



**ASPECTOS DA ENTOAÇÃO NA FALA DE PACIENTES  
COM ESQUIZOFRENIA  
ASPECTS OF INTONATION IN SPEECH OF  
SCHIZOPHRENICS PATIENTS**

*Waldemar Ferreira Netto<sup>1</sup>*  
*Marcus Vinícius Moreira Martins<sup>2</sup>*  
*Ana Cristina Aparecida Jorge<sup>3</sup>*

**RESUMO**

O objetivo deste artigo é desenvolver uma metodologia capaz de analisar se a informação entoacional produzida por falantes com esquizofrenia apresenta características particulares capazes de especificar a fala destes indivíduos. É reportado na literatura médica que pacientes com este transtorno tendem a produzir a entoação de forma específica, seja pela sua falta (aprosódia), seja pelo seu excesso (hiperprosódia). A caracterização dessa entoação pode ser importante facilitador para o diagnóstico da esquizofrenia. Para este fim, realizamos um experimento com 10 pacientes com esquizofrenia e 10 sujeitos-controle, todos do sexo masculino e com idade e escolaridade pareadas. A extração dos dados foi feita pelo aplicativo ExProsodia, tomando exclusivamente a entoação para determinar as fronteiras prosódicas frasais, por esta razão o critério sintático não foi levado em consideração. Após a segmentação, procedemos com o cálculo da quantidade de informação veiculada exclusivamente pelos parâmetros entoacionais, para isso calculamos a entropia de Shannon, para cada uma das frases analisadas. Os resultados indicam diferenças significativas entre os grupos, de modo que o valor de entropia na fala de sujeitos com esquizofrenia foi significativamente menor do que a do grupo de controle:  $t_c(1,81) > t_o(-3,04)$ ,  $p < 0,01$ . A partir desses resultados, propomos que a entoação da fala de sujeitos com esquizofrenia se caracteriza de forma bastante homogênea, com forte tendência à diminuição na quantidade de informação que poderia manifestar por meio da variação de F0.

**PALAVRAS-CHAVE:** Esquizofrenia; Prosódia; Entoação; Análise automática; Entropia.

1 Professor Titular, Departamento de Letras Clássicas e Vernáculas, Universidade de São Paulo, São Paulo, Brasil. E-mail: wafnetto@usp.br

2 Docente, Departamento de Letras e Literatura, Universidade do Estado de Minas Gerais, Belo Horizonte, Brasil. E-mail: marcus.martins@uemg.br

3 Pós-graduanda, Departamento de Letras Clássicas e Vernáculas, Universidade de São Paulo, São Paulo, Brasil. E-mail: anacristinajorge@usp.br

## ABSTRACT

This article aims to develop a methodology capable of analyzing whether the intonation information produced by speakers with schizophrenia presents characteristics capable of specifying these individuals. It is reported in the medical literature that patients with this disorder tend to produce intonation in a specific way, either due to the lack (aprosody) or its excess (hyperprosody) of prosody. The characterization of this intonation can be an important support for the early diagnosis of schizophrenia. For this purpose, an experiment was conducted with 10 patients with schizophrenia and 10 control subjects, all male, and with matched age and education. Data extraction was done by the ExProsodia application, exclusively using intonation to determine the phrasal boundaries, for this reason, the syntactic criterion was not considered. After segmentation, we proceeded with the calculation of the amount of information conveyed exclusively by the intonation parameters, for that we calculated the Shannon's entropy for each of the analyzed sentences. The results indicate significant differences between the groups, such that the entropy value in the speech of subjects with schizophrenia was significantly lower than that of the control group:  $t(1.81) > t(-3.04)$ ,  $p < 0.01$ . Based in these results, we propose that the speech intonation of subjects with schizophrenia is characterized in a very homogeneous way, with a strong tendency to decrease the amount of information that could be expressed through the F0 variation.

**KEYWORDS:** Schizophrenia; Prosody; Intonation; Automatic Analysis; Entropy.

## Introdução

O propósito deste artigo é descrever, a partir do conceito de entropia, aspectos entoacionais que possam caracterizar a fala de sujeitos com esquizofrenia. Do ponto de vista da Teoria da Comunicação, que utilizamos aqui, Shannon (1949) tomava a entropia como medida de imprevisibilidade na escolha de eventos previamente conhecidos, assumindo que o limites máximo denotaria escolhas totalmente aleatórias e o mínimo, estagnação e ausência de variedade. Na literatura médica e linguística, a entoação de pacientes esquizofrênicos é comumente descrita como não dotada de variações melódicas, característica tratada como *aprosódia* (COMPTON et al., 2018; COVINGTON et al., 2005; ROSS, 1981). A etiologia e as causas desse fenômeno são ainda motivo de investigação por médicos e neurobiólogos (MITCHELL; CROW, 2005). Entende-se que a doença causa uma perda de capacidades cognitivas, incluindo a linguagem, de modo que a supressão melódica da fala estaria ligada a este sintoma. Parola e seus colegas (2020), fazendo uma revisão da literatura relacionada a atipias prosódicas entre sujeitos com esquizofrenia, verificaram que o *plan affect* não era a principal atipia atuando como possível marcadora da esquizofrenia. Lado a pausas e velocidade de fala, variações entoacionais eram mais heterogêneas e menos constantes entre pessoas com esquizofrenia. Matínez-Sanchez e seus colegas (2015), numa das pesquisas analisadas em Parola e colegas (2020), verificaram que, utilizando exatamente esses mesmos aspectos — pausa, velocidade e variações entoacionais, todas mostraram diferenças significativas entre na comparação de pessoas com esquizofrenia e grupo de controle. Por sua vez, a pesquisa empreendida por Santos (2012) e por Santos e seus colegas (2014), com falantes de português do Brasil com esquizofrenia, contrastivamente com grupo de controle, verificou que alterações

ocorreram nas “tarefas de evocação lexical com critério semântico, atos de fala indiretos, discurso conversacional e discurso narrativo” eram maiores do que as que “ocorreram nas componentes prosódicas no nível de compreensão, destacando-se a prosódia emocional” (p. 1285-1286). Essas pesquisas apontam para a necessidade de uma análise mais pormenorizada da entoação na fala de pessoas com esquizofrenia, sobretudo falantes do português do Brasil, conforme já salientara Behlau (2001).

Nossa hipótese é a de que seja possível mensurar essas perdas por meio do cálculo da entropia de informação veiculada pela entoação desses pacientes. Dessa forma, os objetivos deste artigo são (i) apresentar um modelo de análise da entoação baseado no conceito de entropia da informação de C. Shannon (1948) e, a seguir, (ii) analisar a fala de pacientes esquizofrênicos a partir deste modelo.

## **Esquizofrenia**

O *Diagnostic and Statistical Manual of Mental Disorders* (AMERICAN PSYCHIATRY ASSOCIATION, 2013) caracteriza a esquizofrenia por um misto de sinais e sintomas disformes. Para o seu diagnóstico, o indivíduo deve exibir pelo menos dois dos seguintes sintomas, sendo que pelo menos um deles deve ser um dos três primeiros: delírios, alucinações, discurso desorganizado, comportamento grosseiramente desorganizado ou catatônico e sintomas negativos, permanentes durante o período de um mês, com alguns sinais deste transtorno evidentes durante 6 meses. No Brasil, o Ministério da Saúde (BRASIL, 2013b) define a esquizofrenia como uma patologia constituída por um grupo de distúrbios mentais graves, sem sintomas patognomônicos. Pode ser caracterizada pela presença de distorções do pensamento e da percepção, por inadequação e embotamento afetivo sem prejuízo na capacidade intelectual, embora, com a prevalência, ocorram danos cognitivos e laborativos. Os primeiros sintomas aparecem entre o final da adolescência e início da idade adulta. Dados estatísticos demonstraram que 0,6 a 3,0% (dependendo dos critérios adotados) da população nacional a desenvolve. Ela é responsável por 20% das internações psiquiátricas, independente de classe social, cultura e etnia, sem distinção entre homens e mulheres. Esse mesmo compêndio psiquiátrico descreve a redução da entoação da fala como um dos possíveis sintomas negativos da esquizofrenia. Porém, destaca-se que esse é um fator difícil de ser analisado devido às diversas variações linguísticas e aos diferentes modos de expressar a emoção (BRASIL, 2013b).

Cooper e Sartorius (1977) propuseram que a forma aguda da esquizofrenia sempre existiu como uma doença mental em todas as culturas e que, em uma proporção considerável de indivíduos vulneráveis, a forma aguda tendia à cronicidade. No entanto, segundo eles, seu reconhecimento deveu-se a um conjunto de mudanças sociais decorrentes da industrialização. Para os autores, os avanços que acompanharam a tecnologia médica levaram à separação dos doentes mentais e dos fisicamente doentes e estabeleceram um cenário propício para a rejeição

e estigmatização dos afetados (COOPER; SARTORIUS, 1977, p. 52). Essa consequente concentração maior de esquizofrênicos em instituições especializadas, decorrente da rejeição pela comunidade, resultou na possibilidade de psiquiatras terem de lidar com um número substancial de pacientes, tornando viável o reconhecimento de traços comuns a todos eles.

Bebbington e seus colegas (2016), a partir de investigação que cotejou dados publicados em anos anteriores com uma pesquisa em duas fases, uma de triagem seguida de uma avaliação clínica um subconjunto de 6% de participantes, verificaram que, embora os testes estatísticos tenham indicado um aumento significativo nas taxas de ocorrência de psicose, de 0,4% em 2007 para 0,7% em 2014, esses números são consistentes com uma tendência contínua de ampla estabilidade nas taxas de psicose. Dessa maneira, ao estabelecer um valor médio para conjunto de dados, a prevalência global do transtorno psicótico na Inglaterra, em 2017, teria sido de 0,5% da população geral adulta. Segundo eles, é provável (95% de confiança) que, se todos na população domiciliar tivessem sido testados, a taxa estaria entre 0,4% e 0,7%. (BEBBINGTON *et al.*, 2016, p. 139)

A incidência da esquizofrenia, segundo Crow (2000), poderia ser explicada por meio das bases neurais da linguagem. Crow (2004) considerou que o conjunto de sintomas apontados poderia ser diluídos em apenas um: os distúrbios de linguagem. Sua hipótese admitia que, dentre outros fatores linguísticos, as alucinações auditivas poderiam ser tomadas como uma anomalia de percepção da palavra falada. Ceccherini-Nelli *et al.* (2007) associaram sintomas nucleares linguísticos decorrentes desse transtorno a um índice de desconexão no hemisfério esquerdo, somada a outra disfunção no hemisfério direito referente a déficits de atenção e imputação de associações. O estudo realizado por Park *et al.* (2008) verificou que pacientes com esquizofrenia apresentaram redução significativa na amplitude de *pitch* em comparação aos participantes bipolares e saudáveis, independentemente do contexto afetivo, e esse intervalo de *pitch* reduzido não se deveu à menor frequência basal. Esses dados corroboraram as hipóteses de Crow (2000, 2004) e de Ceccherini-Nelli *et al.* (2007). Crow (2008) desenvolveu essa hipótese, propondo que a esquizofrenia possuiria uma predisposição genética marcada por um evento epigenético envolvendo ambos os cromossomos sexuais X e Y, o que sugeria, além de alguma alteração na estrutura cerebral, também uma aneuploidia cromossômica, especificamente na região do bloco homóloga Xq21.3/Yp11.2.

