1991.

KUHL, P.; IVERSON, P. Linguistic experience and the perceptual magnet effect. In: STRANGE, W. (Org.). *Speech perception and linguistic experience*: issues in cross-language research. Baltimore: York Press, 1995. p.121–154.

KUHL, P. et al. *Língua, cultura, mente, cérebro:* progresso nas fronteiras entre disciplinas. Tradução de Waldemar Ferreira Netto e Fernanda Consoni. São Paulo: Paulistana, 2006.

LADD, D. R. *Intonational Phonology*. Cambridge: Cambridge University Press, 1996.

MACIEL, F.; ROTHE-NEVES, R. Investigações experimentais da entonação no português brasileiro: revisão de literatura. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE FONÉTICA, 9.; CONGRESSO INTERNACIONAL DE FONÉTICA E FONOLOGIA, 3., 2006, Belo Horizonte. *Anais...* Belo Horizonte, 2006. v.1, p.1-48.

MAEDA, S. A. *Characterization of American English Intonation*. 1976. 332f. Tese (Doutorado em Emgenharia Elétrica) - Massachussetts Institute of Technology (MIT), Cambridge, 1976.

MARTINS, M. V. M.; FERREIRA NETTO, W. Speech intonation and perception: a study of frequency scales for brazilian portuguese. *Journal of Acoustical Society of America*, New York, n.129, v.4, p.2657, 2011.

_____. Prosódia e escalas de frequência: um estudo em torno da escala de semitons. *ReVEL:* Revista Virtual de Estudos da Linguagem, [S.l], v.15, n.8, p.286–296, 2010.

MORETTIN, P. A. *Ondas e Ondaletas:* da análise de Fourier à análise de Ondaletas. São Paulo: EDUSP, 1999.

PERES, D. O.; CONSONI, F.; FERREIRA NETTO, W. A influência da cadeia segmental na percepção de variações tonais. *LL Journal*, Nova York, v.6, p.3, 2011.

PIERREHUMBERT, J. Tonal elements and their alignment. In: HORNE, M. (Org.). *Prosody, theory and experiment*: studies presented to Gösta Bruce. Dodrecht: Kluwer Academic Publisher, 2000. p.11–36.

ROEDERER, J. G. *Introdução à física e psicofísica da música*. Tradução de Alberto Luís da Cunha. São Paulo: EDUSP, 2002.

STEVENS, S. S.; VOLKMAN; J.; NEWMAN, E. A scale for the measurement of the psychological magnitude of pitch. *Journal of the Acoustical Society of America*, New York, v.3, n.8, p.185–190, 1937.

TROUBETZKOY, N. *Principes de Phonologie*. Tradução de Jean Cantineau. Paris: Klincksieck, 1976.

VAISSIÈRE, J. Perception of intonation. In: PISONI, D. B.; REMEZ, R. E. (Org.). *Handbook of speech perception*. Oxford: Blackwell, 2004. p.236-263.

VERNON, M. *Percepção e experiência*. Tradução de Dante Moreira Leite. São Paulo: Perspectiva, 1974.

WERTHEIMER, M. Laws of organization in perceptual forms. In: ELLIS, W. D. (Org.). *A source book of Gestalt Psychology*. London: Routledge & Kegan Paul, 1938. p.71-88.

XU, Y. In defense of lab speech. *Journal of phonetics*, Londres, n.38, p.329-336, 2010.

_____. Speech melody as articulatorily implemented communicative functions. *Speech Communication*, Amsterdan, n.46, p.220–251, 2005.

XU, Y.; WANG, Q. E. What can tone studies tell us about intonation? In: BOTINIS, A.; KOUROUPETROGLOU, G.; CARAYANNIS, G. (Org.). ESCA WORKSHOP, 1997, Atenas. *Intonation*: theory, models and applications. Atenas: European Speech Comunnication Association, 1997. p.337-340.

Recebido em setembro de 2012.

Aprovado em janeiro de 2013.