

OS LIMIARES DE DIFERENCIAÇÃO TONAL DO PORTUGUÊS BRASILEIRO

Marcus Vinicius Moreira MARTINS¹

Waldemar Ferreira NETTO²

Resumo: O objetivo deste trabalho é determinar os limiares de diferenciação tonal (T'HART, 1981) do Português Brasileiro, no que se refere à entoação. Os limiares seriam pontos a partir dos quais os falantes são capazes de identificar diferenças significativas de variação tonal, levando-os a considerar aquele ponto como dotado de alguma informação linguística relevante, como uma ênfase em um vocábulo ou sílaba. Para estabelecer os limiares do Português Brasileiro, aplicamos dois experimentos de percepção com 13 pares de duas orações faladas por uma voz masculina. Os pares consistiam de uma mesma oração neutra pareada 12 vezes com ela mesma, mas com o F0 manipulado em seis semitons positivos e seis negativos, a partir dos valores originais. No Teste I, 16 participantes, sem treinamento musical, respondiam se notavam ou não qualquer diferença entre os pares apresentados. No Teste II, os mesmos 16 participantes podiam responder sim, não ou talvez, o objetivo desta última alternativa era avaliar pontos nebulosos. Para determinar os limiares de diferenciação tonal, utilizamos um modelo de análise de componentes principais, o qual retornou como limiares os valores de -3 e +4, indicando que são estes os valores do limiar de diferenciação tonal do Português Brasileiro.

Palavras-chave: Fonética. Fonologia. Percepção. Entoação. Acústica.

¹ USP – Faculdade de Filosofia, Letras e Ciências Humanas – Programa de Pós-graduação em Filologia e Língua Portuguesa. São Paulo – São Paulo – Brasil. 05508-900 – marcusvmmartins@gmail.com

² USP – Faculdade de Filosofia, Letras e Ciências Humanas – Departamento de Letras Clássicas e Vernáculas. São Paulo – São Paulo – Brasil. 05508-900 – wafnetto@usp.br

<http://dx.doi.org/10.21165/gel.v14i2.1762>

Introdução

Uma discussão central quando se fala em percepção de fala refere-se a qual é o elemento alvo da percepção de um falante: os elementos linguísticos ou a elocução em si? Embora grande parte dos estudos se utilizem de material de laboratório, sem os estímulos externos derivados do contexto de fala, podemos assumir, por ora, que o primeiro passo no processo de percepção é a decodificação do sinal de fala, ainda que no interior de um conjunto maior de estímulos sonoros. A partir deste processo o ouvinte é capaz de detectar os padrões do som da fala, mesmo com os efeitos de outras fontes tomando parte, sendo assim possível rastrear os elementos linguísticos que estão presentes em uma elocução (CHERRY, 1953; KUHL, 2001). Neste sentido, em primeiro nível, o elemento alvo da percepção de um falante seria a elocução, o que explicaria a capacidade de se identificar um sinal de fala de uma língua a qual não dominamos (OHALA, 1996). Em um segundo nível estariam os padrões detectáveis, sendo eles os elementos que compõem o domínio linguístico de um falante (HALLE, 2002).

Organizar a experiência sensorial em padrões é um dos pressupostos básicos seguidos pelos psicólogos da *Gestalt* (WERTHEIMER, 1938); para eles, um ambiente homogêneo e sem variações não pode ser detectado pela percepção, a qual precisa necessariamente que os elementos presentes neste ambiente destaquem-se do fundo. Dentro desta abordagem, a percepção seria capaz de organizar em padrões os estímulos recebidos pelo nosso sistema sensorial, basicamente o visual e o sonoro, e destacá-los do fundo caótico, criando assim os elementos perceptíveis.

Os estudos da Gestalt compreendem que a percepção de um objeto está assegurada pela permanência do mesmo objeto, isto é, independente do circuito de transformações fisiológicas e físicas ocorridas entre a visualização e a manifestação dessa percepção em nosso cérebro, o objeto será o mesmo, desde que o ambiente ou objeto em si não sofram variações. Neste caso, diferente do que pressupõe Bertrand Russell (1978) e seus antecessores, a percepção não é um produto derivado de processos inerentes ao indivíduo e, sim, uma propriedade do percepto. Dentro desta visão, o agrupamento de padrões ou de características se tornou um dos pressupostos básicos dos estudos de percepção, uma vez que a alteração deles ou de um deles pode alterar o modo como o objeto é percebido (WERTHEIMER, 1938).

No caso da percepção da fala, podemos elencar diversos elementos que foram analisados com o intuito de serem compreendidos como aqueles que concorrem para a percepção de determinados elementos linguísticos, tais como a duração e frequência fundamental do acento, bem como a qualidade das vogais (FRY, 1955, 1958, 1965), amplitude da sílaba acentuada (MOL; UHLENBECK, 1956; BOLINGER, 1958) e a pressão subglotal presente na vogal da sílaba acentuada (VAN KATWIJK, 1974). No caso da percepção da entoação, o único correlato acústico a ser explorado é a variação de frequência fundamental e seu correlato perceptivo, o *pitch*, de modo que as análises podem se pautar pela análise do efeito de suas variações no tempo, ou seja, a configuração do *pitch* em uma elocução, correlacionada às suas variações de altura, ou por meio de um modelo que avalie os níveis do *pitch*, sendo que estes dois modelos de análise formam uma grande parte das análises envolvendo o estudo da entoação (PIKE, 1945; BOLINGER, 1961; PIERREHUMBERT, 1980; T'HART *et al.*, 1990; LADD, 1996).

Como aponta Vaissière (2004), o modelo de análise da percepção de F0 pode variar, a parte do modo como se procura compreender as variações de F0. Elas podem se dividir em quatro grandes modelos de análise da percepção: (i) um primeiro modelo compreenderia a entoação como a soma de elementos locais e simbólicos, tais como *pitch accents* e tons de fronteira, (ii) o segundo compreende a entoação como a superposição de diversos eventos entoacionais correlatos, ao passo que (iii) o terceiro modelo veria nesses eventos correlatos funções linguísticas que precisam ser analisadas em conjunto com as modulações de F0, o quarto e último modelo (iv) seria direcionado ao ouvinte e parte da hipótese de que a análise da entoação tem de ser feita a partir dos padrões de F0 relevantes para o ouvinte. Para o presente trabalho utilizaremos a última proposta.

A abordagem orientada ao ouvinte

O modelo direcionado ao ouvinte teve um grande desenvolvimento nas décadas de 80 e 90 no IPO (Institute for Perception Research em Eindhoven), tendo seus achados e propostas condensados no trabalho de T'Hart, Collier e Cohen (1990), no qual propõem que a entoação seja compreendida basicamente como a variação de F0, de modo que certos movimentos tonais (*pitch movements*) são relevantes para o ouvinte. Estes movimentos tonais no plano cognitivo corresponderiam a comandos para movimentos

específicos das pregas vocais, os quais podem ser recuperados no sinal da fala, por meio de testes de percepção ou da avaliação psicométrica da resposta dos falantes aos estímulos.

Para os pesquisadores do IPO, um dos componentes essenciais para compreender melhor as funções linguísticas da variação de F0 era o *limiar de diferenciação tonal* (LDT). A definição do termo envolve duas dimensões básicas: a variação de frequência e a magnitude perceptível dessa variação. Desta maneira, o limiar de diferenciação tonal pode ser definido como um valor z a partir do qual o indivíduo é capaz de notar alguma diferença entre dois tons seguidos, mesmo que eles façam parte de um mesmo segmento sonoro. T'Hart, Collier e Cohen (1990) apresentam como exemplo os estudos de Ritsma (1965 *apud* T'HART; COLLIER; COHEN, 1990), no qual os ouvintes não eram capazes de identificar um tom isolado de 1KHz, porém eram capazes de diferenciar dois tons seguidos, tendo um 1KHz e outro 1,005 KHz. Em outras palavras, era necessário que um tom variasse, ao menos 0,5%, para que a sensibilidade tonal dos ouvintes fosse acionada e eles pudessem notar diferenças entre dois tons com frequências acima de 1KHz.