Behlau (2001) destacou que não há estudos extensos que investigam a expressão emocional através da entoação na esquizofrenia. A autora caracterizou a voz na esquizofrenia através de uma típica impressão de melancolia, associada com uma frequência fundamental constante e com entoação limitada. A dificuldade de realizar pesquisas residia na possibilidade de uma população de sujeitos adequados e a presença de metodologia precisa para avaliar tais dados. Leitman *et al.* (2010) e Mota *et al.* (2014) propuseram que o estudo focalizado nas variáveis linguísticas poderia auxiliar na composição de diagnósticos mais acurados para esse transtorno mental. Esta também foi a hipótese de Santos (2012), ao retratar que indivíduos

acometidos por algum transtorno mental, em especial, pela esquizofrenia poderiam apresentar déficits singulares em quatro níveis comunicativos: léxico-semântico, discursivo, pragmático e no processamento prosódico. Em sua pesquisa, verificou que os sujeitos com esquizofrenia apresentaram dificuldades em compreender e repetir as entoações linguísticas (afirmação, interrogativa e imperativa) e emocionais (alegria, tristeza e raiva).

### **Análise da entoação**

Desde o século XVIII, com o trabalho de Steele (1779), as fronteiras de frases entoacionais são reconhecidas por sua carga informacional na fala. Mas, foi, principalmente, a partir do século XX que critérios para a avaliação dessa carga informacional foram estabelecidos. A noção de *shift* — mudança tonal ascendente que ocorre entre o tom final de uma frase entoacional e o tom inicial da frase seguinte — foi objeto de consideração de vários autores (PRONOVOST, 1939; FAIRBANKS, 1940; SNIDECOR, 1943, 1944, 1951). Esses autores, ainda que mostrassem algumas divergências em seus resultados, verificaram que valores de *shift* dificilmente alcançariam intervalos maiores do que nove semitons, mas se manteriam em um intervalo em torno de cinco semitons ascendentes.

O reconhecimento da pausa como marcador de fronteira de frase entoacional também foi objeto de vários estudos (HOCKET, 1942; SNIDECOR, 1944; VERZEANO E FINESINGER, 1949; HEARGRAVES E STARKWEATHER, 1959; BOOMER E DITMANN, 1962). Esses autores verificaram que pausas com duração maior do que 500 milissegundos eram sempre reconhecidas em seus experimentos, mas pausas com duração menor — de até 150 milissegundos — também poderiam ser reconhecidas. Essa diferença foi tomada como duas categorias em que as pausas breves caracterizariam fronteiras de frases entoacionais e pausas longas caracterizariam, também, hesitações.

A partir do conceito de *breath-group* (JONES, 1922; STETSON, 1927), autores diversos propuseram variações globais para a definição de frase entoacionais. Os conceitos de declinação ou de *baseline* envolviam uma curva entoacional descendente global na entoação frasal (LIEBERMAN, 1967; COHEN E T'HART, 1967; VAISSIÈRE, 1974, 1983; MAEDA, 1976; PIERREHUMBERT, 1979). Essa variação global estabeleceria que as variações pontuais na entoação deveriam ser consideradas relativamente a uma linha de referência tonal descendente que ocorreria conforme a locução se desenvolvesse. Finda a frase entoacional, haveria a retomada do tom inicial, desencadeando um processo a que se chamou *reset* (MAEDA, 1967). Alguns autores (COOPER E SORENSEN, 1977; VAISSIÈRE, 1986) procuraram desvincular essa variação global descendente da noção de *breath-group* da fala, atribuindo-a a uma opção do falante. Ladd (1984) fez uma abordagem detalhada de todas essas propostas.

A correlação entre parâmetros para o estabelecimento de marcadores de fronteira de frases entoacionais foi objeto de várias pesquisas no decorrer do século XX (SNIDECOR, 1944; TRAGER E SMITH, 1957; GARVIN E MATHIOT, 1958, DANÉŠ, 1960; DITTMANN E LLEWELLY, 1967; BOLINGER, 1989). De maneira geral, as variações entoacionais foram associadas a pausas. As pausas, entretanto, mostraram-se como marcas predominantes na definição de juntura frasal (TRAGER E SMITH, 1957; DITTMANN E LLEWELLY, 1967), sobretudo por conta das possibilidades de variação ascendente ou descendente (JONES, 1922; TRUBETZKOY, 1973) em finalização de frases entoacionais. Considerando outros parâmetros, alguns pesquisadores (PRICE ET AL., 1989; CAMPBELL, 1983) verificaram que o alongamento silábico ou segmental final de frases também poderia correlacionar-se como marcador de fronteira, em conjunto com a variação tonal e a presença de pausas.

Baseado nesses critérios, Campbell (1993) verificou que a duração dos segmentos silábicos como uma só unidade variava conforme sua posição sentencial final ou não e que a duração dos segmentos silábicos internos variava em relação à sua posição em fronteira sintagmática. Taylor (1994) procurou estabelecer um modelo de análise prosódica automática baseando-se especialmente na variação tonal. Para tanto, desenvolveu uma rotina de análise que analisava porções de 50ms de  $f_0$  — suavizado (*smoothed*) principalmente quanto às interrupções causadas pela ocorrência de segmentos não vozeados — considerando as variações tonais ascendente, descendente ou nula.

Ao desenvolver um modelo de análise automática da entoação, Taylor (2000, p. 1698) propôs que *eventos entoacionais* se caracterizassem na cadeia entoacional [*intonational stream*] como *pitch accents*, *boundary tones* e uma sequência contígua de ambos: *pitch accents* seriam os deslocamentos de  $f_0$  usados pelo falante para dar algum grau de ênfase a alguma sílaba ou palavra e *boundary tones* também seriam deslocamentos de  $f_0$ , mas para marcar fronteiras de sintagmas, bem como para marcar efeitos de afirmação, continuação ou de questionamento. Segundo ele, eventos entoacionais não poderiam se interromper e, portanto, teriam de ser tratados como uma unidade. Embora propusesse que eventos entoacionais fossem unidades alinhadas com sílabas ou palavras, o autor propôs que, diferentemente da cadeia segmental, a cadeia entoacional fosse um fenômeno contínuo, mesmo na sua relação com os possíveis significados que poderia apresentar. Saliente-se que, nesse modelo de análise automática, intervalos entre eventos entoacionais foram considerados somente como referências para os deslocamentos de  $f_0$ .

A discussão de Taylor (2000) quanto à impossibilidade de se tratar as variações de  $f_0$  de uma argumentação convincente até então, parece não justificar a crítica que faz às pesquisas que demonstraram que a entoação poderia ser tratada por padrões pontuais fixos — ou até pelo modelo de *perceptual magnet effect* (KUHL, 1991), como o fez para a cadeia lexical. As fronteiras semânticas associadas a padrões entoacionais que ele alegou não terem sido demonstradas, ao contrário, podem ser encontradas, por exemplo no português, no trabalho

de Cagliari (1981) ou, posteriormente, no de Moraes (2007). Ao que parece, a diferença entre as cadeias lexical (segmental) e prosódica não poderia se resolver somente pelo aspecto da arbitrariedade e da continuidade, ainda que se possa encontrar essa diferença entre elas. A cadeia prosódica é contínua, na medida em que tem uma fonte contínua de alimentação (FERREIRA NETTO et al., 2017). Os resultados dessa mesma pesquisa mostraram que tais unidades podem ser padronizadas por diferentes variações de qualidade e, assim, serem tratadas descontinuamente. No entanto, a cadeia prosódica, seja no que diz respeito à entoação ou ao ritmo, pode ter um uso contínuo, em que os “significados” de suas unidades sonoras variam concomitantemente — isso é bem claro na entoação emocional (mais medo-mais tensão e, portanto, mais agudo e maior taxa de enunciação; menos medo-menos tensão e, portanto, menos agudo e menor taxa de enunciação) — mas não em expressões padronizadas (e talvez por isso mesmo) como a manifestação de dúvida, de surpresa etc. De certa maneira, o autor deixou que essas características prosódicas escapassem de sua atenção.

### **Entoação e entropia**

Pierrehumbert (1980, p.59) destaca que a análise da entoação deve se preocupar, inicialmente, com a gramática e as regras de combinação tonal que definem contornos melódicos e estruturam os constituintes prosódicos, como as fronteiras prosódicas, as proeminências tonais e o *pitch accent*. Nesta perspectiva, o caráter informacional da entoação, isto é, as variações de *pitch* que contribuem exclusivamente com informações semântico-pragmáticas, precisam ser levadas em consideração. Essa abordagem parece ser recorrente em maior ou menor grau nos modelos baseados no método de *close-copy*, como os de Pierrehumbert (1983), t’Tart; Collier; Cohen (1990), Silverman *et al.* (1992), Ladd (1996) e Hirst; Cristo (1998), nos quais há uma tendência a simplificar a informação entoacional, buscando, em partes, demonstrar a estrutura dos contornos de  $f_0$  subjacente aos constituintes prosódicos.

Xu (2005) sugere que, para além dessas funções, a entoação também tem a capacidade de codificar significados comunicacionais, centrados no falante ou no ouvinte, por meio de operações fonéticas primitivas. Para o autor, a entoação operaria em dois níveis, no primeiro deles teríamos o significado comunicativo a ser transmitido, no segundo estariam os recursos a serem utilizados na transmissão desse significado. Neste caso, não se trata somente de uma operação correlata a outra, ou seja, mapear os correlatos acústicos não seria suficiente, uma vez que essas medidas podem encobrir operações mais abstratas que ocorrem em um nível intermediário, em que se codificam significados comunicativos em fenômenos acústicos.

