

# A Percepção Musical da Fala e o Problema do Alinhamento da Curva de Frequência Fundamental

André Ricardo de Souza  
Faculdade de Artes do Paraná  
UNESPAR  
Curitiba - PR  
arsdicendi@ig.com.br

**Resumo**—Este trabalho apresenta alguns resultados e conclusões de uma abordagem musical do problema da entoação, mais especificamente da relação entre a curva de frequência fundamental e a cadeia segmental do enunciado. Com o apoio de alguns resultados das investigações sobre a percepção e a cognição da altura na fala e na música, o texto analisa a questão comparando o modelo de aproximação ao alvo (*target approximation*) de Yi Xu ao fenômeno do portamento em música, e levantando outras questões relacionadas com o papel da estabilidade da altura ao longo da sílaba no reconhecimento de diferentes marcas prosódicas.

**Palavras-chave:** entoação; percepção; cognição; altura; música; fala

## I. INTRODUÇÃO

Apesar de nos referirmos com frequência à entoação como sendo “a melodia da fala”, é praticamente impossível confundirmos um enunciado cantado com um enunciado falado. Tampouco é possível, malgrado as inúmeras tentativas já feitas, representar de maneira satisfatória a entoação de um enunciado por meio de notação musical convencional. De todas as características que distinguem a fala do canto, certamente a mais notável é o fato de que quando cantamos uma melodia com texto a frequência fundamental se mantém razoavelmente constante ao longo de cada sílaba, ao passo que na fala isso não acontece.

No contexto dos estudos prosódicos, há uma série de evidências que indicam que o alinhamento da curva de frequência fundamental com a cadeia segmental é um aspecto significativo na formação do sentido que a entoação atribui ao enunciado [1]. A questão do alinhamento envolve o problema de se determinar que pontos da cadeia são marcos temporais da entoação e qual o papel da sílaba nesta relação. Colocado em outros termos, o problema poderia ser descrito da seguinte forma: podemos associar uma altura ou um contorno melódico a uma sílaba, ou o posicionamento da curva de frequência no interior da sílaba também é significativo?

Esta questão está indissociavelmente ligada a outra discussão fundamental no estudo da entoação, que é o que Anthony Fox [2] chama de controvérsia “patamares versus

configurações”, ou seja, a questão de se determinar se os elementos distintivos da entoação são pontos correspondentes a patamares da frequência fundamental ou são contornos característicos da curva de  $f_0$ . O presente trabalho pretende trazer contribuições do conhecimento musical para estas indagações e faz parte de uma pesquisa que busca investigar as semelhanças e diferenças entre fala e canto, visando a interpretação da música vocal e do texto dramático, bem como suscitar discussões sobre as particularidades da linguagem oral.

## II. PERCEPÇÃO DA ALTURA: CONTRIBUIÇÕES DAS NEUROCIÊNCIAS

Nos últimos vinte anos, estudos das neurociências apoiados em técnicas de imagem computadorizada do cérebro, tais como os exames de tomografia e ressonância magnética, têm trazido novas ideias a respeito da especialização lateral dos processos cognitivos. A tradicional concepção de que o hemisfério esquerdo (HE) responde pelos processos envolvidos no uso da linguagem e o hemisfério direito (HD) pelas faculdades musicais deu lugar à concepção modular dos sistemas responsáveis pelos processos envolvidos na atividade linguística e na musical, distribuídos nos dois hemisférios. Zatorre e colegas mostram, a partir dos inúmeras evidências reunidas, que o hemisfério direito tem, de fato, uma capacidade superior de processamento da informação relativa à altura, mas que este processamento pode tanto atender a necessidades linguísticas com musicais [3]. Esta especialização dos hemisférios cerebrais quanto ao processamento da informação proveniente do nervo auditivo é decorrente do que os autores chamam de “princípio da incerteza da percepção auditiva”: para obter maior precisão na determinação da altura, o circuito responsável deve trabalhar com intervalos de tempos maiores e, conseqüentemente, perde precisão no domínio temporal; por outro lado, uma maior precisão na discriminação de eventos muito breves acaba limitando a precisão na identificação de alturas e intervalos, e na percepção do espectro de maneira geral. Assim, os dois hemisférios estabelecem uma divisão de tarefas, ficando o HD especializado na percepção de alturas e intervalos e o HE como responsável pela discriminação de eventos muito rápidos tais como os transientes que caracterizam consoantes oclusivas, *glides*, e até mesmo variações bruscas de altura no interior da sílaba.