T'Hart, Collier e Cohen (1990) justificam a definição dos limiares de diferenciação tonal por meio de um conceito bastante difundido nos estudos de acústicas, as *just noticeable difference*³ (j.n.d.). A j.n.d. é uma medida psicofísica proporcional, construída sobre a lei de Weber e definida como:

- (1) Dado um determinado estímulo y fixado, deve-se variar seu par associado (y_1) até o ponto em que o sujeito seja capaz de notá-los como distintos.

A divisão $\frac{y}{y_1}$ retorna um valor proporcional capaz de expressar o quão acurada é a percepção de um sujeito dado um conjunto de estímulos. Roederer (2003) demonstra que as j.n.d.'s possuem uma alta resolução, em torno dos 3%, na faixa dos 500 Hz aos 1000 Hz, sendo que tal resolução diminui com o acréscimo de frequência, tornando-se da ordem de 0,5% para frequências entre 1KHz e 2KHz, isto é, tons na faixa de produção da frequência fundamental da fala humana necessitam de uma variação maior para serem

³ A tradução de *just noticeable difference* é dada como *diferença no limite do observável* em Roederer (2003). Embora seja uma tradução técnica, ela não traduz com fidelidade o termo em inglês, por esta razão optamos por utilizar o termo original. Recomendamos o artigo de Zwicker, Flottorp e Stevens (1957) para maiores esclarecimentos.

percebidos como diferentes, ao passo que sons em altas frequências não dependem dessa variação mais ampla.

T'Hart (1981) aplicou estas propostas aos estudos de fala a fim de determinar os limiares de diferenciação tonal no que se refere à entoação do holandês. Para o autor, o falante poderia produzir modalizações nas curvas entoacionais, como uma marcação de foco, controlando a glote a fim de que os valores de frequência fundamental ultrapassassem os limiares de diferenciação tonal. Em seu estudo, sílabas de uma lista de palavras isoladas foram manipuladas dentro de um espectro de um a seis semitons ascendentes e descendentes. Os testes indicaram que para que um falante de holandês note qualquer diferença tonal entre dois estímulos, é necessário que se varie ao menos três semitons ascendentes ou três semitons descendentes.

Consoni (2011) aplicou os mesmos princípios experimentais de T'Hart (1981) para o português brasileiro. A autora aplicou dois testes de percepção distintos, porém complementares, a fim de determinar os limiares de diferenciação tonal para o Português Brasileiro e a percepção do foco, utilizando palavras isoladas. Nos dois testes da autora, a tarefa dos participantes era apontar se havia ou não diferenças entre os pares de estímulos apresentados, ou seja, testes de discriminação. No primeiro teste de Consoni (2011), foram utilizados dez pares de palavras trissílabas, todas as palavras tiveram suas frequências fundamentais variadas em um a quatro semitons ascendentes e um a seis semitons descendentes; o objetivo era verificar se as diferenças tonais eram perceptíveis pelos falantes neste contexto. No segundo teste, também de percepção, os pares de sílaba foram agrupados em dois conjuntos: (i) cuja sílaba tônica havia sido manipulada; (ii) cuja palavra toda havia sido manipulada. Todas as palavras e frases foram gravadas com uma voz feminina. Em sua análise geral, integrando as respostas dos dois testes, Consoni (2011) chegou aos valores de três semitons ascendentes e quatro semitons descendentes.

Em ambos os experimentos, as janelas temporais de análise foram da ordem de uma palavra trissílaba ou maior, por volta dos 60-100 milissegundos. Nossa proposta é a de se avaliar os mesmos parâmetros, contudo tendo como parâmetro de análise a frase, com a intenção de verificar se os limiares de diferenciação tonal sofrem alguma influência direta da duração. A hipótese desse trabalho se sustenta no efeito *tau*, pois segundo os experimentos de Shigeno (1986) a distância temporal entre dois estímulos afeta a percepção tonal: quanto menor o tempo entre dois tons, menor a frequência do segundo

tom é percebida pelos sujeitos. A fim de verificar estas questões, aplicamos um teste de percepção, seguindo os princípios de T'Hart (1981) e Consoni (2011).

Método: fase preparatória

Na primeira etapa de nosso experimento, um sujeito do sexo masculino com 23 anos, natural de Mogi das Cruzes, região metropolitana de São Paulo, foi gravado em uma sessão voltada à coleta dos áudios necessários para a manipulação. Nesta seção, foram gravadas três repetições de leitura com cinco frases distintas. As gravações foram feitas em três séries de cinco gravações, gravando-se as orações de maneira sequenciada, isto é, gravavam-se todas as frases em uma série e não as três repetições de uma mesma frase em uma série. Ressaltamos que as gravações foram feitas em ambiente de ruído controlado, mas não issonorizado.

Entendemos que a terceira repetição seria a com menor variação de frequência fundamental, uma vez que as duas primeiras seriam como treinos para nosso informante. De maneira geral, o principal problema, no que se refere à entoação, foi o fato de o informante focalizar itens lexicais com um aumento da frequência fundamental, afetando diretamente o nosso objeto de análise. Com relação aos segmentos, o problema foi o apagamento ou a curta duração de determinados segmentos, como a supressão de vogais átonas em final de palavra ou frase. Tendo em vista que o tempo é uma de nossas variáveis, destacamos que as orações selecionadas possuem uma duração média de 1,7 segundos. O conteúdo lexical e informacional das frases não era relevante, no entanto, também não se tratavam de frases sem sentido.

Fase de manipulação

A segunda etapa de nossa pesquisa consistiu na manipulação do estímulo a ser utilizado no experimento de percepção. Optamos por manipular as frequências fundamentais no intervalo de uma oitava, isto é, de seis semitons ascendentes a seis semitons descendentes. A intenção ao se optar pela manipulação por oitavas é a de ter um maior controle sobre a banda de frequências na qual os valores de F0 estarão, uma vez que os valores de F0 do estímulo manipulado com +6 semitons deverão ser o dobro

daqueles manipulados com -6 semitons. Na Tabela 1, exemplificamos esse pressuposto, apresentando os valores finais de F0 do estímulo “O João mora na casa verde”, manipulado em seis semitons positivos e negativos, além dos valores naturais. Destacamos que todas as manipulações foram executadas com a função *To Manipulation* do Praat (BOERSMA; WEENIK, 2009).

Tabela 1: Tabela contrastiva dos valores de frequência fundamental, após manipulação com o Praat

	- 6 semitons	+ 6 semitons
Freq. Fund. Média	104,49 Hz	209,65 Hz
Freq. Fund. Mínima	91,47 Hz	182,97 Hz
Freq. Fund. Máxima	112,60 Hz	225,89 Hz

Como pode se notar na Tabela 1, os valores de média, mínimo e máximo das manipulações de seis semitons ascendentes são o dobro dos valores das manipulações de seis semitons descendentes, o que demonstra a eficiência da manipulação. O restante das variáveis foi manipulado pelo mesmo processo. Para essa manipulação, usamos como parâmetro os valores de 0,01 segundos para *time step* e uma janela de 50 Hz a 600 Hz como valores de mínimo e máximo, respectivamente. Desta forma, valores abaixo e acima destes limites não foram manipulados e a variação da frequência fundamental era feita de 10 em 10 milissegundos; destacamos que toda a extensão de F0 foi manipulada.