Em nosso entender, o objetivo de uma análise da entoação seria compreender os modos de funcionamento destes dois níveis. Além disso, se faz necessário considerar um terceiro nível de análise, o que compreende a codificação dos parâmetros acústicos na construção do significado comunicativo. Desta forma, os recursos acústicos estão relacionados a primitivos

fonéticos da produção de  $f_0$ , ao passo que a análise do conteúdo desta informação veiculada seria do domínio da semântica ou pragmática. Com relação ao terceiro nível, o de codificação, entendemos que uma análise utilizando pressupostos da teoria da comunicação e da transmissão de informação possam ser válidas. Neste artigo, utilizaremos uma medida de entropia como parâmetro de teste.

Transpor um domínio comunicativo para um domínio linguístico é supor que a mensagem é afetada diretamente pelo código, isto é, que a forma como o código se estabelece afeta o modo como a mensagem é produzida pelo emissor (recursos fonéticos) e é percebida pelo falante (significados comunicacionais). Desta forma, um código de tamanho  $n$  permite um total de combinações  $n!$ , enquanto um código  $m > n$ , permite um número  $m!$  de combinações. O que se tem, portanto, é que a extensão combinatória do código é uma medida relativa da capacidade de informação a ser transmitida, ou seja, a informação é o grau de liberdade que se tem ao selecionar um código. No entanto, esse grau de liberdade é limitado pelas combinações impossíveis ou sem sentido, também chamadas de redundâncias, a entropia mede justamente o grau de aleatoriedade que se tem ao construir uma informação (SHANNON, 1948).

Shannon (1948) define a entropia (H) em seu Teorema 2 representado na equação (1):

$$H = -K \sum_{i,j} p_i \log p_i \quad (1)$$

Onde, temos que  $K$  é uma constante definida para tornar o resultado maior que zero e  $p$  é a probabilidade do sistema estar no ponto  $i$  em seu espaço de fase.

A entropia pode ser definida como a quantidade de informação que pode ser produzida e transmitida por uma fonte a partir da combinação dos símbolos disponíveis em um código. Por ser um modelo probabilístico, a entropia leva em consideração a probabilidade de um dado conjunto de elementos se organizar uma sequência. Deste modo, uma sequência invariável do símbolo  $\phi$ , terá entropia igual a zero, visto que a probabilidade será sempre igual a 1 ( $\log \log 1 = 0$ ). De outra forma, a entropia máxima será aquela em que são sempre iguais, ou seja, no inverso do espaço amostral  $H_{\max} = \log \log 1/n$ , quando a probabilidade de ocorrência de qualquer símbolo é igual a 0,5.

Em Ferreira Netto e Martins (2017) demonstramos que a entoação de falantes esquizofrênicos tende a suprimir diversas variações de  $f_0$  presentes na fala normal. Nossa hipótese é a de que a tentativa de tornar sua fala “normal” leva este falante a supor uma total ausência das variações de *pitch* e a produzir aquilo que chamamos de simulacro de fala neutra, ou seja, uma curva entoacional destituída das variações mínimas de  $f_0$ . Diversos estudos, como os citados no início desta seção, demonstram que mesmo a fala normal é carregada de variações mínimas de  $f_0$ , sejam elas derivadas de regras fonológicas subjacentes, sejam elas por variações intrínsecas da frequência fundamental.



Nesse sentido, pretendemos avaliar, por meio do cálculo de entropia, como essa redução na variação de  $f_0$  pode afetar a quantidade de informação veiculada pela fala esquizofrênica. Shannon (1948) demonstra que a entropia de uma fonte pode ser dada pela média ponderada das entropias  $H$  de acordo com a probabilidade de ocorrência delas em cada um dos estágios. O valor  $n$  por nós indicado busca justamente estabelecer esse valor ponderado, baseando-se no número de unidades entoacionais que ocorrem no decurso de uma sentença. Caso os parâmetros sejam equivalentes teremos , ou seja, um estado no qual a entropia é mínima, bem como quando  $p$  for o inverso de  $n$  teremos a máxima entropia.

## Coleta de Dados

Para essa pesquisa, foram usadas informações coletadas em duas instituições de saúde mentais diferentes: uma localizada em um bairro da Capital do Rio de Janeiro e outra situada no interior do Estado de São Paulo. O mesmo procedimento descrito abaixo foi adotado para ambas as instituições. As gravações ocorreram em sala reservada sem isolamento acústico, em que estavam presentes a psicóloga responsável pelo experimento e um participante, por vezes, um responsável legal e/ou um profissional da equipe das instituições. O roteiro executado foi o seguinte: (i) uma entrevista baseada em um questionário semiestruturado, em que se buscavam informações relativas ao contexto de vida, rotina laboral e acadêmica. Na sequência, solicitou-se que os participantes (ii) descrevessem uma atividade artística. Dada a dificuldade que alguns apresentaram para a realização dessa tarefa, solicitou-se que falassem a respeito da Figura 1, disponível na internet, ou que descrevessem algum filme. Este procedimento foi mais evidente no caso dos sujeitos do grupo de controle. Em seguida, os participantes eram convidados a (iii) narrar uma experiência pessoal feliz e uma triste. As gravações finalizavam com a solicitação de (iv) leitura de um trecho narrativo de história infantil (BAUM, 2011, p. 95), o texto pode ser visto no anexo I. Apenas um participante não quis realizar essa última etapa.

**Figura 1.** Imagem utilizada na coleta de dados. Era solicitado aos participantes que descrevessem a imagem.



Fonte: [www.1freewallpapers.com](http://www.1freewallpapers.com)

As gravações da voz foram realizadas em etapas sucessivas individualmente com cada participante. O tempo de duração da coleta de dados variou entre 10 minutos e uma hora conforme a necessidade de cada partícipe. De maneira geral, os pacientes com esquizofrenia

sentiam maior necessidade de expressar-se e, assim, demoravam-se mais. O equipamento utilizado foi um gravador de áudio da marca H4 ZOOM.

### **Descrição dos sujeitos**

Participaram desse estudo 10 sujeitos com esquizofrenia (SE), segundo os critérios definidos pelo DSM V (APA, 2013) e 10 sujeitos que formaram o grupo de controle (SC). Os SEs eram 4 residentes no Estado do Rio de Janeiro e 6 moradores do Estado de São Paulo, entre eles 2 conviviam em uma casa de apoio. Todos eram do sexo masculino, entre 25 e 65 anos, alfabetizados e estavam em tratamento ambulatorial frequente com médicos, psicólogos e demais técnicos; além disso, realizavam cotidianamente atividades terapêuticas nas instituições de saúde mental a qual estavam vinculados. O grupo de controle se formou também com 10 participantes, todos do sexo masculino, entre 24 e 63 anos, alfabetizados. Nenhum deles havia realizado tratamento psiquiátrico ou psicológico. Todos eram falantes nativos do português brasileiro.

Todas as informações sobre os objetivos e procedimentos realizados foram fornecidas para a equipe de profissionais locados na instituição de saúde mental. Com o intuito de que os riscos para a participação nessa pesquisa fossem mínimos, uma vez que os pacientes poderiam, eventualmente, ter exibido alguma ansiedade emocional durante os procedimentos, a psicóloga responsável pelo experimento permaneceu 15 dias sucessivos em cada estabelecimento, acompanhada diretamente pelos profissionais das unidades. Dessa maneira, foi possível interagir e estabelecer vínculo afetivo com os pacientes e colaboradores das instituições, o que facilitou a interação durante as gravações.

Todos os participantes aceitaram contribuir voluntariamente com a pesquisa. As gravações ocorreram conforme o roteiro descrito acima com o acompanhamento dos profissionais da equipe multiprofissional por meio de reuniões diárias ou semanais. Em alguns casos específicos, algum profissional e/ou o responsável legal acompanhava no próprio local enquanto os procedimentos que estavam sendo desenvolvidos. Nesses casos, foi solicitado que não interrompessem o diálogo com o participante. Porém, em algumas circunstâncias, o profissional ou responsável legal também se comunicava.

Aos participantes e aos seus responsáveis legais foram garantidos sigilo e respeito com as informações coletadas. Antes de se iniciar as gravações foram fornecidas todas as orientações e solicitado que todos assinassem o Termo de Consentimento Livre e Esclarecido (TCLE) antes de iniciarem a coleta. Esses documentos foram impressos em duas vias — uma delas foi entregue ao participante ou a seu responsável legal e a outra será arquivado pela psicóloga responsável, conforme a Resolução 466/12 do Conselho Nacional de Saúde.

## Teoria e cálculos

A partir dos arquivos sonoros obtidos durante a coleta de dados e já processados por outros instrumentos de análise acústica (FERREIRA NETTO, 2016), o aplicativo toma o Tom Médio ( $TM$ ), extraído de  $f_0$  pela equação (1):

$$TM_{(t)} = \frac{(t - 1) * TM_{(t-1)} + Z_{(t)}}{t} \quad (2.1)$$

Na equação (1.1),  $t$  corresponde a cada um dos momentos de  $f_0$  em que se faça a medição das frequências e  $Z(t)$  é cada um dos valores obtidos nesse momento  $t$ . A seleção desses momentos  $t$  faz-se pelas restrições adequadas a cada língua quanto à duração mínima, à intensidade mínima e ao intervalo de frequência. O multiplicador  $(t - 1)$  estabelece o número total dos momentos anteriores atual como peso na média geral de modo a calcular a média acumulada no tempo. A cada uma das frequências  $Z$  obtidas nos momentos  $t$  convencionou-se chamar de UBI (*unidade básica de entoação*). Conforme vai na equação (A) acima, o Tom Médio é o conjunto da série de UBIs que se obteve.

