


**unesp**  **UNIVERSIDADE ESTADUAL PAULISTA**  
**“JÚLIO DE MESQUITA FILHO”**  
**Faculdade de Ciências e Letras**  
**Campus de Araraquara - SP**

André Luiz Machado

## **Fala, canto e música: Limites e interseções**



ARARAQUARA – S.P  
2022

ANDRÉ LUIZ MACHADO

## **Fala, canto e música: Limites e interseções**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras – Unesp/Araraquara, como requisito para obtenção do título de Doutor em Linguística e Língua Portuguesa.

**Linha de pesquisa: Análise Fonológica, Morfossintática, Semântica e Pragmática**

**Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Gladis Massini-Cagliari**

**Bolsa: Capes**

ARARAQUARA – S.P.  
2022

M149f Machado, André Luiz  
Fala, canto e música: : Limites e interseções / André Luiz Machado.  
-- Araraquara, 2022  
153 f.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),  
Faculdade de Ciências e Letras, Araraquara  
Orientadora: Gladis Massini-Cagliari

1. Linguística. 2. Música. 3. Fonética. 4. Entonação. 5. Emoções. I.  
Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências e Letras, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

ANDRÉ LUIZ MACHADO

## **FALA, CANTO E MÚSICA: Limites e interseções**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras – UNESP/Araraquara, como requisito para obtenção do título de Doutor em Linguística e Língua Portuguesa.

**Linha de pesquisa: Análise fonológica, Morfossintática, Semântica e Pragmática.**

**Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Gladis Massini-Cagliari**

**Bolsa: Capes**

Data da defesa: 23/05/2022

### **MEMBROS COMPONENTES DA BANCA EXAMINADORA:**

---

**Presidente e Orientador: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Gladis Massini-Cagliari**  
Universidade Estadual Paulista “Júlio de Mesquita Filho”

---

**Membro Titular: Prof. Dr. Alexsandro Rodrigues Meireles**  
Universidade Federal do Espírito Santo

---

**Membro Titular: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Anise de Abreu Gonçalves d'Orange Ferreira**  
Universidade Estadual Paulista “Júlio de Mesquita Filho”

---

**Membro Titular: Prof. Dr. Daniel Soares da Costa**  
Universidade Estadual Paulista “Júlio de Mesquita Filho”

---

**Membro Titular: Prof. Dr. Marcus Vinícius Moreira Martins**  
Universidade do Estado de Minas Gerais

**Local:** Universidade Estadual Paulista  
Faculdade de Ciências e Letras  
UNESP – Campus de Araraquara

## AGRADECIMENTOS

Uma das maiores lições que eu aprendi é a necessidade de se cercar de pessoas que se destacam naquilo que você gostaria de fazer melhor, e valorizar cada vez que alguém generosamente dedica seu recurso mais precioso, o tempo, para ajudar você na sua jornada.

Willian, por ter me encorajado aceitar o desafio de um Doutorado, por me acompanhar ao longo do caminho e me colocar de pé mais uma vez em todos os momentos em que eu estava fragilizado. Eu te amo!

Gladis, por anos de convivência e de uma orientação impecável. Você me inspira a dar sempre o meu melhor, e espero ter sido digno, em algum grau, da sua confiança.

Obrigado, Daniel e Alessandro, pela leitura e pelas sugestões na minha banca de qualificação.

Agradeço muito às vozes maravilhosas que viabilizaram esta pesquisa: Caio Ceragioli, Carol Gierwiatowski, Guilherme Conradi, Lorena Rebello, Luciano Pacchioni, Luiz Zaggue e Rodrigo Vulcano. Agradeço a cada um de vocês por terem se disposto a me ajudar nesta empreitada. Ao muso inspirador Conrado Carmven, que tem um talento que não cabe em uma pessoa (ou área) só, e novamente ao Rodrigo Vulcano, com sua animação e disposição, minha gratidão por me emprestarem seus ouvidos e garantirem que as gravações não estavam (muito) artificiais. Obrigado a cada um que respondeu um questionário opinando sobre gravações em línguas exóticas.

Marcus Sene, meu polímato e salvador estatístico! Eu não sei como teria conseguido sem a sua ajuda. Natalia Sartori, eu nunca vou conseguir agradecer o suficiente pela sua contribuição: quando o Louco, primeiro arcano maior, parte despreocupado em sua viagem, cedo ou tarde acaba precisando de ajuda para não desmoronar junto com a Torre. Felizmente, no meio do meu caminho não tinha uma pedra, e sim uma tagarela.

Eu preciso agradecer também a cada pessoa que me ajudou a não perder totalmente a sanidade, não só nesse turbilhão pandêmico que começou em 2020 e ainda continua, mas também antes dele e, com sorte, depois. Thaynary, minha carneirinha com o melhor senso de humor do mundo; Michele Delbon, que poderia se chamar Alicerce por não deixar a gente cair e afundar; Carlos Nascimento, muito obrigado por não deixar de acreditar em mim, mesmo que eu nem sempre faça por merecer, e por sempre me mostrar um caminho novo. Adriel Silva, Aline Ferreira, Angelina Morozova (мальчик хочет в тамбов!), Ben-Hur Pilotti, Deivid Filippin (promesssa é dívida, e eu devo a você e a mim mesmo confiar mais na minha capacidade), Gabi Orselli, Juliana Rabatone, Katia Yendo, Maurício Caero, Rômulo Filho (mais de quatro anos depois, ainda preciso da sua ajuda)... Puxa, são tantos! Obrigado a cada um de vocês.