Obtivemos assim 13 amostras, uma natural (não-manipulada) e outras 12 manipuladas de -6 semitons até +6 semitons. A partir destas gravações, montamos o teste de percepção utilizando 13 pares: a gravação não-manipulada (zero) pareada com todas as demais doze manipuladas (de +6 a -6) e uma não-manipulada pareada com uma não-manipulada (par 0-0).

Design Experimental

Uma das questões relativas à percepção está associada aos pontos de inflexão, isto é, pontos a partir dos quais alteração das características naturais do percepto são tantas que levam o indivíduo a notar a diferença entre eles. No entanto, estes pontos não são

categóricos e, sim, gradientes (SCHOUTEN, 1987), de tal modo que é natural que esses perceptos sejam confundíveis entre si, pois nesses pontos nebulosos o participante não sabe, ao certo, se há ou não diferenças entre os estímulos apresentados. Pensando nessa questão, nosso delineamento tem como pressuposto usar dois estímulos, mas com procedimentos diferentes. No primeiro teste, o participante seria forçado a responder “sim” ou “não”, ao passo que no segundo haveria a opção “neutra”. A ideia por detrás desse delineamento é diminuir os erros do participante, bem como procurar um número balanceado de respostas.

Os fundamentos dos testes (Teste I e Teste II), bem como os cuidados com o delineamento e as questões a serem respondidas, foram feitos com base em Rea e Parker (2000), Cozby (2003) e Dancey e Reidy (2006). O Teste I foi inspirado nos chamados *de two-alternative forced choice tasks* e comumente abreviado por *2AFC tasks*. A lógica destes testes refere-se à capacidade de os indivíduos fazerem escolhas com base em julgamentos daquilo que consideram adequado para uma determinada tarefa, sendo que destas respostas pode-se analisar a lógica das tomadas de decisão (HILL, 1898; BOGACZ *et al.*, 2006). O teste II foi inspirado nos testes de *multiple forced choice* e a utilização de três alternativas foi feita baseada na escala de Ramsay, a qual avalia o nível de sedação em pacientes internados em UTI's com seis níveis de avaliação. Cada nível corresponde a um grau de sedação e eles se opõem internamente (1 se opõe a 6, 2 se opõe a 3, 4 se opõe a 5) e são obtidas a partir de uma avaliação feita com o próprio paciente. Para o desenvolvimento da escala de três níveis, baseamo-nos na escala Likert (LIKERT, 1932) de cinco níveis.

Para calcularmos o tamanho da amostra, utilizamos o *software G*Power* (ERDFELDER; FAUL; BUCHNER, 1996) e como teste de hipótese optamos por utilizar a ANOVA dentre fatores, isto é, no teste de percepção todos os falantes participariam de todas as condições experimentais, sendo que essa escolha também perpassa o objetivo inicial de nossos testes, o qual era avaliar os pontos nebulosos de variação. Neste caso, para avaliar a homogeneidade das respostas e encontrar os pontos de gradiência seria necessário que os sujeitos participassem dos dois testes.

Deste modo, para satisfazer estes pressupostos, utilizando um $\alpha = 0,05$ e um $\beta = 0,95$, tendo um grupo de análise e 13 repetições (neste caso a avaliação é feita com relação à sentença 0, a qual repete 13 vezes), o G*Power retornou um $n = 17$. O resultado

apresentado demonstra a necessidade de apenas 17 sujeitos para que o teste de hipótese tenha um tamanho de efeito significativo. Neste caso específico, o tamanho de efeito escolhido foi de $f = 0,25$, o que é considerado médio. A Figura 1 apresenta os valores de significância a serem assumidos de acordo com o tamanho da amostra coletada.

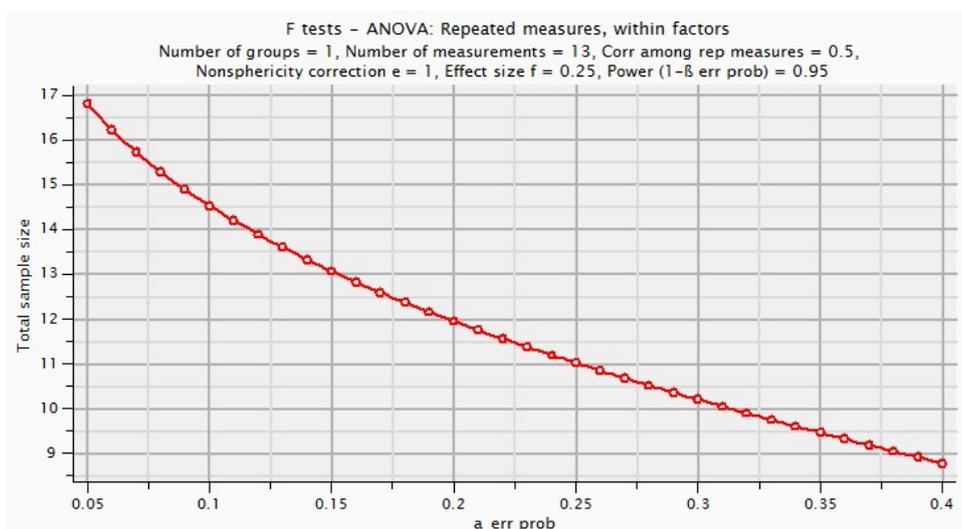


Figura 1: Relação entre o valor de α e o tamanho n da amostra

Fase de execução

Um teste piloto a fim de averiguar a reação dos participantes foi aplicado em cinco pessoas, que não participaram do teste final, sendo que a principal dúvida referia-se ao que devia ser julgado: o domínio semântico, sintático ou fonético. Para os participantes, encontrar erros de gramática ou diferenças segmentais mínimas, ou ainda significados diferentes era mais importante do que observar a entoação, uma vez que a percepção deles, como falantes, está mais centrada nestes aspectos. Deste modo, foi necessário que estabelecêssemos orientação detalhada, seguida de um teste com uma frase aleatória, com sua versão neutra e outros três pares manipulados apenas para que os participantes se habituassem à interface e tirassem dúvidas. Os dois testes foram montados a partir da pergunta “As sentenças são parecidas?”, a qual poderia ser respondida com um “sim”, “não” ou “talvez”.

O Teste I e o Teste II eram aplicados em sequência. A frase manipulada no teste I era “O João mora na casa verde”, enquanto no Teste II os participantes ouviam a frase “O gato está em cima do muro”. A aplicação se deu em ambiente de ruído controlado, mas

não issonorizado. O teste foi aplicado em 16 pessoas, satisfazendo o tamanho de efeito calculado, de acordo com nossa estimativa apresentada na Figura 1. Não avaliamos, neste caso, a variedade regional de cada um dos falantes, a faixa etária se distribuía entre 20 e 50 anos e obedecia a uma distribuição balanceada de gêneros. Para a aplicação, utilizamos o *script ExperimentMFC* do Praat, cuja interface pode ser vista na Figura 2. Foram utilizados um *notebook* de 14” para a realização da tarefa, um *mouse* óptico com fio e um fone de ouvido *Sennheiser* HD-280. Em cada teste, os participantes eram avaliados isoladamente, isto é, apenas um sujeito respondia ao teste por vez.