Uma vez definido o Tom Médio, estabelece-se uma faixa de 3 semitons acima e 3 semitons abaixo, isto  $1,19 * TM_{(t-1)}$  e  $0,84 * TM_{(t-1)}$  respectivamente, para cada uma das UBIs — ou  $Z(t)$  (MARTINS, 2011; MARTINS, FERREIRA NETTO, 2017). Esses valores correspondem à expectativa que se tem da faixa de frequência que será recuperada após a interrupção de produção sonora realizada no intervalo entre  $Z_{(t-1)}$  e  $Z_{(t)}$ . Convencionou-se chamar essa alternância entre a realização de UBIs e o seus intervalos de Ritmo Tonal. Os intervalos entre UBIs que ultrapassem valores adequados, caracterizando uma interrupção do Ritmo Tonal, são tomados como uma pausa intencional. As UBIs que alcancem valores que extrapolem o limite superior da faixa de frequência do Tom Médio caracterizam-se pela maior saliência tonal e, portanto, estabelecem o que se convencionou chamar de Foco/Ênfase superior ( $F/Esup$ ). As UBIs cujos valores ocorram abaixo do limite inferior da faixa de frequência do Tom Médio caracterizam-se pela atenuação da sonoridade da série tonal (Kohler, Niebuhr, 2007); e estabelecem o que se convencionou chamar de Foco/Ênfase inferior ( $F/Einf$ ). As inequações (B) e (C) definem esses parâmetros.

$$F/Esup(t) > 1,19 \times TM_{(t-1)} \quad (2.2)$$

$$F/Einf(t) < 0,84 \times TM_{(t-1)} \quad (2.3)$$

Os valores 1,19 e 0,84 nas equações (1.2) e (1.3) correspondem, respectivamente, a 3 semitons e estão de acordo com o trabalho de t'Hart e seus colegas (1990). Há que se salientar, entretanto, que tais valores podem ser estabelecidos conforme requeriam outras avaliações. Martins e colegas (2017) realizaram uma análise preliminar cotejando a percepção das variações entoacionais no inglês e no português e verificaram que os sujeitos anglofalantes, diferentemente dos falantes do Português do Brasil, eram sensíveis a mudanças de *pitch* a partir de 1 semitom, indiferentemente ao fato de serem músicos ou não. Consoni (2011), Peres e colegas (2011) encontraram a mesma variação de sensibilidade de 3 semitons, indiferentemente do fato de serem músicos. Esses resultados são corroborados com a pesquisa de Madsen e colegas (2017).

Para a implementação de um processo automático de segmentação de frases entoacionais, foi necessário estabelecer alguns parâmetros liminares. A partir do levantamento bibliográfico, foi possível destacar alguns parâmetros que podem ser usados como marcadores de fronteiras de frases entoacionais, quais sejam: pausa, entoação final descendente, *shift*, duração silábica, e declinação frasal (ou baseline).

Em relação à duração da pausa, pesquisas anteriores estabeleceram diferentes medidas temporais para a percepção de pausas. Alguns trabalhos demonstraram que a partir de 100ms, as pausas já eram reconhecidas como tais (BOOMER, DITTMANN, 1962; LÖVGREN, DOORN, 2005). De maneira geral, mesmo para esses autores, pausas mais longas foram uma opção mais segura, a partir de 200ms (Duez, 1985; Fletcher, 1987; Friedman, Oconnell, 1991; Oliveira, 2002; Silva, 2017). Para este ensaio, optamos por um limite de 200ms como duração mínima de pausa.

Quanto à entoação final descendente, seguimos o princípio estabelecido pela proposta teórica do programa ExProsodia de pesquisa para a entoação de fala, em que o limite inferior abaixo do tom médio perceptível para a língua portuguesa falada no Brasil é de três semitons. Critério semelhante foi usado para a definição do *shift*, ora interpretado de forma semelhante ao de *pitch reset*. A proposta teórica do programa de pesquisa ExProsodia para a análise da entoação preconiza uma variação máxima de três semitons acima do tom médio como variação necessária para a percepção de movimento entoacional.

Boomer e Dittmann (1962), como já vimos, optaram por diferenciar pausas de hesitação, que ocorriam com maior duração de, pelo menos, 50ms, e pausas delimitadoras frasais, que se manifestavam já a partir dos 200 ms. Para este ensaio, optamos pela reprodução desse princípio, valendo-nos de pausas com duração de 200ms como marcadores prosódicos liminares de fronteiras interfrasais e pausas como duração maior ou igual a 500ms como marcadores intencionais de segmentação frasal. Critério semelhante foi utilizado em Silva (2017).

Tomando esses parâmetros foi possível estabelecer dois critérios de segmentação para estabelecimento de *Pitch Reset* ou renovação do tom médio. O primeiro critério segue o princípio proposto por Daneš (1960) da juntura formada pela redundância dos critérios de duração de

pausa e direção entoacional descendente. O trabalho de Moraes (1998) corrobora essa afirmação, na medida em que, segundo ele, o português, como na maioria das línguas conhecidas, o padrão declarativo neutro caracteriza-se pela queda na frequência fundamental ( $f_0$ ) no final do enunciado (mais precisamente, na tônica final) enquanto o tom inicial está no nível médio. Com base nas pesquisas realizadas no âmbito do programa ExProsodia (FERREIRA NETTO, CONSONI, 2008; GARCIA, 2015; COSTA, 2016), essa característica parece vincular-se mais a aspectos sociais e não propriamente linguísticos. O segundo critério considera especialmente a duração da pausa, ou da interrupção do Ritmo Tonal.

Neste ensaio pretende-se descrever os padrões entoacionais de frases assertivas em que se considerem diferenciadamente frases lidas, espontâneas e, neste caso, letradas e não letradas. Assim, foi necessário estabelecerem-se três parâmetros, a serem usados alternativamente, mas não exclusivos entre si:

(i) ocorrência de intervalo ascendente com valor maior do que  $F/Esup_{(t)}$  entre os momentos  $Z_{(t-1)}$  e  $Z_{(t)}$ , em que o momento  $Z_{(t-1)}$  seja menor ou igual a  $F/Einf(t)$ ;

(ii) interrupção do Ritmo Tonal com duração maior ou igual ao menor valor definido para pausa seguida de um intervalo tonal maior ou igual a  $F/Esup_{(t)}$  ou menor ou igual a  $F/Einf_{(t)}$ ;

(iii) interrupção do Ritmo Tonal com duração maior ou igual ao dobro da duração definida para uma pausa.

Um intervalo ascendente (*shift*) com valor maior do que  $F/Esup_{(t)}$  entre dois momentos consecutivos na série temporal de  $f_0$  associa-se ao fato de que os limites do Tom Médio são, segundo o modelo ExProsodia, os limites para a percepção da variação tonal quando há cadeia segmental (léxico) presente. Aparentemente a distância de 3 semitons entre uma UBI final e uma UBI inicial em fronteiras frasais não é suficiente para caracterizar uma segmentação, pois esses intervalos ocorrem com muita frequência na fala espontânea. Dessa maneira, houve a opção por conjugar esse parâmetro com o fato de que o *shift* deveria partir de um tom perceptivelmente grave passível de definir uma finalização autêntica.

Uma vez estabelecidos os momentos liminares de segmentação da série temporal, pode-se questionar sua relação com outras possibilidades de segmentação que envolvam critérios sintáticos e semânticos, por exemplo. No âmbito do programa ExProsodia de pesquisa, a relação entre esses dois aspectos da linguagem, prosódia e léxico, foi discutida anteriormente (FERREIRA NETTO, 2017) e se reporta à concatenação de dois sistemas sonoros diferentes. O sistema o prosódico tem fonte contínua de alimentação que se relaciona diretamente com necessidades expressivas do falante, envolvendo, por exemplo, seu estado emocional; o sistema lexical tem fonte descontínua de alimentação que se relaciona diretamente com as seleções lexicais do falante, envolvendo instâncias especialmente linguísticas da cognição. De maneira geral, essa proposta segue o princípio estabelecido por Shannon e Weaver (SHANNON, 1948) para a comunicação em geral e, bem posteriormente, por Taylor (2000, p. 1698), quando *Diadorim*, Rio de Janeiro, vol. 24, número 1, p. 336 - 365, 2022.

propôs que a fala pudesse ser tomada como uma sequência de eventos entoacionais, que ele chamou de *intonational stream* conectada com uma sequência de eventos segmentais, que ele chamou de *segmental stream*. A concatenação desses sistemas envolveria várias instâncias da análise automática. Neste ensaio, é nossa preocupação a interpretação das segmentações interfrasais liminares.

Conforme o levantamento que se apresentou no início deste ensaio, a segmentação de frases entoacionais se estabelece por meio de diferentes critérios, quer sejam interrupções no fluxo de informações sonoras, variações de frequência pontuais ou contínuas, ascendentes ou descendentes, duração segmental, com quaisquer combinações entre eles. Mesmo em se tratando de descrições de sentenças assertivas, não há um critério de segmentação que, por si só, estabeleça uma marca de finalização sentencial, mas, pode-se propor que o conjunto das interrupções de fala, variações de frequência, alongamento de segmento final e declinação forme um sistema redundante no sistema prosódico na ocorrência de pontos de finalização frasal. Para esta pesquisa, estabelecemos os seguintes critérios para a avaliação das fronteiras interfrasais: movimento da entoação, (movimento) direção da entoação, (direção) alvo da entoação, (alvo) interrupção do ritmo tonal (pausa  $>0,05$ ), alongamento de UBI, (alongamento) correlação entre tom médio e tom médio total, (correlação), diferença entre o tom médio e o tom médio total (diferença) e dispersão de entoação acima do tom médio. (F/E)

Os critérios de *movimento*, *direção* e *alvo* da entoação estão diretamente relacionados com a variação de F0 em relação ao tom médio considerando seus intervalos laterais. Dessa maneira, entende-se que o movimento final de F0 pode ser *oblíquo* ou *pontual*, isto é, no caso de um movimento oblíquo, os valores finais de F0 serão próximos entre si; no caso de um movimento pontual, os valores finais de F0 serão distantes entre si, ou seja, ocorrerão por salto entre UBIs contíguas. A direção de F0 pode ser *ascendente* ou *descendente*, isto é, as UBIs finais contíguas estarão em diferentes alturas, apresentando um movimento por salto ou oblíquo ascendente ou descendente. O alvo de F0 poderá ser *autêntico* ou *plagal*. No caso do alvo autêntico, espera-se que um movimento entoacional descendente que atinja um ponto grave, abaixo do limite inferior de F0, calculado a partir da inequação (C) descrita mais acima. O alvo plagal caracteriza-se por manter a curva entoacional dentro dos limites de F0, ou mesmo acima disso. Nesse caso, frases interrogativas caracterizam-se como finalizações plagais com marcação de foco/ênfase. Nesse processo, entende-se que a finalização por movimento descendente, pontual e autêntico seja a forma mais marcada da finalização de frases assertivas.