A comparação entre resultados de investigações da atividade cerebral em casos de amusia (incapacidade de identificar e memorizar melodias) e afasia mostram relações interessantes com respeito à lateralização de processos cerebrais. Isabelle Peretz relata, a partir de sua investigação sobre a amusia, que pacientes com lesões bilaterais do cérebro são em geral apresentam amusia mas não afasia, enquanto que afásicos não-amúsicos apresentam em geral lesões somente no HE [4]. Por outro lado, observações feitas em casos de amusia causada por lesões somente no hemisfério direito indicam que nestas situações as faculdades linguísticas são mantidas, embora haja indícios de que a discriminação de nuances prosódicas é prejudicada em sujeitos com amusia congênita.

O treinamento musical também influi na distribuição do processamento da altura entre os hemisférios, conforme mostram estudos que empregam medidas do fluxo sanguíneo no cérebro, indicando que a percepção de altura é acentuadamente lateralizada à direita em não músicos, mas não em músicos [5]. Em um estudo mais recente, baseado em testes perceptivos, evidências apontam para o possível compartilhamento de algumas funções básicas que atendem a cognição musical e a linguística, dentre as quais podemos supor que esteja o processamento da altura [6].

Outros resultados relevantes quanto à percepção de altura na linguagem são os resultados de experiências comparando atividade cerebral de falantes de línguas tonais e não-tonais [7]. De maneira geral pode-se dizer que falantes de línguas não tonais apresentam atividade cerebral maior no HD ao ouvirem palavras de uma língua não tonal, enquanto que, nas mesmas condições, falantes de línguas não-tonais apresentam atividade cerebral maior no HE. A consequência desta constatação seria hipótese de que a percepção do contorno melódico fora do contexto linguístico é uma tarefa executada no HD, o que poderia explicar a aparente contradição com os resultados que apontam para a localização do processamento da percepção de contorno no HD. A participação significativa HE provavelmente está relacionada com o reconhecimento de traços distintivos dos tonemas. Diana Deutsch e colegas apontam relações entre o fenômeno do ouvido absoluto e a recorrência das alturas nos tons do mandarim [8]. Segundo os autores, o desenvolvimento do ouvido absoluto depende de associações verbais com determinadas alturas estabelecidas nos estágios iniciais de desenvolvimento do indivíduo e, por este motivo, há uma maior probabilidade desta habilidade particular ser encontrada em falantes de línguas tonais.

Uma primeira consequência da especialização hemisférica da percepção temporal e espectral é a conclusão de que a discriminação de um contorno melódico e a identificação dos intervalos são processadas independentemente, em locais diferentes do córtex auditivo. A identificação de intervalos exige uma maior precisão na definição da frequência fundamental e, conseqüentemente, emprega uma janela temporal maior, enquanto que para a percepção da forma de um contorno melódico não é necessário levar em conta frequências exatas. Além disso, podemos supor que a discriminação de um contorno melódico exige mais do processamento temporal que do processamento espectral.