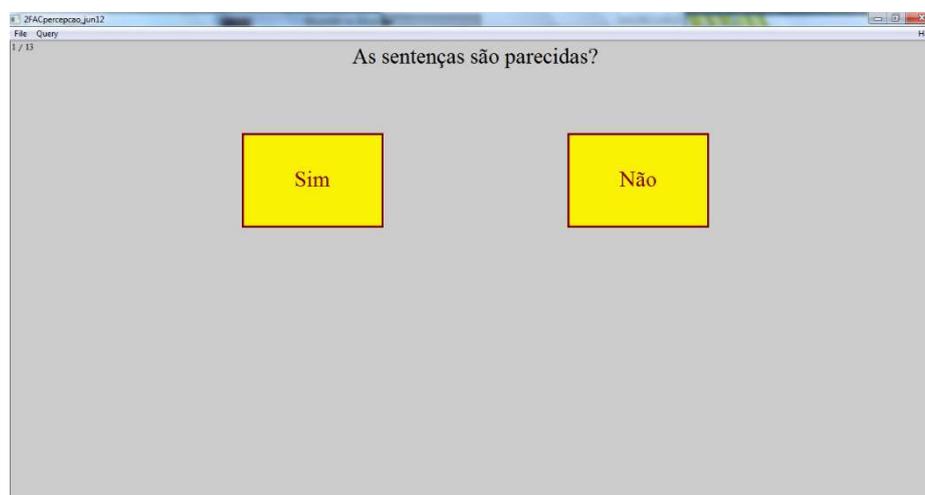


Figura 2: Interface usada para o experimento de percepção

Importante destacar que não havia intervalo entre os estímulos, uma vez que havia um silêncio anterior ao início da fala na gravação original não manipulada. O tempo total entre o início do primeiro estímulo e o fim do segundo era da ordem de cinco segundos para os dois testes, o que nos dá uma janela de processamento temporal da ordem de um segundo para cada estímulo, mais um segundo de intervalo entre elas. Além disso, foi escolhida a ordem randômica de apresentação dos pares, isto é, o Praat escolhia quais seriam os pares apresentados de maneira arbitrária, com base em um algoritmo próprio. O tempo médio total para aplicação do teste, somado o tempo para as respostas, mais o de instrução e o de adaptação ao equipamento pelo participante foi da ordem de nove minutos. A Tabela 2 resume os valores obtidos na fase de verificação do teste. Importante destacar que valores de resposta fora desses padrões nos ajudam a detectar *outliers*, por

exemplo, tempos de respostas maiores que 15 ou 20 segundos demonstram que o participante em algum ponto confundiu os pares a serem analisados.

Tabela 2: Tabela com os valores de tempo para cada etapa do teste

Duração (tempo gasto)	
Duração de cada estímulo	5,17s (Teste I) e 5,36s (Teste II)
Duração de cada par de estímulo	10,34s (Teste I) e 10,72s (Teste II)
Duração média total do teste mais treino	9 min.
Duração média do tempo de respostas	5 min.
Duração média para instrução e treino	4 min.

A aplicação do teste final seguiu o seguinte roteiro, tendo sido aplicado em 16 pessoas:

1. Instrução oral, em que o informante recebia do analista as orientações do que deveria ser feito;
2. Teste de familiarização. Momento no qual era apresentada a janela de respostas como aquela reproduzida na Figura 2. Na etapa de familiarização, tinham-se apenas três pares de estímulos diferentes daqueles utilizados no teste;
3. Teste I e II, o informante respondia ao teste completo com os 13 estímulos (sendo o par zero-zero, mais 12 pares zero mais as manipulações). Cada estímulo tinha uma duração de cerca de dez segundos.
4. Entre os testes I e II havia uma pausa, para alertar o informante que a partir daquele ponto teria início o teste com três alternativas.

Análise – Teste I: 2AFC - 2 Alternative Forced-Choice

Ao final do teste, obtivemos um total de 208 respostas, sendo 13 respostas por 16 sujeitos ($13 \times 16 = 208$, logo $n=16$ e $n'=208$). As Figuras 3 e 4 apresentam a distribuição de respostas. A Figura 3 demonstra a quantidade de respostas “não” para cada uma das variáveis, enquanto a Figura 4 apresenta a quantidade de respostas “sim” para cada variável. Importante ressaltar que a resposta “não” é a resposta de controle, isto é, os cálculos estatísticos são feitos a partir da quantidade de “não” dados, ao passo que a resposta “sim” serve como dado complementar.



Figura 3: Número de respostas “não” obtidas no teste 2AFC para cada variável

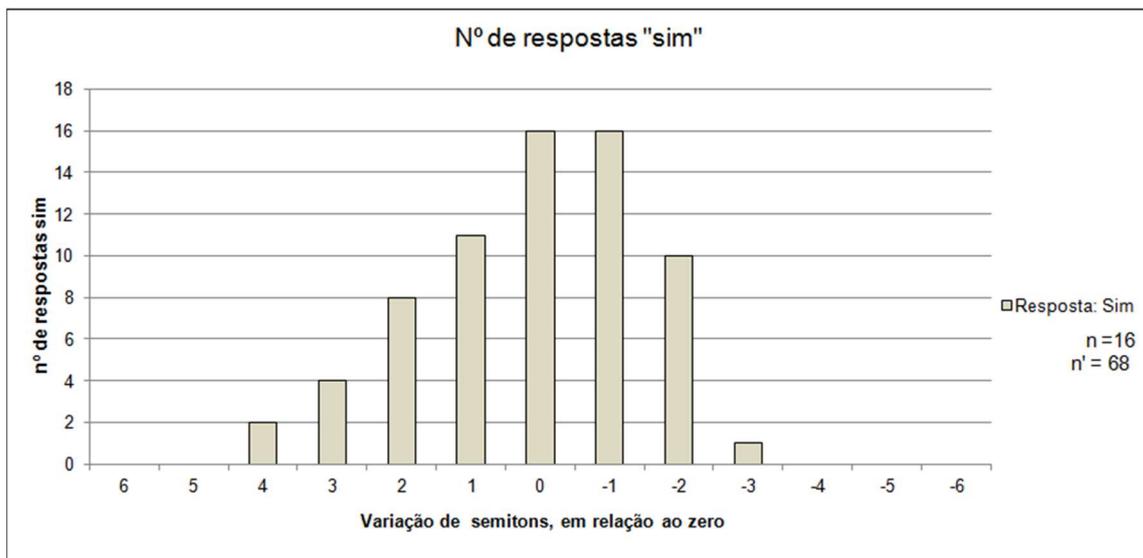


Figura 4: Número de respostas “sim” obtidas no teste 2AFC para cada variável

Ambos os gráficos demonstram que temos uma tendência à não diferenciação nos valores mais centrais, em especial, na faixa que vai de -3 a + 4 semitons. Note-se que obtivemos um maior número de respostas “não” (67,3%, $n'=140$), demonstrando que o índice de diferenciação é relativamente maior do que o de não diferenciação (32,7%, $n'=68$). A seguir, executamos uma ANOVA entre todas as variáveis de análise para $n = 16$. A hipótese nula assumida é a de que não há diferenças entre as respostas dadas para cada um dos tons variados, isto é, que os falantes não notariam diferenças entre eles. Obtivemos $p < 0,001$, ($F = 31,71$, para um $F_{critico} = 1,84$), para $\alpha = 5\%$ e $n = 16$, $gl = 12$, tomando como base uma ANOVA sem repetição e de múltiplos fatores dentre sujeitos.

Tal *p-value* permite-nos rejeitar a hipótese nula e supor *H1*, uma vez que ele é estatisticamente significativo, temos também o valor F ($F = 33,07$) maior que seu valor crítico ($F_{crítico} = 1,81$), o que corrobora a hipótese alternativa, dizendo que há diferenças entre as variáveis selecionadas. Os cálculos foram executados com o *software* MsExcel.