Finalização	Descrição	Inequação	
Movimento pontual	Valores finais de F0 distantes entre si, com movimento realizado por salto, ascendente ou descendente, entre UBIs contíguas.	$\frac{F0_{(t-1)}}{F0_{(t)}} < 0,84$ ou $\frac{F0_{(t-1)}}{F0_{(t)}} > 1,19$	(2.4)
Movimento oblíquo	Valores finais de F0 próximos entre si, com movimento realizado oblíquo, ascendente ou descendente, entre UBIs contíguas.	$\frac{F0_{(t-1)}}{F0_{(t)}} > 0,84$ e $\frac{F0_{(t-1)}}{F0_{(t)}} < 1,19$	(2.5)
Direção ascendente	UBI final que se atinge por movimento ascendente.	$\frac{F0_{(t-1)}}{F0_{(t)}} > 1$	(2.6)
Direção descendente	UBI final que se atinge por movimento descendente.	$\frac{F0_{(t-1)}}{F0_{(t)}} < 1$	(2.7)
Alvo plagal	UBI final que se atinge por movimento reto ou ascendente.	$\frac{F0_{(t)}}{TM_{(t)}} < 0,84$	(2.8)
Alvo autêntico	UBI final, abaixo do limiar inferior do TM, que se atinge por movimento descendente.	$\frac{F0_t}{TM_t} > 0,84$	(2.9)

A *interrupção do ritmo tonal* caracteriza-se mais comumente pela supressão intencional e demorada da fala do locutor. Tal supressão de fala pode ser causada tanto pela interferência de um interlocutor, em decorrência de troca ou de roubo de turno, como pela necessidade do próprio locutor. Nesse caso, pode tanto ocorrer a formação de sons de vária natureza, como “pausas preenchidas” ou *fillers*, como a formação de intervalos de silêncio chamados *pausas*. Uma vez que as pausas que ocorrem na fala podem corresponder a fronteiras de frases entoacionais ou a momentos em que o locutor procura dar alguma expressividade a sua fala, ou ainda a uma tentativa de reorganização de seu discurso, elas não se permitem serem tomadas como marcadores de finalização de frases entoacionais, por si só; mas podem ser tomadas como cruciais para o conjunto de marcadores que atuam como fronteiras de frases entoacionais. Mais que marcadores de finalização, pausas também atuam como barreiras para o fluxo da entoação justamente porque interrompem o ritmo tonal. Ainda que ora isso seja especulativo, pode-se assumir que uma finalização de frase entoacional que prescindia de uma pausa exigirá uma compensação quanto às formas marcadas de movimento, direção e alvo tonal, bem como do alongamento de UBI. Para este ensaio, entretanto, as pausas serão consideradas obrigatórias na formação de fronteiras de frases entoacionais. Nesse caso, segue-se o critério de que já se

tratou anteriormente, entendendo que pausas longas, mais de 0,5 segundos, são mais facilmente percebidas do que as com menor duração; no entanto, as pausas longas ocorrem principalmente em momentos de hesitação. Assim, pressupõe-se que pausas longas sejam mais marcadas como fronteiras do que pausas com menor duração em frases assertivas. Embora seja comum a inserção de expressões hesitativas que não se realizam como palavras lexicalizadas, tais como “os *áh, éh, ahm, mm*, quase sempre alongados e preenchendo pausas (...) matéria prima das pausas preenchidas” (MARCUSCHI, 1999, p. 165-166)”, elas não serão objeto de discussão neste ensaio.

A *renovação* do tom médio, que se caracteriza pela presença de *shift* ou de *reset*, pode se estabelecer pela diferença entre a frequência final de uma frase entoacional e a frequência inicial da frase seguinte. O alcance dessa diferença, por sua vez, pressupõe que o locutor tenha abandonado o tom médio da frase entoacional anterior, ou porque já a tinha finalizado, ou por conta de alguma interrupção do ritmo tonal, e promoveu seu reinício. Essa renovação pode ser definida na equação (A) pelo reinício do momento  $t$  do  $TM$  no multiplicador  $(t - 1)$  do  $TM$ , bem como no do denominador  $t$ . Na medida em que a renovação interrompe o fluxo entoacional da fala, entende-se tratar-se de uma forma mais marcada do que sua incoerência em frases assertivas.

Embora uma variação de duração segmental, por sua vez, não envolva qualquer relação com a variação entoacional da fala, pode ser tomada como um aspecto interno da formação das UBIs, que, justamente, portam a entoação frasal. Essa variação de duração de UBI final, ora chamada de *alongamento*, recebeu pouca atenção na prosódia do português falado no Brasil. Assim, esse alongamento se permite avaliar somente em comparação com as demais durações de UBI; caso a UBI final tenha uma duração maior do que a média esperada para as UBIs da mesma frase entoacional em que se encontra será tratada como forma marcada de finalização de frase assertiva.

A correlação entre tom médio e tom médio total, a diferença entre o tom médio e o tom médio total e a dispersão de entoação acima do tom médio são fatores que, diferentemente dos critérios anteriores usados como fronteiras de frases entoacionais, não se baseiam nos momentos finais dessas frases. Na medida em que o tom médio se diferencia do tom médio total pelo fato de que o tom médio total não sofre interferências de fronteiras frasais, ele serve como uma referência fixa para a avaliação do tom médio, definido para cada frase entoacional. Assim, havendo uma declinação frasal bem definida ou muitas inflexões, a correlação entre tom médio total e tom médio será muito baixa ou inexistente. Uma correlação negativa também mostra diferenças marcadas entre eles. Assim, entende-se que a ocorrência de correlação forte entre tom médio e tom médio total é a forma menos marcada para a definição de uma frase assertiva. Relação semelhante pode ser definida pela razão entre tom médio e tom médio total. Na medida em que essa razão se afaste da igualdade, que seria uma razão = 1, entende-se que houve uma tendência maior a uma mudança de registro em relação ao que era esperado. A diferença em



relação à variação do tom médio demonstra, por exemplo, variação de manifestação emocional na fala. Essa variação de registro emocional já foi verificada em outras pesquisas no âmbito do projeto ExProsodia (VASSOLER; MARTINS, 2013; FERREIRA NETTO et al., 2014). Nesse caso, ainda que esse critério também se defina em relação às variações de registro frasal no período, entende-se que, havendo variação, essa seria a forma marcada. A dispersão de entoação de acima do limite superior do tom médio também tem de ser caracterizada como uma tendência à mudança de registro frasal. Essa dispersão também tem de ser definida a partir da comparação com as variações de registro frasal no período. Em trabalho relativo a gênero (FERREIRA NETTO et al, 2010), foi possível verificar que a variação média das frequências acima do tom médio tem caráter significativo e pode apresentar-se como critério diferenciador de frases entoacionais num mesmo período. Assim, apesar de esse também ser um aspecto pouco explorado na prosódia do português falado no Brasil, entende-se que uma variação média maior do que 3 semitons de foco/ênfase é mais marcada do que uma variação média menor do que essa.

### **Resultados: variação de entropia relativa entre os sujeitos**

Com base nos resultados obtidos na tabela dos nove parâmetros binários, estabelecemos a probabilidade de ocorrência desses parâmetros com valor 1 em cada frase entoacional que ocorria em cada discurso. Para obter a probabilidade de ocorrência do resultado obtido em cada frase entoacional, calculamos um espaço amostral equivalente a 29 (=512), tomamos a probabilidade de cada ocorrência (=  $1/512 = 0,01953$ ) e atribuímos um peso a cada um dos valores obtidos, segundo uma distribuição normal, conforme a tabela 1.

**Tabela 1:** Probabilidades calculadas de acordo com as medidas obtidas para o espaço amostral dos nove parâmetros descritos.

Valor obtido	Peso	$p()$
0	1	$1,95*10^{-3}$
1	9	$1,76*10^{-2}$
2	36	$7,03*10^{-2}$
3	84	$1,64*10^{-1}$
4	126	$2,46*10^{-1}$
5	126	$2,46*10^{-1}$
6	84	$1,64*10^{-1}$
7	36	$7,03*10^{-2}$
8	9	$1,76*10^{-2}$
9	1	$1,95*10^{-3}$

Com base nessas probabilidades, obtivemos a soma total das probabilidades no discurso de cada sujeito ( $p_{\text{suji}}$ ) que aplicamos à equação (3) de entropia relativa de Shannon ( $H_r$ ) (SHANNON, 1948, p. 13), sendo o total de frases entoacionais que ocorreram em cada discurso.

$$H_r = \frac{\log_2 \frac{1}{p_{suji}}}{n_i} \quad (3)$$

Os valores obtidos com base na aplicação dessa fórmula vão dispostos na tabela abaixo:

**Tabela 2:** Entropias Relativas referentes aos discursos dos sujeitos com esquizofrenia e dos sujeitos do grupo de controle

Sujeitos com esquizofrenia	Sujeitos do grupo de controle
0,144	0,216
0,079	0,228
0,075	0,343
0,208	0,182
0,052	0,152
0,041	0,258
0,206	0,187
0,066	0,099
0,104	0,089
0,075	0,279

A comparação entre as entropias relativas entoacionais de sujeitos com esquizofrenia e de sujeitos do grupo de controle apresentou resultados significativos. Os valores médios das entropias relativas e dos desvios-padrão foram mais baixos para os sujeitos com esquizofrenia ( $\bar{x} = 0,105$  e  $\sigma^2 = 0,004$ ) do que para os membros do grupo de controle ( $\bar{x} = 0,203$  e  $\sigma^2 = 0,006$ ).

Na avaliação estatística comparada entre os dados apresentados na Tabela 2, o teste-z para comparação entre as médias apresentou um resultado significativo com  $p < 0,01$  e  $z_{(1,645)} = -3,126$ . Esses resultados apontam para o fato de que a entropia relativa entoacional dos sujeitos com esquizofrenia é menor do que a do grupo, caracterizando o discurso esquizofrênico com tendência forte para a diminuição na quantidade de informação entoacional. Com base nessa tabela realizamos testes comparativos. O teste  $F$  para comparação de variâncias não foi significativo. O teste  $t$  para amostras com variâncias iguais também foi significativo tanto unicaudal —  $p < 0,01$  e  $t_{(1,734)} = -3,126$  — como bicaudal —  $p < 0,01$  e  $t_{(2,101)} = -3,126$ .

### Varição de entropia relativa entre os parâmetros

A avaliação da entropia relativa dos parâmetros considerou a probabilidade de ocorrência das opções marcadas em relação às não marcadas, estabelecendo, portanto, uma probabilidade de 0,5 para a ocorrência de uma opção marcada e, em seguida, aplicando à soma de todas as probabilidades em cada parâmetros a mesma fórmula descrita em (J), considerando-se apenas tratar-se de entropia relativa de parâmetros e não de discurso.