A partir das evidências acima listadas, podemos propor a hipótese de que haja um sistema especializado em identificar alturas e variações de altura com grande precisão, com os circuitos localizados predominantemente no HD, e que é acionado tanto na percepção da fala como da música, enquanto que o processamento da hierarquia entre as alturas (intervalos, classes de altura, escalas, etc.) envolve um outro sistema, com circuitos provavelmente localizados no HE (embora possa lançar mão de informações do processamento espectral, provenientes, portanto, do HD), e que não é acionado quando da percepção da fala. Esta linha de raciocínio explicaria porque quando direcionamos nossa percepção musical para identificar os intervalos da melodia da entoação de um enunciado, perdemos a informação entoacional propriamente dita, e vice-versa. Por algum motivo, ainda a ser descoberto, os dois circuitos (o que processa melodia e o que processa entoação) são mutuamente exclusivos, pelo menos em algum de seus componentes.

Considerando-se uma possível validação desta hipótese, seria bastante improvável que as alturas empregadas na formação da entoação de um enunciado fossem baseadas em um sistema de referência como o da música ocidental, formado por classes de altura (que correspondem ao nomes das notas independentemente da oitava em que se encontram), ou mesmo de intervalos harmônicos, apesar de alguns estudos apontarem evidências de que isto ocorre sistematicamente na fala [9]. O mais provável é que haja uma parte do processamento de altura que seja exclusiva da fala, que pode até ter uma precisão menor que um semitom, mas que não faça associações por classes de altura.

### III. ALVOS PERCEPTUAIS

Buscando uma solução que conciliasse os dois pontos de vista da controvérsia “patamares *versus* configurações”, Yi Xu propõe um modelo para a entoação, baseado no conceito de “aproximação ao alvo” (target approximation, em inglês), segundo o qual o aparelho vocal busca, ao longo do enunciado, ajustar a frequência fundamental a níveis específicos (os “alvos”), mas seu comportamento dinâmico e suas restrições físicas impedem uma mudança imediata, discreta, desta frequência [10]. Assim, as configurações observadas na curva de frequência fundamental são o resultado do empenho do aparelho vocal em atingir esses alvos, que podem ser estáticos, quando a curva se dirige a um patamar, ou então dinâmicos, quando a curva tende a uma determinada inclinação. Xu argumenta também que o modelo também apresenta solução para o problema do alinhamento, uma vez que os marcos temporais que atrelam a curva de  $f_0$  à cadeia segmental são os instantes em que a curva começa a se dirigir ao próximo alvo, os quais, segundo os resultados apresentados pelo autor, estão sistematicamente ligados a pontos específicos da sílaba.

A noção de alvo estático assim proposta se assemelha muito ao que em música se chama portamento: uma nota que se inicia com uma altura ligeiramente mais baixa (ou, mais raramente, mais alta) que a esperada e aos poucos se aproxima da altura “real”. O ouvinte, neste caso, não interpreta isso como uma desafinação, mas sim como um efeito expressivo, e entende como altura real aquela à qual o intérprete se dirige (o alvo). A semelhança do fenômeno com o caso do portamento na música

parece reforçar a hipótese da importância perceptual dos patamares para onde a curva de frequência se dirige no alinhamento da curva de frequência com a cadeia segmental, tendo a sílaba como unidade básica de percepção da altura.

Esta semelhança entre o portamento na música e a aproximação ao alvo na fala também lança alguma luz sobre as iniciativas que músicos ocidentais têm realizado no sentido de registrar em linguagem musical (escrita ou executada em um instrumento) a “melodia” da entoação. O compositor Leos Janacek, por exemplo, chegou a reunir mais de 5.000 exemplos de entoação do tcheco e de outros idiomas, anotados por ele na pauta musical [11].

Mais recentemente, o músico brasileiro Hermeto Pascoal desenvolveu uma técnica de composição baseada na mesma ideia, que ele chama de “som da aura” [12]. O processo é simples: a partir de gravações de voz falada, Hermeto grava com um teclado (sintetizador) uma melodia que corresponde, sílaba por sílaba, à entoação do enunciado, e em seguida grava um acompanhamento harmônico. O primeiro registro dessa ideia aparece no álbum “Lagoa da Canoa Município de Arapiraca”, de 1984, empregando gravações de locutores esportivos; mais tarde, aparece de forma mais elaborada no álbum “Festa dos Deuses” [13], com exemplos de leitura de poesia, discurso político e fala espontânea; também foi realizada em outras situações tais como entrevistas em televisão e documentários, inclusive com enunciados em outras línguas.