Neste ponto, é preciso categorizar as variações em função das respostas dadas pelos participantes, com o intuito de facilitar a aplicação de novos testes estatísticos. O primeiro passo é estabelecer as variâncias, pois o seu valor nos permite estabelecer o padrão de respostas dos participantes. A partir dos gráficos podemos sugerir a sistematização dos dados em dois grupos distintos de variáveis: um grupo de *variáveis variável* e outro de *variáveis invariável*. No primeiro, temos que a variância é maior que zero ($Var(X) > 0$), isto é, obtivemos ao menos uma resposta “sim” ou “não”, enquanto no segundo temos que a variância é igual a zero ($Var(X) = 0$), ou seja, nestes casos obtivemos apenas respostas “sim” ou “não”. Deste modo, as variáveis podem ser organizadas da maneira apresentada na Tabela 3:

Tabela 3: Agrupamento das variáveis com base na Variância

Condição	
Invariáveis	6 ,5, -1, -4, -5 e -6
Variáveis	4, 3, 2, 1, -2 e -3
Fixo	0

As variáveis mais6, mais5, bem como a variável neutra e as variáveis menos1, menos4, menos5 e menos6 podem ser entendidas como *variáveis categóricas*, isto é, para os falantes, as alterações feitas não afetam o modo como são percebidas. Desta maneira, podemos dividi-las em duas grandes categorias: *categóricas por semelhança*, quando são percebidas como iguais a partir do estímulo neutro e *categóricas por diferença*, quando são percebidas como diferentes. Na primeira categoria, encontram-se as variáveis neutra e menos1, as quais obtiveram 100% de respostas “sim” no teste de percepção, ao passo que na segunda categoria estão as variáveis mais6, mais5, menos4, menos5 e menos6, as quais obtiveram 100% de respostas “não”.

Para a categorização das demais variáveis, precisamos executar uma nova ANOVA levando em conta apenas as seis variáveis do grupo variável. Neste teste, obtivemos um *p-value* $< 0,001$ ($F = 9,69$, para um *F* crítico = 2,33), $\alpha = 5\%$ e $n = 16$, gl

= 12. Com base nestes números podemos, novamente, excluir a hipótese nula de não-diferenças entre grupos e assumir que haja uma sub-organização entre os elementos que compõem o grupo variável, isto é, que a percepção deles é distinta, mesmo que não seja homogênea por parte dos ouvintes.

A fim de tratar da questão da subsistematização, fizemos uma análise dos intervalos de confiança, apresentados na Figura 5. Este tipo de análise nos permite encontrar a média amostral dentro de um determinado intervalo de valor. No caso desta análise, a média central seria de 1,5, uma vez que as respostas analisadas estão opostas entre 1 e 2, valores por nós estabelecidos. Por exemplo, a média de um grupo cujas respostas foram unicamente “sim” seria 1, por outro lado, um grupo com respostas “não” teria sua média em 2.

Deste modo, intervalos com centro (o pequeno círculo na barra de erros) em, aproximadamente, 1,5 podem ser entendidos como elementos de confusão, pois contém respostas “sim” e “não” de forma quase equilibrada, ou seja, seriam variáveis de confusão. Intervalos que possuem seu centro deslocado de 1,5 têm uma maior tendência a serem quasi-categóricos, pois quando o centro da barra de erros for maior que 1,5 e próximo a 2 é possível assumir que eles tendem a ser compreendidos como diferentes, embora não sejam percebidos como diferentes de maneira categórica; por outro lado, quanto mais próximos de 1, mais as variáveis tenderão a ser compreendidas como semelhantes. O tamanho da barra de erros tem importante papel, pois ela nos demonstra a probabilidade de que uma determinada variável seja percebida como estímulo não manipulado, sendo assim, quanto maior a barra, maior é essa probabilidade e vice-versa.

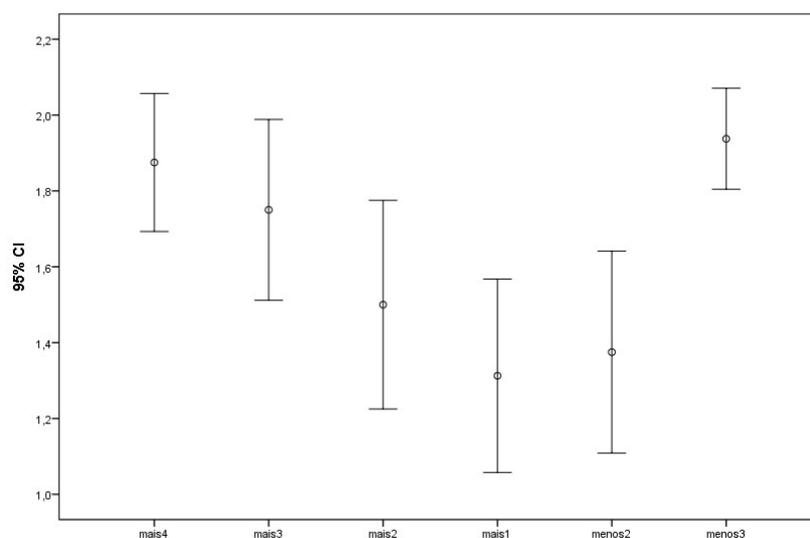


Figura 5: Gráfico barras de erros com médias de respostas

Nesse sentido, o gráfico da Figura 5 permite-nos dizer que as variáveis mais4 e menos3 são *quasi-catóricas por diferença*, uma vez que possuem a média próxima a 2 e têm comprimentos intervalares pequenos, em relação aos demais. Neste caso, a probabilidade de que um indivíduo as confunda com o zero é menor, ou seja, estas variáveis tendem a ser notadas como diferentes em grande parte das vezes. As variáveis mais2 e mais3 podem ser analisadas como *variáveis de confusão*, pois possuem seus centros próximos a 1,5 e dimensões intervalares mais extensas que as variáveis mais4 e menos3, sendo assim, existe a possibilidade de que elas sejam ora vistas como diferentes, ora como semelhantes, sendo esta uma variação inerente ao sujeito e não ao percepto.

Já as variáveis mais1 e menos2 podem ser consideradas *quasi-catóricas por semelhança*, dados a proximidade do centro com 1 e o tamanho dos intervalos. Neste caso, temos que a probabilidade de se associá-las com o zero é maior do que se comparado com as variáveis de confusão, em outras palavras, estas variáveis são notadas como semelhantes a zero em grande parte das vezes. O grupo de variáveis invariáveis pode ser considerado *catóricas por semelhança e catóricas por diferença*, posto que a probabilidade de se associá-las ou diferenciá-las do zero é igual a 0%, ou seja, é bastante improvável que um sujeito venha a confundi-las.

Estes dados nos permitiram aplicar uma análise de fatores, baseada em uma análise de componentes principais, a qual indicaria como deveríamos organizar e agrupar nossas variáveis. A análise de fatores é um teste estatístico de hipótese baseado em testes

de correlações. Tal teste tem por pressuposto o de comparar valores dados a uma determinada variável com os valores atribuídos a todas as demais. A Tabela 4 apresenta os valores de correlação das variáveis de nosso teste, as quais tiveram variância maior que o ($Var > 0$), isto é, que obtiveram ao menos uma resposta não.