**Tabela 3:** Comparação dos parâmetros entre os grupos. A primeira coluna descreve cada um dos parâmetros utilizados para a avaliação. Na segunda e na terceira coluna vão os valores de entropia relativa encontrados para cada um dos parâmetros discriminados por grupos, e os testes estatísticos correspondentes. Na terceira e na quarta colunas, vão os resultados dos testes estatísticos correspondentes

Parâmetros	Sujeitos com esquizofrenia	Sujeitos do grupo de controle	teste F	teste t
<b>Direção</b>	0,970	0,878	F = 0,064 p ≤ 0,001	t = 2,109 p ≤ 0,05
<b>Movimento</b>	0,694	0,805	F = 0,951 N.S.	t = -0,824 N.S.
<b>Alvo</b>	0,573	0,819	F = 0,56 N.S.	t = -2,214 p ≤ 0,05
<b>Alongamento</b>	0,628	0,819	F = 1,451 N.S.	t = -2,670 p ≤ 0,01
<b>Correlação</b>	0,971	0,848	F = 0,022 p ≤ 0,001	t = 1,874 p ≤ 0,05
<b>Diferença</b>	0,454	0,175	F = 0,674 N.S.	t = -0,146 N.S.
<b>F/E</b>	0,564	0,684	F = 0,513 N.S.	t = -1,470 N.S.
<b>Pausa&gt;0,5s</b>	0,353	0,521	F = 0,423 N.S.	t = -1,092 N.S.
<b>Shift</b>	0,783	0,943	F = 8,935 p ≤ 0,01	t = -1,893 p ≤ 0,05

A comparação de cada um dos parâmetros entre os grupos mostrou que *movimento*, *diferença*, *F/E* e *pausa>0,05* não foram significativamente diferentes.

A comparação entre os grupos quanto aos parâmetros *alvo* e *alongamento* mostrou que há diferença significativa quanto a entropia relativa média, mas não quanto à dispersão. Os dois parâmetros tiveram uma entropia relativa mais baixa entre os sujeitos com esquizofrenia.

A comparação dos parâmetros *direção*, *correlação* e *shift* entre os grupos mostrou uma diferença significativa quanto à variação da entropia relativa e quanto à dispersão dos valores dentro de cada grupo. Os parâmetros *direção* e *correlação* tiveram uma entropia relativa mais alta entre os sujeitos com esquizofrenia. A variância ( $\sigma^2 = 0,001$  para *direção* e  $\sigma^2 = 0,0009$  para *correlação*) desse grupo para o parâmetro *direção* em relação à do grupo de controle ( $\sigma^2 = 0,018$  para *direção* e  $0,042$  par *correlação*) mostra maior homogeneidade entre os sujeitos com esquizofrenia. Quanto ao parâmetro *shift*, a entropia relativa dos sujeitos com esquizofrenia foi mais baixa, apresentando maior dispersão ( $\sigma^2 = 0,064$  para o grupo os sujeitos com esquizofrenia e  $\sigma^2 = 0,007$  para os sujeitos do grupo de controle).

## Discussão

Ao lidar com a análise automática da entoação do ponto de vista da Teoria Matemática da Comunicação de Shannon, pressupõe-se que as informações entoacionais decorrem de escolhas feitas pelos sujeitos em estágios sucessivos de suas locuções. Desse ponto de vista, cabe entender que os limites extremos dessas escolhas são, de um lado, a ausência de opções e, de outros, a absoluta ausência de regras ou tendências capazes de orientar os sujeitos nesse processo de escolha. A medida oferecida por Shannon, no âmbito de seu modelo teórico, é a aplicação do cálculo de entropia desenvolvido por Ludwig Boltzmann no século XIX. Esse cálculo estabelece limites máximo e mínimo para a entropia de um sistema. O grau zero de entropia representaria a completa estagnação do sistema ou, no caso da análise automática da entoação, representaria um procedimento categórico que se repetia continuamente na locução produzida por um sujeito. O grau máximo de entropia decorre da total aleatoriedade do sistema e está diretamente relacionado ao número de opções disponíveis em cada estágio do sistema. Assim, um sistema com duas opções terá entropia máxima igual a um, um sistema com três opções terá entropia máxima igual a 1,58, um sistema com cinco opções terá uma entropia máxima igual a 2,38, e assim em diante.

No intervalo que se estabelece entre entropia zero e entropia máxima, ocorrem os fenômenos entoacionais que interessam à análise linguística. Entretanto, por se tratar de um sistema contínuo, a entoação recebeu um tratamento automático que procurou seguir a proposta de Shannon.

To a considerable extent the continuous case can be obtained through a limiting process from the discrete case by dividing the continuum of messages and signals into a large but finite number of small regions and calculating the various parameters involved on a discrete basis. (SHANNON; WEAVER, 1949, p. 81) <sup>4</sup>

A divisão do sinal que procedemos decorreu especialmente da análise empreendida no que chamamos Ritmo Tonal (FERREIRA-NETTO, 2016). Essas variações rítmicas, que pressupõem alternância entre porções do sinal em que há unidades de suporte para ondas periódicas, como segmentos soantes, e porções em que não há, segmentos obstruintes e silêncios, serviram de base para toda a análise automática que foi realizada nos experimentos descritos neste artigo. Os nove parâmetros utilizados dizem respeito à qualidade, à distribuição e à interação entre as unidades que formam o Ritmo Tonal.

Entendemos que os estágios propostos na teoria de Shannon seriam cada uma das frases entoacionais definidas automaticamente e que a fonte de informação seriam os próprios locutores

---

<sup>4</sup> Até um certo nível os casos contínuos podem ser equacionados através de um processo limitativo dos casos distintos, dividindo-se o “continuum” de mensagens ou sinais em número grande, não obstante, finito de regiões e calculando-se os vários parâmetros resultantes como se estivéssemos operando com um sistema distinto. (SHANNON; WEAVER, 1975, p. 87)

de tais frases entoacionais. Com base nessa interpretação foi possível constatar que os grupos de sujeito cuja entoação foi analisada apresentaram entropia significativamente diferente. Esse fato pode ser interpretado sob a luz do teorema 9 do modelo teórico de Shannon.

Let a source have entropy  $H$  (bits per symbol) and a channel have a capacity  $C$  (bits per second). Then it is possible to encode the output of the source in such a way as to transmit at the average rate  $C - \epsilon$  symbols per second over the channel where  $\epsilon$  is arbitrarily small. It is not possible to transmit at an average rate greater than  $C$ . (SHANNON; WEAVER, 1949, p. 59)<sup>5</sup>

Desse ponto de vista, a diferença que se constatou entre os sinais *output* das fontes em questão apontou, por exemplo, para a dificuldade de se adequar os dois sistemas a uma situação dialógica. Nesse caso, o sinal *output* com maior entropia não poderá ser reproduzido de maneira tão exata quanto possível, mas será somente uma aproximação do sinal original. (SHANNON; WEAVER, 1949, p. 31)

Não se trata, portanto, de se considerar que os sujeitos com esquizofrenia tenham dificuldade no manejo da entoação, mas sim de caracterizar que o manejo da entoação é feito com possibilidades de escolha diferentes daquela do grupo de controle.

No conjunto dos parâmetros que estabelecemos para fazer essas medições, as diferenças mais significativas ocorreram entre *alvo*, *alongamento*, *direção*, *correlação* e *shift*. Na medida em que os sujeitos com esquizofrenia tendem a não marcar limites de entoação frasal, mesmo por movimento descendente, pressupõe-se que tais limites possam codificados pela seleção lexical e sua formulação sintática. Essa dissociação entre entoação e cadeia léxico-sintática corrobora as discussões apresentadas de Crow (2000, 2004), Ceccherini-Nelli *et al.* (2007).

Recentemente, Ernala e seus colegas (ERNALA *et al.*, 2017) verificaram uma mudança significativa nas manifestações escritas de sujeitos com esquizofrenia antes e depois de se autodeclararem com esquizofrenia, publicamente, em redes sociais. Dentre as várias hipóteses para suas análises, os autores tomaram a coerência do discurso (ANDREASEN, 1982), especificamente considerando a falta de coerência sintática, semântica e pragmática (NASLUND *et al.* 2017). Todorov (1978) propôs que a incoerência no discurso dos sujeitos com esquizofrenia poderia ser causada por disfunções metalinguísticas, dentre as quais ele salienta anáforas, conjunções e sinalizadores da hierarquia discursiva, atentando para a entoação. Especialmente quanto às anáforas, estas caracterizam as progressões referenciais que estruturam um texto, dando-lhe coesão temática e referencial (KOCH, 2005; 2008; LEITÃO; SIMÕES, 2011, ARAÚJO; FERRARI-NETO, 2020). Ainda que se reportem a um transtorno

---

5 Consideremos uma fonte que tenha entropia  $H$  (bits por símbolo) e um canal que tenha a capacidade  $C$  (bits por segundo). Então será possível codificar o “output” da fonte de maneira tal que esta transmita uma média proporcional  $C - \epsilon$  símbolos por segundo sobre um canal onde  $\epsilon$  é indefinidamente pequeno. E não é possível transmitir a uma razão proporcional maior do que  $C$ . (SHANNON; WEAVER, 1975, p. 63)

diferente, os resultados da pesquisa empreendida por Albuquerque e seus colegas (2012) trazem informações que nos indicam uma direção complementar aos resultados que obtivemos. Essa pesquisa verificou que sujeitos com TDAH apresentam maior lentidão na leitura das correferências, particularmente em relação às formas pronominais. A mesma lentidão não se apresentou de forma significativa quando o processamento recebia estímulo externo. Pesquisa semelhante poderia verificar se o mesmo fenômeno ocorre com pessoas que têm esquizofrenia.

Nossa pesquisa, corrobora também com a hipótese de que uma dissociação entre entoação e cadeia léxico-sintática possa estar correlacionada com dificuldades na associação inter-hemisférica tal como propuseram Ceccherini-Nelli e colegas (2007). Pesquisas que correlacionem música, entoação e cadeia léxico-sintática, tais como as que empreenderam Kantrowitz e colegas (2013), Kantrowitz e colegas (2014) e Hatada e colegas (2014), poderão trazer informações que permitam verificar se a percepção de uma sequência de tons diferentes capazes de desencadear diferentes sensações emocionais ou de expressão linguística socialmente codificada apresenta-se com variações significativas entre pessoas com esquizofrenia e grupos de controle.