O interesse despertado por estas gravações reside no fato de que Hermeto resolve de uma maneira peculiar, porém aparentemente sistemática, o problema da atribuição de uma altura a uma sílaba. Entendemos que estas gravações oferecem uma situação privilegiada para estudarmos as relações entre os modos de escuta de fala e música. Várias questões podem ser levantadas para serem investigadas com este material, como, por exemplo, se o músico “arredonda” as alturas para a escala cromática temperada, qual a relação da frequência com o ritmo percebido pelo músico, entre outras. Decidimos então investigar se há alguma evidência acústica que apoie as escolhas das alturas feitas por Hermeto Pascoal, empregando métodos e conceitos derivados dos estudos prosódicos. As primeiras avaliações da comparação entre as notas escolhidas por Hermeto Pascoal para cada sílaba do enunciado dos seus “homenageados” apontam para uma correlação significativa com a frequência dos alvos estáticos, com desvio médio de cerca de um quarto de tom. Para prosseguirmos, porém, nesta investigação ainda é necessário um levantamento de dados mais detalhado, e a elaboração de testes estatísticos que possam não só corroborar a hipótese, mas também revelar outros aspectos do alinhamento que contribuam para o avanço na compreensão do fenômeno da entoação.

#### IV. UMA TIPOLOGIA DA ENTOAÇÃO

Supondo que as questões acima discutidas apontem realmente para a conclusão de que o alinhamento da curva de frequência fundamental com a cadeia segmental esteja ancorado nos pontos de inflexão em que a curva começa a se dirigir ao alvo, e que, desta forma, os alvos (sejam frequências ou velocidade de variação de frequência) são os elementos

significativos da entoação como propriedade de cada sílaba, restaria analisar qual o papel da estabilidade da altura na composição do sentido da entoação, uma vez que contornos iguais do ponto de vista dos alvos resultam sentidos diferentes conforme a estabilidade da altura ao longo da sílaba.

Em um artigo singular, George List aborda esta questão na investigação da entoação do ponto de vista da etnomusicologia [14]. O autor comenta a diversidade que existe na maneira como a altura é empregada nos enunciados dependendo do contexto social em que. Ele chega a propor uma classificação para os diferentes tipos de realização prosódica (no caso, entoacional), dependendo da variação (extensão) da melodia e da estabilidade da altura ao longo da sílaba. A classificação de List emprega dois parâmetros básicos: conformação das alturas a um dado sistema escalar e estabilidade das alturas. Aparentemente há uma contradição neste caso, já que para verificarmos a conformação à escala, as alturas têm de apresentar um mínimo de estabilidade. E esta fragilidade da proposta coincide exatamente com o ponto crucial de nossa investigação, a saber: como atribuímos a altura a uma sílaba?

Em sua taxonomia List cita como tipos o salmodiar, a declamação recitada, o pregão, a fala do leiloeiro, fala espontânea, canto, *sprechgesang*, entre outros, estabelecendo as distinções entre eles segundo o comportamento da curva de frequência, tornando mais claras as nuances que distinguem um tipo de outro. A ampliação do conjunto de parâmetros e adaptação metodológica propostos por List para o estudo da entoação propriamente dita está entre os próximos passos da nossa pesquisa.