Tabela 4: Tabela de correlação entre as variáveis do grupo com variância maior que zero

Matriz de Correlação							
		mais4	mais3	mais2	mais1	menos2	menos3
Correlação	mais4	1,000	,655	,378	,255	,293	,683
	mais3	,655	1,000	,577	,389	,447	,447
	mais2	,378	,577	1,000	,405	,516	,258
	mais1	,255	,389	,405	1,000	,313	,174
	menos2	,293	,447	,516	,313	1,000	,200
	menos3	,683	,447	,258	,174	,200	1,000

Note-se, por exemplo, que na primeira linha das correlações a variável “mais4” tem uma tendência a ser mais correlacionada a suas variáveis próximas (como o mais3) e simétricas (menos3) e menos correlacionada com as variáveis centrais. No caso de uma variável central, como o mais1, vemos a tendência a se correlacionar com suas variáveis mais próximas e também centrais, ao passo que possui uma menor correlação com aquelas que estão nas extremidades (mais4 e menos3).

O próximo passo foi agrupar os valores de correlação em fatores, para isso executamos a rotação da matriz apresentada na Tabela 5. Não entraremos nos detalhes matemáticos sobre a rotação de matrizes pelo processo *varimax* por acharmos não pertinentes ao trabalho, no entanto, recomendamos a leitura de Kaiser (1958). A vantagem de uma análise usando uma matriz rotacionada é a de se ter uma maior simplicidade na leitura da união de variáveis. No caso do método *varimax*, por exemplo, tende-se a maximizar as máximas semelhanças e a minimizar as mínimas semelhanças, obtendo assim uma separação mais clara das correlações entre variáveis.

Tabela 5: Matriz rotacionada e cargas fatorais⁴

Rotated Component Matrix^a		
	Factor	
	1	2
mais4		,887
mais3	,596	,608
mais2	,797	
mais1	,700	
menos2	,765	
menos3		,897

Podemos notar, pela Tabela 4 que a análise fatorial nos retornou duas componentes distintas. Chamaremos estes componentes, por ora, de Fator 1 e Fator 2. O Fator 1 tem como característica o fato de agrupar os valores mais centrais ou próximos de zero ao seu redor, tais como mais1, menos2 e mais2. Por outro lado, o Fator 2 tem uma maior dispersão e reúne em seu entorno as variáveis localizadas mais próximas das extremidades, mais4 e menos3. Como pode se ver na Tabela 4, a variável mais3 possui dois valores de carga fatorial, tendo apenas 0,012 de diferença entre uma e outra carga. Podemos considerá-la parte dos dois fatores, isto é, ela seria uma *variável de confusão*, ao passo que podemos considerar a variável mais2 como quasi-categórica. O teste de componentes principais demonstra que as variáveis mais4 e menos3 seriam os *limiares de diferenciação tonal*, isto é, a partir deles existe uma grande probabilidade de o ouvinte dizer que se tratam de dois sons diferentes quando comparados ao valor 0.

Análise – Teste II: 3MFC – 3 Multiple Forced Choice

O objetivo do teste com três alternativas era o de avaliar os pontos nebulosos, isto é, aqueles em que o ouvinte não sabe ao certo se existem ou não diferenças significativas. Note-se que neste ponto da análise, como base no teste de componentes principais, apresentado na Tabela 4, é possível determinar que as variáveis mais4 e menos3 sejam os

⁴ Extraction Method: PrincipalComponent Analysis. Rotation Method: Varimax with Kaiser Normalization. a. Rotation covered in three iterations.

limiars de diferenciação tonal, com a variável mais3 exercendo um papel de elemento de confusão. As Figuras 6 e 7 apresentam os resultados do teste II. As siglas n' , nt , ns e nn significam, respectivamente, número de participantes no teste, número de respostas “talvez”, número de respostas “sim” e número de respostas “não”.

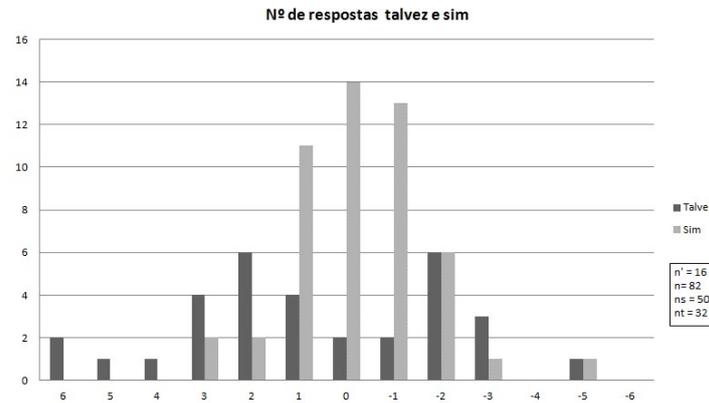


Figura 6: Número de respostas “sim” e “talvez” obtidas no teste 3MFC

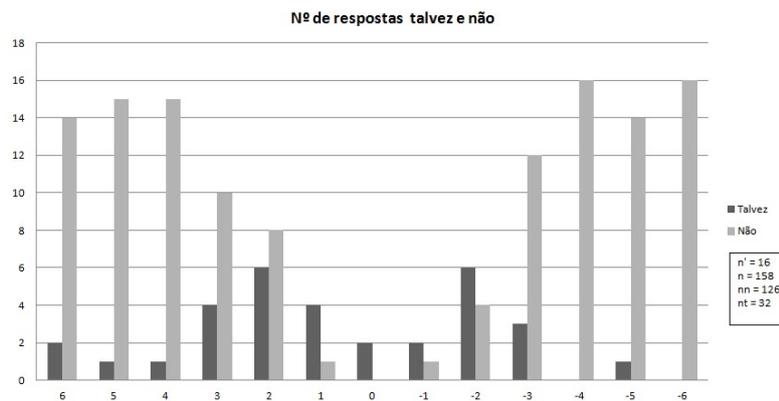


Figura 7: Número de respostas “não” e “talvez” obtidas no teste 3MFC

A Figura 6 demonstra que há uma maior distribuição dos valores de “sim” ($ns = 50$) e um número de respostas “talvez” relativamente baixo ($nt = 32$), ao passo que há um total de 126 respostas “não” ($ns = 126$), totalizando 208 ($16 \cdot 13 = 208$) respostas, como esperado. No teste I, com duas respostas apenas, os números de “sim” e “não” foram de 68 e 140, respectivamente, o que demonstra uma diminuição de 26,4% de respostas no primeiro caso e um total de 10% no segundo caso.

As variáveis que sofrem alterações no quesito da variância, em comparação ao Teste I e apresentadas na Tabela 3, foram as variáveis mais6, mais5, zero, menos1 e

menos5, ou seja, elas obtiveram ao menos uma resposta “talvez”, o que poderia invalidar a noção de que elas são variáveis categóricas e serem vistas como sensíveis às condições do falante. Com o intuito de verificar essa hipótese, aplicamos uma ANOVA entre estas variáveis utilizando o número de respostas “sim” e “não” comparadas ao de respostas “talvez”. O teste nos retornou um $p = 0,33$ ($F = 1,28$, para um $F_{crítico} = 4,46$), para $\alpha = 5\%$ e $n = 96$, ou seja, a quantidade de respostas “talvez” para estas variáveis não foi significativa para nossa análise, de modo que podemos considerá-las categóricas.

Nossa interpretação é a de que a alternativa “talvez”, no caso das variáveis acima elencadas (mais6, mais5, zero, menos1 e menos5), foi utilizada pelos participantes não como uma alternativa de respostas, mas, sim, como uma alternativa compensatória. Analisando-se os tempos de resposta para cada alternativa, nota-se que ao responder “talvez” o participante demorou cerca de 50% a mais de tempo do que nas alternativas “sim” e “não”, cerca de 15 segundos para “talvez” e 10 segundos para “sim” e “não”. Ao nosso ver, isso demonstra uma desatenção ao par apresentado ou uma leve fuga da tarefa, sendo que a maneira de se compensar é utilizando-se da resposta “talvez”.