Obrigado *ma and pa* por me darem a oportunidade de chegar até aqui, e obrigado Karin por me dar o privilégio de te ver crescer.

Este trabalho tem um pouco de cada um de vocês.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## RESUMO

Os limites e as interseções entre fala, canto e música – em especial no que se refere à expressão das emoções – são objeto deste trabalho. Pensando especificamente no canto, a literatura da área começa a questionar se cantar e falar são fazeres distintos que se utilizam da mesma materialidade, ou então polos opostos em um contínuo. Nesse sentido, o que faz com que o rap ou o recitativo de uma ópera sejam interpretados como “não-canto”, enquanto alguns sotaques, por exemplos, são considerados “mais cantados” do que outros? De um ponto de vista teórico, a discussão e a busca por parâmetros objetivos que nos permitam demarcar os limites entre música, canto e fala nos levam, a princípio, a pensar na estabilidade das frequências sequenciais e na regularidade rítmica do enunciado, bem como nas próprias definições de língua, linguagem e significado. Com base em experimentos já realizados com falantes de inglês americano, holandês e japonês (que verificaram coincidências acústicas na expressão de emoções tanto na fala quanto na música), amostras bissilábicas de atores e músicos expressando quatro emoções básicas (raiva, alegria, tristeza e simpatia) foram analisadas através dos programas Praat e Melodyne, de modo a se obter estilizações das curvas melódicas, as quais correspondem a intervalos musicais. Assim, pode-se verificar se os intervalos musicais também podem ser considerados parâmetros acústicos confiáveis na expressão de carga emocional nos enunciados, e se a ocorrência desses intervalos na fala também corresponde ao uso desses intervalos na música. Em outro experimento, através da aplicação de um questionário online, informantes foram expostos a duas versões da mesma gravação de um enunciado cantado, sendo uma delas alterada digitalmente de modo a mudar os intervalos da melodia. Ao pedir que os informantes associassem o estímulo sonoro a uma dentre quatro imagens de expressões faciais médias das emoções da coleção Averaged Karolinska Directed Emotional Faces – AKDFE, buscamos verificar se a manipulação dos intervalos cantados acarretaria associações distintas. Por fim, também através de questionário online e da exposição a amostras de fala em diferentes idiomas, tentamos lançar luz à percepção de alguns idiomas como mais melódicos do que outros. Os resultados obtidos neste trabalho podem ser resumidos em três pontos: 1) a expressão das emoções por parte dos falantes de português brasileiro não apresentou correspondências ao uso de intervalos musicais; 2) a percepção da carga emocional de um enunciado melódico não foi correlacionada com a presença de intervalos específicos, mas com a percepção de tonalidade expressa pela melodia; 3) um menor número de consoantes e, principalmente, maior variação da altura melódica, são fatores que parecem estar ligados à percepção de uma língua como mais musical e melódica. Para além destes resultados, argumentamos que vários trabalhos anteriores comparando a expressão na música e na fala não apresentam a clareza necessária na definição de conceitos, e apontamos a investigação da percepção estética das características fonéticas e fonológicas dos idiomas como um campo potencialmente contencioso, mas promissor.

**Palavras-chave:** Fonética. Música. Emoções. Percepção.

## ABSTRACT

The limits and intersections between speech, singing and music – especially with regard to the expression of emotions – are the object of this work. Specifically, as far as singing is concerned, the literature in the area begins to question whether singing and speaking are distinct activities that use the same materiality, or rather two opposite poles in a continuum. In this sense, what makes rap or opera recitative interpreted as “non-singing”, while some accents, for example, are considered to sound “more like singing” than others? From a theoretical point of view, the discussion and the search for objective parameters that allow us to demarcate the boundaries between music, singing and speech lead us, at first, to think about the stability of a sequence of frequencies and the rhythmic regularity of the utterance, as well as the definitions of language and meaning themselves. Based on experiments already carried out with speakers of American English, Dutch and Japanese (which verified acoustic coincidences in the expression of emotions in both speech and music), bisyllabic samples of actors and musicians expressing four basic emotions (anger, joy, sadness and sympathy) were analysed using the programs Praat and Melodyne, in order to obtain stylisations of the melodic curves corresponding to musical intervals. Thus, we can verify whether musical intervals can also be considered reliable acoustic parameters in the expression of emotional charge in utterances, and whether the occurrence of these intervals in speech corresponds to the use of these intervals in music. In another experiment, via an online questionnaire, informants were exposed to two versions of the same recording of a sung utterance, one of them digitally altered in order to change the melody intervals. By asking the informants to associate the sound stimuli to one of four images of average facial expressions of emotions from the Averaged Karolinska Directed Emotional Faces – AKDFE collection, we sought to verify whether the manipulation of the sung intervals would lead to distinct associations. Finally, also via an online questionnaire and exposure to speech samples in different languages, we tried to shed some light on the perception of some languages as more melodious than others. The results reached in this enterprise can be summarised as follows: 1) the emotional expression in the speech of Brazilian Portuguese speakers showed no correspondence to the use of musical intervals; 2) the perception of the emotional content of a melodic utterance had no correlation with the presence of specific musical intervals, rather, it correlated with the perception of tonality expressed by the melody; 3) a lower number of consonants and, more remarkably, greater pitch variation seem to lead a given language to be perceived as more musical and melodious. These results aside, we also argue that many works comparing emotional expression in music and speech do not present the necessary clarity in the definition of concepts, and we put forward the investigation of the aesthetic perception of languages' phonetic and phonological features as a potentially contentious, yet promising, field.