### **Procedimentos éticos**

Para esse artigo, foram usadas informações coletadas em duas instituições de saúde mentais diferentes: uma localizada em um bairro da Capital do Rio de Janeiro e outra situada no interior do Estado de São Paulo. Aos participantes e aos seus responsáveis legais foram garantidos o sigilo e o respeito com as informações coletadas. A pesquisadora fornecia orientações e sanava possíveis dúvidas e solicitava que todos assinassem o Termo de Consentimento Livre e Esclarecido (TCLE) antes de iniciarem a coleta. Esses documentos impressos em duas vias, uma delas foi entregue ao participante ou a seu responsável legal e a outra será arquivado pelos pesquisadores, bem como designa a Resolução 466/12 do Conselho Nacional de Saúde. Para a realização dessas coletas foi obtido o consentimento do Comitê de Ética em Pesquisa através da Plataforma Brasil. O trabalho de campo realizado no Rio de Janeiro, teve aprovação concedida pelo Comitê de Ética em Pesquisa do Instituto de Psicologia da Universidade de São Paulo, através do Parecer nº 2.085.673.; o trabalho de campo realizado no interior de São Paulo, teve concordância do Comitê de Ética em Pesquisa da Faculdade de Saúde Pública da Universidade de São Paulo; Parecer nº 2.301.603.

### **Bolsas**

Além dos recursos obtidos na própria Universidade de São Paulo, essa pesquisa obteve apoio da Capes, processo 1721399 e do CNPq processos 401145/2009-0 e 300235/2010-0.

## Referências

ALBUQUERQUE, G. et al. Processamento da linguagem no transtorno do déficit de atenção e hiperatividade (TDAH), *D.E.L.T.A.*, v. 28, n. 2, p. 245-280; 2012.

AMARANTE, P. *Saúde Mental e Atenção Psicossocial*. Rio de Janeiro: Fio Cruz; 2007.

AMERICAN PSYCHIATRY ASSOCIATION. *Diagnostic and Statistical Manual of Mental disorders - DSM-5*. 5th.ed. Washington: American Psychiatric Association; 2013.

ANDREASEN, N. C. Negative symptoms in schizophrenia: definition and reliability. *Archives of general psychiatry*, v. 39, n. 7 , p. 784-788, 1982.

ARAUJO, M. A. FERRARI-NETO, J. Aquisição da correferência anafórica e sua relação com o desenvolvimento da memória de trabalho. *Veredas – Revista de Estudos Linguísticos*, v. 24, n.1, p. 222-257, 2020.

BEBBINGTON, P. et al. (ed.) *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. Leeds: NHS Digital, 2016. p. 131-152

BEHLAU, M. *Voz. O livro do especialista*. Vol. I. Rio de Janeiro: Revinter; 2001.

BOLINGER, D. *Intonations and its Uses*. Melody in grammar and discourse. Stanford: Stanford University Press; 1989.

BOOMER, D. S.; DITTMANN, A. T. Hesitation pauses and juncture pauses in speech. *Language and Speech*, 5(4):215-220, 1962.

BRASIL. CONSELHO NACIONAL DE SAÚDE. *Resolução n° 466, de 12 de dezembro de 2012*. Aprova normas regulamentadoras de pesquisas envolvendo seres humanos. Brasília: Diário Oficial da União; 2013a.

BRASIL. MINISTÉRIO DA SAÚDE. *Protocolo Clínico e Diretrizes Terapêuticas*. Portaria SAS/MS n° 364, de 9 de abril de 2013b.

CAGLIARI, L. C. *Elementos de Fonética do Português Brasileiro*. [Tese]. Campinas: Instituto de Estudos da Linguagem, Universidade de Campinas; 1981.

CAMPBELL, N. Automatic detection of prosodic boundaries in speech. *Speech Communication*, 13(3-4):343-354, 1993.

CECCHERINI-NELLI, A.; TURPIN-CROWTHER, K.; CROW, T. J. Schneider's first rank symptoms and continuous performance disturbance as indices of dysconnectivity of left- and right-hemispheric components of language in schizophrenia. *Schizophrenia Research*, 90:203-213, 2007.

COHEN A, HART JT. Anatomy of intonation. *Lingua*,19(2):177-192, 1967.

CONSONI. F. *Aspectos da percepção da proeminência tonal em português brasileiro*. 2011. [Tese]. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

COOPER, J.; SARTORIUS, N. Cultural and temporal variations in schizophrenia: a speculation on the importance of industrialization. *The British Journal of Psychiatry*, 130:50-55, 1977.

COOPER, W. E.; SORENSEN, J. M. Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America*, 62(3):683-692, 1977.

COOPER, W. E.; SORENSEN, J. M. *Fundamental Frequency in Sentence Production*. New York: Springer-Verlag; 1981.

CONSONI. F. *Aspectos da percepção da proeminência tonal em português brasileiro*. 2011. [Tese]. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

CROW, T. J. Schizophrenia as the price that Homo sapiens pays for language: a resolution of the central paradox in the origin of the species. *Brain Research Reviews*, 31:118–129, 2000.

CROW, T. J. Auditory hallucinations as primary disorders of syntax: an evolutionary theory of the origins of language. *Cognitive Neuropsychiatry*, 9(1-2):125-45, 2004.

CROW, T. J. The ‘big bang’ theory of the origin of psychosis and the faculty of language. *Schizophrenia Research*, 102 (1-3):31-52, 2008.

DANEŠ, F. Sentence intonation from a functional point of view. *Word-Journal of the International Linguistic Association*, 16(1):34-54, 1960.

DITTMANN, A. T.; LLEWELLYN, L. G. Phonemic clause as a unit of speech decoding. *Journal of Personality and Social Psychology*, 6(3):341-349, 1967.

DUEZ, D. Perception of silent pauses in continuous speech. *Language and Speech*, 28:377-389, 1985.

ERNALA, S. K. et al. *Linguistic Markers Indicating Therapeutic Outcomes of Social Media Disclosures of Schizophrenia*. PACM on Human-Computer Interaction, v. 1, n. CSCW, Article 43. Publication date: November 2017.

FAIRBANKS, G. Recent experimental investigations of vocal pitch in speech. *Journal of the Acoustical Society of America*, 11(4):457-466, 1940.

FERREIRA-NETTO, W et al. Relações entre variação de gênero e variação tonal na fala de língua portuguesa do Brasil. *Revista Symposium (Lavras)*, v. 7, n. 1, p. 70-73, 2010.



FERREIRA-NETTO, W .et al. Tentativa de disposição de registros entoacionais num eixo horizontal organizado pela tensão entoacional. *Gradus*, v. 2,. n. 1, p. 14-29, 2017.

FERREIRA-NETTO, W. Análise automática de manifestações emocionais em PB: aplicações do programa ExProsodia. In: Ferreira-Netto, W. (org). *ExProsodia*. Resultados Preliminares. São Paulo: Paulistana; 2016. p. 11-28

FERREIRA-NETTO, W. *ExProsodia*. Revista de Propriedade Industrial. PAULO, U. D. S. Brasil. RS 08992-2 2010.

FERREIRA-NETTO, W.; CONSONI, F. Estratégias prosódicas na leitura em voz alta e da fala espontânea. *Alfa*. Revista de Linguística, v. 52, n. 2, p. 521-534, 2008.

FERREIRA-NETTO, W.; MARTINS, M. V. M. Proposal of description for an intonation pattern: the Simulacrum of Neutral Intonation. *Journal of Acoustical Society of America*, n. 141, p. 3701, 2017.

FERREIRA-NETTO, W.; MARTINS M. V. M.; VIEIRA, M. F. Efeitos da entoação e da duração na análise automática das manifestações emocionais. *Estudos Linguísticos*, n., 43, p. 22-32, 2014.

FLETCHER, J. Some micro and macro effects of tempo change on timing in French. *Linguistics*, 25(5):951-967, 1987.

FOUCAULT, M. *História da loucura na idade clássica*. Trad. José Teixeira Coelho Neto. São Paulo: Ed. Perspectiva; 1978.

FRIEDMAN, L. A.; OCONNELL, D. C. Pause reports for spontaneous dialogic speech. *Bulletin of the Psychonomic Society*, 29(3):223-225, 1991.

GARCIA, R. R. *A entoação do dialeto caipira do Médio Tietê: reconhecimento, características e formação*. 2015. [Tese]. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

GARVIN, P. L, MATHIOT, M. Fused units in prosodic analysis. *Word-Journal of the International Linguistic Association*, 14:(2-3):178-186, 1958.

HARGREAVES, W. A.; Starkweather, J. A. Collection of temporal data with the duration tabulator. *Journal of the Experimental Analysis of Behavior*, 2(3):179-183, 1959.

HATADA, S. *et al*. Impaired musical ability in people with schizophrenia. *Journal of Psychiatry and Neuroscience*, v. 39, n. 2, p. 118-126, 2014.

HIRST, D.; CRISTO, A. D. A survey of intonation systems. In: Hirst D, Cristo A. D (org.). *Intonation Systems*. Cambridge: Cambridge University Press; 1998. p. 1-44

- HOCKETT, C. F. A system of descriptive phonology. *Language*, 18(1):3-21, 1942.
- JONES, D. *An Outline of English Phonetics*. New York: G. E. Stechert & Co.; 1922.
- KANTROWITZ, J. T. *et al.* Reduction in Tonal Discriminations Predicts Receptive Emotion Processing Deficits in Schizophrenia and Schizoaffective Disorder. *Schizophrenia Bulletin*, v. 39, n. 1, pp. 86–93, 2013
- KANTROWITZ, J. T. *et al.* Amusia and protolanguage impairments in schizophrenia. *Psychological Medicine*, v. 44, n. 15 p. 2739-2748, 2014.
- KOCH, I. V. Como se constroem e reconstroem os objetos-de-discurso. *Revista Investigações*, v. 21, n. 2, p. 99-114, 2008.
- KOCH, I. V. Léxico e progressão referencial. In: RIO-TORTO, G. M.; FIGUEIREDO, O. M.;/ SILVA, F. (org.) Estudos em Homenagem ao Professor Doutor Mário Vilela. Porto: Faculdade de Letras da Universidade do Porto, 2005. p. 263-276.
- KOHLER, K. J.; NIEBUHR, O. The phonetics of emphasis. In: Trouvain, J. E.; Barry, W. J. *International Congress of Phonetic Sciences (ICPhS)*, 2007, Saarbrücken. Jürgen Trouvain; William J. Barry. p. 2145-2148. Disponível em: [http://www.ipds.uni-kiel.de/kjk/pub\\_exx/kk2007\\_4/emphasis\\_paper.pdf](http://www.ipds.uni-kiel.de/kjk/pub_exx/kk2007_4/emphasis_paper.pdf)
- KUHL, P. K. Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2):93-107, 1991.
- LADD, D. R. Declination: a review and some hypotheses. *Phonology Yearbook*, 1:53-74, 1984.
- LADD, D. R. *Intonational Phonology*. Cambridge: Cambridge University Press; 1996.
- LEITÃO, M. M.; SIMÕES, A. B. G. A influência da distância no processamento correferencial de pronomes e nomes repetidos em português brasileiro. *Veredas Online*, Atemática, n. 1, p. 262-272, 2011.
- LEITMAN, D. I *et al.* Getting the cue: sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophrenia Bulletin*, 36(3):545-556, 2010. <https://academic.oup.com/schizophreniabulletin/article/36/3/545/1879036>
- LIEBERMAN, P. *Intonation, Perception, and Language*. Cambridge: The M.I.T. Press; 1967.
- LÖVGREN, T.; DOORN, J. V. Influence of manipulation of short silent pause duration on speech fluency. DiSS'05, *Disfluency in Spontaneous Speech Workshop*, 2005, Aix-en-Provence. ISCA Archive. p.123-126. Disponível em: [https://www.isca-speech.org/archive\\_open/archive\\_papers/diss\\_05/dis5\\_123.pdf](https://www.isca-speech.org/archive_open/archive_papers/diss_05/dis5_123.pdf)

MADSEN, S. M. K.; WHITEFORD, K. L.; OXENHAM, A. J. Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific Reports*, 7, 2017. Disponível em: <https://www.nature.com/articles/s41598-017-12937-9>

MAEDA, S. *A characterization of American English intonation*. 1976. 332 [Thesis]. Massachusetts: Electrical Engineering, Massachusetts Institute of Technology.