Entendemos que, no caso das variáveis mais2 e menos2, por exemplo, se tem justamente o oposto, em que os participantes demonstraram ter algum tipo de dúvida com relação a semelhança ou diferenças dos pares. Devido ao fato de que grande parte das demais respostas “talvez” encontrem-se no intervalo que cremos ser o dos limiares de diferenciação tonal, não iremos prosseguir com a aplicação de testes de hipóteses. Por ora, podemos dizer que as variáveis mais2 e menos2 aparentam ser os indícios de variação, isto é, o valor a partir do qual os ouvintes têm a percepção de que há uma diferença, embora ela esteja ainda muito difusa.

Conclusões

Os resultados, por ora obtidos, nos permitem esboçar algumas conclusões. As *variáveis quasi-categóricas por semelhança* (mais2, mais1 e menos2) são aquelas as quais possuem diferenças perceptíveis, embora apenas alguns sujeitos sejam capazes de notá-las. As *variáveis categóricas por diferença* (mais6, mais5, zero, menos1, menos4, menos5 e menos6) formam o conjunto das variáveis com diferenças perceptíveis e percebidas de maneira categórica pelos participantes de nossa pesquisa. A *variável de*

confusão (mais3) é a variável cuja percepção é confundida ora com as variáveis quasi-categóricas, ora como um limiar de diferenciação tonal. Os *limiões de diferenciação tonal (mais4 e menos3)* são os valores por nós pesquisados e representam os limites, a partir dos quais a percepção do ouvinte passa a ser modalizada. Ou seja, para que se note alguma diferença de tom entre dois estímulos é necessário que estes variem, ao menos, quatro semitons ascendentes ou três semitons descendentes a partir do referencial estabelecido.

Este referencial estabelecido é chamado por Ferreira-Netto (2008) de tom médio e representa uma unidade prototípica de valores F_0 , por meio da qual o falante poderá estabelecer os valores de frequência fundamental, bem como as variações por ele previstas ou desejadas. Por definição, o tom médio é o valor de frequência fundamental que o falante irá estabelecer durante o início da sua fonação e percebido pelo ouvinte, as variações que ultrapassem os limiões de diferenciação tonal podem ser compreendidas como foco, por exemplo, para o Português Brasileiro.

Por fim, podemos estabelecer, de acordo com a Tabela 6, os limiões de diferenciação tonal e as demais variáveis:

Tabela 6. Tabela representando os limiões de diferenciação tonal para o Português Brasileiro, como relação a um tom natural

Variáveis categóricas por diferença	mais5, mais6, menos4, menos5, menos6
Limiar de diferenciação tonal	mais4
Variável de confusão	mais3
Variáveis centrais	mais2, mais1, menos1, menos2
Limiar de diferenciação tonal	menos3

Com relação à nomenclatura, existe também uma questão semântica, pois nomeá-las por critérios de semelhança ou diferença nos levaria a criar limiões de limiões, isto é, passaríamos a analisar micro-limiões, pontos mínimos a partir dos quais a variação poderia ser percebida pelo ouvinte. Sendo assim, quanto mais próximos estivermos de um limiar, mais próximos estaremos de um outro limiar ainda não pesquisado, o qual tem uma natureza distinta do primeiro. Obviamente, não chegaremos nunca ao infinito, pois a percepção humana possui algum grau de finitude.

Wertheimer (1983) discute esta questão em seu artigo “Laws of Organization in Perceptual Forms”. Para o autor, a percepção não poderia ser entendida como a coleção de pequenas percepções de inúmeros objetos, mas, sim, um conjunto uno, regulado por certas regras. Por outro lado, os fisiologistas do século XIX, como van Helmholtz (1895), acreditavam no contrário, que embora houvesse uma organização externa, haveria uma lógica intrínseca ao objeto percebido. A proposta adequada para nosso trabalho parece ser a de Patricia Kuhl (KUHL, 1991; IVERSON; KUHL, 1995; KUHL *et al.*, 2006). Nesta abordagem, haveria formas prototípicas, as quais por uma questão de gradiência iriam se tornando “não prototípicas”, mas mantendo alguma conexão perceptiva com aquela considerada prototípica.

Por este motivo, agrupamos as variáveis *quasi-categóricas*, *mais2*, *mais1* e *menos2* e a variável categórica *menos1* como variáveis centrais, por compreendermos que elas estão sob o *efeito magnético de percepção* (KUHL, 1991; IVERSON; KUHL, 1995). Assim definimos por variáveis *quasi-categóricas*: *mais2*, *mais1*, *menos1* e *menos2*. Entendemos que estas variáveis estão, de fato, sob a influência do efeito magnético de percepção (PME), isto é, a percepção delas tende a ser categorizada de acordo com os fatores de proximidade e similaridade (WERTHEIMER, 1938). Dada a proximidade e a semelhança com o tom neutro (zero), tendem a ser compreendidas como esta, embora elas tenham de 6% a 13% de variação acústica, este fator não afetou de maneira absoluta a percepção dos tons. Neste sentido, não vimos a necessidade de distingui-las entre variáveis *quasi-categóricas* e *categórica*; por esta razão, as nomearemos unicamente de *variáveis centrais* e incluiremos nesta categoria a variável *menos1*.

Aparentemente, a duração tem um papel importante no que se refere à percepção do tom. Os valores encontrados corroboram os encontrados por Consoni (2011) para palavras isoladas. Podemos supor um espectro de frequências não-linear que varie dos +4 semitons aos +3 semitons e dos -4 semitons aos -3 semitons, a partir do tom médio, cujo limiar de diferenciação tonal seja uma variável do tempo. De acordo com Shigeno (1986), a distância temporal entre dois estímulos afeta, de fato, a percepção tonal: quanto menor o tempo entre dois tons, menor a frequência do segundo tom é percebida pelos sujeitos.

Em nosso entender, uma maior janela temporal para a percepção tonal permite uma refinação na acurácia, o que é levado para a tarefa de responder se há ou não diferenças entre os pares tonais. Embora os estudos que tomem isso como objeto levam

em conta apenas a distância temporal entre estímulos, acreditamos que podemos melhor definir esse tema. Por um viés fisiológico, encontramos respostas na comparação entre a teoria tonotópica e a teoria temporal. Para os primeiros, a percepção de frequências é determinada por espaços definidos na membrana basilar, enquanto a segunda afirma que a percepção da frequência é baseada no tempo em que os neurônios respondem a um som na cóclea, a variação destas janelas temporais altera o processo de percepção (MOORE, 2005).

Uma segunda hipótese refere-se ao fato de que os estímulos utilizados em Consoni (2011) teriam uma referência fixa para todos os estímulos, o tom da frase portadora. Neste sentido, os participantes da pesquisa da autora tinham uma percepção abrupta da variação tonal, ao passo que nossos participantes poderiam realizar o julgamento durante o tempo de realização da frase. Roederer (2003) sugere que variações súbitas são detectadas com uma *j.n.d.* mais baixa do que aquelas percebidas em um tempo normal de processamento. Isto é, dado um pulso com um tom invariável, a primeira variação súbita será detectada, mesmo que esta variação seja pequena em relação ao tom original. Contudo, esta percepção de variação súbita tende a diminuir conforme os tons se tornam mais graves. Roederer (2002) sugere que a resolução para tons em torno de 500 Hz é de 3%, ao passo que, para tons em torno de 2000Hz, esta resolução é de apenas 0,5%. Dessa forma, a diminuição do valor do limiar de diferenciação tonal inferior pode ser explicada pelo fato de que no caso da percepção abrupta, os participantes têm uma pior definição na percepção do tom grave, o que leva à diferença entre os dados aqui encontrados e os encontrados em Consoni (2011).