#### V. CONSIDERAÇÕES FINAIS

Até que ponto a metáfora “melodia da fala” pode ser estendida? As diferentes modalidades do discurso diferenciam-se de maneira notável pelo emprego da altura na constituição dos seus enunciados. As características “melódicas” indicam a função com que cada enunciado deve ser entendido ou contextualizado, se é uma canção, um salmo, um recitativo, um *teasing*, um discurso jurídico, uma locução de rádio, um pregão, e assim por diante. Podemos estender este princípio para dentro da conversa, quando a melodia também atua como uma marca de contexto, como é o caso das atitudes proposicionais, por exemplo. É bem possível que haja uma conexão entre estes dois fenômenos, mas para investigar esta conexão, precisamos estabelecer critérios para descrever estas “características melódicas”.

Acreditamos, de qualquer modo, que a investigação da percepção da altura de um ponto de vista musical aplicado aos estudos prosódicos pode esclarecer muitos aspectos ainda obscuros da misteriosa melodia da fala.

#### AGRADECIMENTOS

Agradeço a todos os participantes do III Colóquio Brasileiro de Prosódia da Fala, que com seus comentários e críticas contribuíram para a definição da forma final do presente trabalho; agradeço especialmente ao meu orientador no programa de pós-graduação em Filologia e Língua Portuguesa (FFLCH/USP), prof. Waldemar Ferreira Netto, pelo

contínuo incentivo com relação à abordagem interdisciplinar (musical e lingüística) do estudo da entoação.

#### REFERÊNCIAS

- [1] LADD, D. Robert. *Intonational Phonology*. Cambridge: Cambridge University Press, 1996, p. 51 e ss.
- [2] FOX, Anthony. *Prosodic features and prosodic structure: the phonology of suprasegmentals*. New York: Oxford University Press, 2000, p. 298 e ss.
- [3] ZATORRE, Robert; BELIN, Pascal; PENHUNE, Virginia B. Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*. Nº1, vol. 6, pp. 37-46. Elsevier, 2002
- [4] PERETZ, Isabelle; HYDE, Krista L. What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*. Nº8, v. 7, pp. 362-367. Elsevier, 2003
- [5] PERETZ, Isabelle. Brain specialization in music. *The Neuroscientist*. Nº 8, v. 4, pp. 374-382. Sage, 2002 [cita dificuldade dos amúsicos com relação à entoação quando outras marcas prosódicas estão ausentes (p.377)]
- [6] FEDORENKO, Evelina; PATEL, Aniruddh; CASASANTO, Daniel; WINAWER, Jonathan; GIBSON, Edward. Structural integration in language and music. *Memory & Cognition*. Nº 37, v.1. pp. 1-9. 2009
- [7] KLEIN, Denise; ZATORRE, Robert; MILNER, Brenda; ZHAO, Viviane. A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage*. Nº 13, pp. 646-653. Academic Press, 2001
- [8] DEUTSCH, Diana; HENTHORN, Trevor; DOLSON, Mark. Absolute pitch, speech and tone language: some experiments and a proposed framework. *Music Perception*. vol. 21, nº 3, pp- 339-356. The University of California Press, 2000
- [9] ROSS, Deborah; CHOI, Jonathan; PURVES, Dale. Musical intervals in speech. *PNAS*, vol. 104, nº 23. Disponível em <[www.pnas.org/cgi/doi/10.1073/pnas.0703140104](http://www.pnas.org/cgi/doi/10.1073/pnas.0703140104)>
- [10] XU, Yi. Timing and coordination in tone and intonation – an articulatory-functional perspective. *Lingua*, 119: 906-927, 2009
- [11] PEARL, Jonathan Geoffrey Secora. *The Music of Language: The Notebooks of Leoš Janáček*. Tese de Doutorado. Santa Barbara: University of California, 2005
- [12] PASCOAL, Hermeto. *O som da aura*. Disponível em <[www.hermetopascal.com.br](http://www.hermetopascal.com.br)>. Acesso em 24/04/2011
- [13] PASCOAL, Hermeto. *Festa dos Deuses*. CD. Polygram, 1992
- [14] LIST, George. *The Boundaries of Speech and Song*. *Ethnomusicology*, vol. 7, no. 1, pp. 1-16 University of Illinois Press, Jan. 1963