MARTINEZ-SANCHEZ, F. *et al.* Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia? *The Spanish Journal of Psychology*, n. 18, e86, 1–9, 2015.

MARTINS, M. V. M. *Aspectos da percepção e do controle entoacional do Português Brasileiro*. 2011. [Dissertação]. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

MARTINS, M. V. M.; FERREIRA-NETTO W. Os limiões de diferenciação tonal do português brasileiro. *Revista do GEL*, v. 14, n. 2, p. 157-182, 2017.

MARTINS, M. V. M.; FERREIRA-NETTO, W.; PERES, D. O. Diferença tonal mínima perceptível em português e inglês. In: SEMINÁRIO DO GEL, 65. Assis: GEL 2017. [citado em 22 ago. 2018]. Disponível em: [https://www.researchgate.net/profile/Waldemar\\_Ferreira\\_Netto/publication/322931630\\_Diferenca\\_tonal\\_minima\\_perceptivel\\_em\\_portugues\\_e\\_ingles/links/5a783589aca2722e4df15115/Diferenca-tonal-minima-perceptivel-em-portugues-e-ingles.pdf](https://www.researchgate.net/profile/Waldemar_Ferreira_Netto/publication/322931630_Diferenca_tonal_minima_perceptivel_em_portugues_e_ingles/links/5a783589aca2722e4df15115/Diferenca-tonal-minima-perceptivel-em-portugues-e-ingles.pdf)

MITCHELL, R. L. C.; CROW, T. J. Right hemisphere language functions and schizophrenia: The forgotten hemisphere? *Brain*, v. 128, n. 1, p. 963–978, 2005.

MORAES, J. A. Intonational Phonology of Brazilian Portuguese. In: Workshop on Intonational Phonology: understudied or fieldwork languages, *ICPHS 2007 Satellite meeting. Saarbrücken 2007*. Disponível em: <http://linguistics.ucla.edu/people/jun/Workshop2007ICPhS/Moraes-BP.pdf>

MORAES J. A. Intonation in Brazilian Portuguese. In: Hirst, D. E.; Cristo, A. D. (Eds.). *Intonation System. A survey of twenty languages*. Cambridge: Cambridge; 1998. p. 179-194.

MOTA, N. *et al.* Graph analysis of dream reports is especially informative about psychosis. *Scientific Reports*, 4:3691, 2014.

NARDI, A. E.; QUEVEDO, J.; SILVA, A. G. *Esquizofrenia: teoria e clínica*. Porto Alegre: Artmed; 2015.

NASLUND, J. A. *et al.* Health Behavior Models for Informing Digital Technology Interventions for Individuals With Mental Illness. *Psychiatric rehabilitation journal*, v. 40, n. 3, p. 325–333, 2017

OLIVEIRA, M. The Role of Pause Occurrence and Pause Duration in the Signaling of Narrative Structure. In: Ranchhod, E.; Mamede, N. J. (Eds). *Advances in natural language processing: third international conference; proceedings*. Berlin: Springer; 2002. p. 43-51.

PARK, S. *et al.* Abnormal production of prosody and reduced range of pitch in schizophrenia; towards quantification of flat affect. *Schizophrenia Research*, 98:189-190, 2008.

PAROLA, A. *et al.* Voice patterns in schizophrenia: A systematic review and Bayesian metaanalysis. *Schizophrenia Research*, v 216, p, 24-40, 2020.

PERES, D. O.; CONSONI, F.; FERREIRA-NETTO, W. A influência da cadeia segmental na percepção de variações tonais. *LLJournal*, v. 6, n. 1, 2011. Disponível em: <https://lljournal.commons.gc.cuny.edu/2011-1-deoliveira-texto>.

PIERREHUMBERT, J. B. Automatic recognition of intonation patterns. In: Marcus, M. (Eds.). Annual Meeting of the Association for Computational Linguistics, 1983, Cambridge. *Proceedings...*1983, p. 85-90.

PIERREHUMBERT, J. B. *The Phonology and Phonetics of English Intonation*. 1980. [Thesis] Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge.

PIERREHUMBERT, J. The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, 66(2):363-369, 1979.

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

PRICE, P. J.; OSTENDORF, M.; WIGHTMAN, C. W. Prosody and parsing. In: Linguistics, A. F. C., Workshop on Speech and Natural Language-HLT, 1989, *Cape Cod. Proceedings...* p. 5-11.

PRONOVOST, W. *An experimental study of the habitual and natural pitch levels of superior speakers*. 1939. [Thesis]. Iowa City: State University of Iowa.

SANTOS, A. E. *Comportamento comunicativo de indivíduos com diagnóstico de esquizofrenia*. 2012. [Dissertação] Ribeirão Preto: Escola de Enfermagem de Ribeirão Preto Universidade de São Paulo, 2012.

SANTOS, E *et al.* Comportamento comunicativo de indivíduos com diagnóstico de esquizofrenia. *Revista CEFAC*, vol. 16, núm. 4, p. 1283-1293, 2014.

SHANNON, C. E. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379-423, 1948.

SILVA, E. W. R. D. *A relação entre produção e percepção de pistas prosódicas na segmentação de narrativas espontâneas*. 2017. [Thesis] Maceió: Letras, Universidade Federal de Alagoas.

SILVERMAN, K. *ET AL.* ToBI: a standard for labeling English Prosody. In: The Second International Conference on Spoken Language Processing – ICSLP, 1992, Banff, Alberta. *Proceedings...* p. 867-870.

SNIDECOR, J. C. A comparative study of the pitch and duration characteristics of impromptu speaking and oral reading. *Speech Monographs*, 10:50-56, 1943.

SNIDECOR, J. C. An objective study of phrasing in impromptu speaking and oral reading. *Speech Monographs*, 11:97-104, 1944.

SNIDECOR, J. C. The pitch and duration characteristics of superior female speakers during oral reading. *Journal of Speech and Hearing Disorders*, 16(1):44-52, 1951.

STEELE, J. *Prosodia Rationalis: or an essay towards establishing the melody and measure of speech, to be expressed and perpetuated by peculiar symbols.* Second edition amended and enlarged. London: J. Nichols; 1779.

STETSON, R. H. *Motor Phonetics. A study of speech movements in action.* Amsterdam: Springer; 1927.

TAYLOR, P. Analysis and synthesis of intonation using the Tilt model. *Journal of the Acoustical Society of America*, 107(3):1697-1714, 2000.

TAYLOR, P. The rise fall connection model of intonation. *Speech Communication*, 15(1-2):169-186, 1994.

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

TODOROV, T. *Les genres du discours.* Paris: Édition de Seuil, 1978.

TRAGER, G. L.; SMITH, H. L. *An Outline of English Structure.* Washington: American Council of Learned Societies; 1957.

TRUBETZKOY, N. S. *Principios de fonologia.* Trad. Delia Garcia Giordano. Madrid: Editorial Cincel; 1973.

VAISSIÈRE, J. Language Independent Prosodic Features. In: Cutler, A.; Ladd, R. (Eds.) *Prosody: Models and Measurements.* Heidelberg: Springer Verlag; 1983. p. 53-65.

VAISSIÈRE, J. On French prosody. *Quarterly Progress Report, Research Laboratory of Electronics*, Massachusetts Institute of Technology, 212- 223, 1974.

VASSOLER, A. M. O.; MARTINS, M. V. M. A entoação em falas teatrais: uma análise da raiva e da fala neutra. *Revista Estudos Linguísticos*, v. 42, n. 1, p. 9-18, 2013.

VERZEANO, M.; FINESINGER, J. E. An automatic analyzer for the study of speech in interaction and in free association. *Science*, 110(2845):45-46, 1949.

XU Y. Speech Melody as articulatorily implemented communicative functions. *Speech Communication*, 2005;46:220–251.

## ANEXO I

Texto Lido pelos pacientes e controles no experimento de coleta de dados

“Quando todos estavam sentados em fila no alto do muro, olharam para baixo e viram uma cena muito estranha.

À frente deles se estendia uma vasta região em que o solo era liso, reluzente e branco como o fundo de uma travessa de louça. Espalhadas por ele, havia muitas casas, todas feitas de porcelana e pintadas das cores mais vivas. As casas eram todas bem pequenas, e a maior delas mal chegava à cintura da menina.

Ao lado, havia pequenos celeiros rodeados por cercas de louça, reunindo grupos de vacas, carneiros, cavalos, porcos e galinhas, todos feitos de porcelana. Mas o mais diferente de tudo eram as pessoas que viviam naquela terra estranha. Eram camponesas que tiravam leite das vacas, ou pastorinhas com vestidos de cores vivas e aventais de bordas douradas; e princesas com lindos trajes prateados, dourados e púrpura; e pastores usando suspensórios e calças curtas com listras cor-de-rosa, amarelas e azuis, e fivelas de ouro nos sapatos; e príncipes com coroas na cabeça, cravejadas de pedras preciosas, usando mantos de arminho e calções de seda; e palhaços engraçados com roupas enfeitadas por muitas rendas, com manchas vermelhas pintadas nos rostos e chapéus compridos e pontudos. E, o mais estranho de tudo, essas pessoas eram todas feitas de porcelana, inclusive as suas roupas, e eram tão pequenas que a maior delas não passava da *altura* do joelho da menina.” (BAUM, 2011, p. 95)