Nesta linha de raciocínio, podemos sugerir que os valores de +3 e -3 semitons são os limiares inferiores dessa função para janelas temporais menores, ao passo que os valores de +4 e -4 semitons são os limiares para janelas temporais maiores.

Os próximos passos desta pesquisa são: (i) analisar os valores temporais mínimos para a modalização de frequência, de acordo com os valores do limiar de diferenciação tonal. (ii) Além disso, pretendemos estabelecer um modelo matemático, o qual preveja essas variações de frequência tomando a variação temporal (∂_t) como função relativa da variação dos limiares tonais.

MARTINS, Marcus Vinicius Moreira; NETTO, Waldemar Ferreira. The tonal differentiation thresholds of Brazilian Portuguese. **Revista do Gel**, v. 14, n. 2, p. 157-182, 2017.

Abstract: *The purpose of this work is to determine the tonal differentiation thresholds (THART, 1981) in Brazilian Portuguese. The thresholds are points from which speakers are able to identify significant differences in tonal variation, leading them to consider that point as endowed with some relevant linguistic information, such as an emphasis on a word or syllable. To define the Brazilian Portuguese tonal differentiation thresholds we applied two perception experiments with 13 pairs of two sentences spoken by a male voice. The pairs consisted of a neutral sentence paired with another twelve, those second sentences had their F0 manipulated into six positive and six negative semitones from the original values. In Test I, 16 participants, without musical training, answered whether or not they noticed any difference between the presented pairs. In Test II, the same 16 participants could answer yes, no or maybe, the goal of the latter alternative was to evaluate fuzzy points. To determine the tonal differentiation thresholds we used a principal components analysis (PCA) model. Accordingly to the results, the Brazilian Portuguese tonal differentiation thresholds are -3 and +4 semitones, which indicates that speakers are unable to identify differences prior to these values.*

Keywords: *Phonetic. Phonology. Perception. Intonation. Acoustics.*

Submetido em: 15/11/2016.

Aceito em: 19/06/2017.

Referências

BOERSMA, P.; WEENINK, D. **Praat:** doing phonetics by computer (Version 5.1.05) [Programa de Computador]. Disponível em: <<http://www.praat.org/>>. Acesso em: 10 jan. 2016.

BOGACZ, R. *et al.* The physics of optimal decision-making: A formal analysis of models of performance in two-alternative forced-choice tasks. **Psychological Review**, v. 113, n. 4, p. 700-765, 2006.

BOLINGER, D. L. On intensity as a qualitative improvement of pitch accent. **Lingua**, v. 7, p. 175-182, 1958.

_____. Intonation: Level vs. configuration. **Word**, v. 7, n. 3, p. 191-210, 1961.

CHERRY, E. C. Some Experiments on the Recognition of Speech, with One and with Two Ears. **The Journal of the Acoustical Society of America**, v. 25, n. 5, p. 975-979, 1953.

CONSONI, F. **Aspectos da proeminência tonal em Português Brasileiro**. 2011. 119 f. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011.

COZBY, P. **Métodos de Pesquisa em Ciências do Comportamento**. São Paulo: Editora Átlas, 2003.

DANCEY, C. P.; REIDY, J. **Estatística Sem Matemática para Psicologia**. Porto Alegre: Artmed – Bookman, 2006.

ERDFELDER, E.; FAUL, F.; BUCHNER, A. Gpower: A general power analysis program. **Behavior Research Methods, Instruments, & Computers**, v. 28, p. 1-11, 1996.

FERREIRA-NETTO, W. Decomposição da entoação frasal em componentes estruturadoras e em componentes semântico-funcionais. In: **X Congresso Nacional de Fonética e Fonologia / IV Congresso Internacional de Fonética e Fonologia**. Niterói: UFF, 2008.

FRY, D. B. Duration and intensity as physical correlates of linguistic stress. **Journal of the Acoustical Society of America**, v. 27, p. 765-768, 1955.

_____. Experiments in the perception of stress. **Language and Speech**, v. 1, p. 126-152, 1958.

_____. The dependence of stress judgments on vowel formant structure. **Proceedings of the Fifth International Congress of Phonetic Sciences** Basel/New York: Karger, 1965. p. 306-311.

HALLE, M. Speculations about the representations of Words in memory. In: _____. **From memory to speech and back: papers on Phonetics and Phonology 1954-2002**. Nova Iorque: Mouton de Gruyter, 2002. p. 122-149.

HELMHOLTZ, H. L. F. **On the sensations of tone as a physiological basis for the theory of music**. Londres: Longmans, Green and Co., 1895.

HILL, C. M. On choice. **The American Journal of Psychology**, v. 9, n. 4, p. 587-590, 1898.

IVERSON, P.; KUHL, P. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. **Journal of Acoustical Society of America**, v. 97, p. 553-562, 1995.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v. 23, n. 3, p. 187-200, 1958.

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

KUHL, P. *et al.* Language/Culture/Mind/Brain. Progress at the Margins between Disciplines. **Annals of the New York Academy of Sciences**, v. 935, p. 36-74, 2001.

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

LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, v. 140, p. 1-55, 1932.

MOL, H. G.; UHLENBECK, G. M. The linguistic relevance of intensity in stress. **Lingua**, v. 5, p. 205-213, 1956.

MOORE, B. C. J. **An introduction to the psychology of hearing**. Londres: Academic Press, 2005.

OHALA, J. Speech perception is hearing sounds, not tongues. **The Journal of Acoustical Society of America**, v. 99, n. 3, p. 1718-1725, mar. 1996.

PIERREHUMBERT, J. **The Phonology of Phonetic of English Intonation**. 1980. 401 f. Tese (Doutorado em Linguística) – Massachusetts Institute of Technology, Cambridge, Massachusetts, 1980.

PIKE, K. L. **The intonation of American English**. Michigan: Ann Arbor: University of Michigan Press, 1945.

RUSSELL, B. A análise da matéria (cap. xvii a xx). In: _____. **Ensaio Escolhidos**. São Paulo: Ed. Abril, 1978 (Os Pensadores). p. 161-202.

REA, L. M.; PARKER, R. A. **Metodologia de pesquisa: do planejamento à execução**. São Paulo: Editora Pioneira, 2000.

ROEDERER, J. **Introdução à física e psicofísica da música**. São Paulo: Edusp, 2003.

SCHOUTEN, M. (Ed.). NATO Advanced Research Workshop on "The Psychophysics of Speech Perception". **North Atlantic Treaty Organization**. Utrecht: Kluwer Academic Publisher, 1987.

SHIGENO, S. The auditory tau and kappa effects for speech and nonspeech stimuli. **Perception & Psychophysics**, v. 40, p. 9-19, 1986.

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

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

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

Van KATWIJK, A. **Accentuation in Dutch: An Experimental Linguistic Study**. Assen: Van Gorcum, 1974.

WERTHEIMER, M. Laws of organization in perceptual forms. In: ELLIS, W. D. (Ed.). **A source book of Gestalt psychology**. London: Routledge & Kegan Paul, 1938.

ZACKSENHOUSE, M.; BOGACZ, R.; HOLMES, P. Robust versus optimal strategies for two-alternative forced choice tasks. **Journal of Mathematical Psychology**, v. 54, p. 230-246, 2010.

ZWICKER, E.; FLOTTORP, G.; STEVENS, S. S. Critical band width in loudness summation. **The Journal of the Acoustical Society of America**, v. 29, n. 5, p. 548-557, 1957.