**Keywords:** Phonetics. Music. Emotions. Perception.

## LISTA DE FIGURAS

<b>Figura 1</b> - Uma das gravações de “Sou eu” visualizada no <i>software</i> Praat. ....	41
<b>Figura 2</b> - Uma das gravações de “Eu sei” visualizada no <i>software</i> Praat. ....	43
<b>Figura 3</b> - Tela do módulo interativo do Prosogram. ....	44
<b>Figura 4</b> - Tela do <i>software</i> Melodyne .....	45
<b>Figura 5</b> - Escala de Dó maior, representada em teclas de piano e em partitura.....	56
<b>Figura 6</b> - Escala de Dó menor, representada em teclas de piano e em partitura.....	56
<b>Figura 7</b> - Exemplos de expressões faciais médias das emoções; cada imagem representa a expressão média de 35 atores para uma emoção específica. Da esquerda para a direita: raiva, alegria, neutralidade e tristeza .....	62
<b>Figura 8</b> - Captura de tela do <i>software</i> BioEstat, com os dados e resultados do teste Exato de Fisher .....	67
<b>Figura 9</b> - Correção de afinação no <i>software</i> Melodyne .....	68
<b>Figura 10</b> - Partitura de <i>Aja tutaja</i> . Os intervalos de terça menor aparecem destacados em azul, e os intervalos de terça maior aparecem destacados em laranja. ....	70
<b>Figura 11</b> - Partitura da versão alterada de <i>Aja tutaja</i> . Os intervalos de terça maior aparecem destacados em laranja. ....	71
<b>Figura 12</b> - Partitura original de <i>Itsuki no Komori-uta</i> do site Lullabies of the World.....	73
<b>Figura 13</b> - Partitura adaptada de <i>Itsuki-no Komori-uta</i> , com a letra do segundo verso Os intervalos de terça menor aparecem destacados em azul, e os de terça maior, em laranja. ....	74
<b>Figura 14</b> - Partitura da versão alterada digitalmente de <i>Itsuki-no Komori-uta</i> . Os intervalos de terça menor aparecem destacados em azul, e os de terça maior, em laranja .....	75
<b>Figura 15</b> - Partitura de <i>Lala, mtoto lala</i> . Os intervalos de terça menor aparecem destacados em azul, e os intervalos de terça maior aparecem destacados em laranja.....	77
<b>Figura 16</b> - Partitura da versão alterada de <i>Lala, mtoto lala</i> . Os intervalos destacados em azul são intervalos de terça menor.....	78
<b>Figura 17</b> - Partitura de <i>Tuu, tuu, tupakarulla</i> . Os intervalos de terça menor aparecem destacados em azul, e os intervalos de terça maior aparecem destacados em laranja. ....	80
<b>Figura 18</b> - Partitura da versão alterada de <i>Tuu, tuu, tupakarulla</i> . Destacado em azul, o intervalo de terça menor, e em laranja, os intervalos de terça maior.....	81
<b>Figura 19</b> - Tela do <i>software</i> Praat. ....	106
<b>Figura 20</b> - Variação da altura melódica (em Hz, no eixo vertical) nas fonações de vozes masculinas ao longo do tempo (no eixo horizontal). Alemão em vermelho, húngaro em verde e norueguês em azul.....	107
<b>Figura 21</b> - Variação da altura melódica (em Hz, no eixo vertical) nas fonações de vozes femininas ao longo do tempo (no eixo horizontal). Armênio em vermelho, hindi em ocre e islandês em azul.....	107

## LISTA DE GRÁFICOS

<b>Gráfico 1</b> - Intervalos observados na fala de atores e músicos expressando tristeza .....	47
<b>Gráfico 2</b> - Intervalos observados na fala de atores músicos expressando raiva. ....	48
<b>Gráfico 3</b> - Intervalos observados na fala de atores expressando simpatia .....	49
<b>Gráfico 4</b> - Intervalos observados na fala de atores expressando alegria.....	50
<b>Gráfico 5</b> - Percepção de sentimentos expressos na versão original de <i>Aja Tutaja</i> , em que os intervalos de terça são predominantemente menores. ....	71
<b>Gráfico 6</b> - Percepção de sentimentos expressos na versão alterada de <i>Aja Tutaja</i> , em que todos os intervalos de terça são maiores. ....	72
<b>Gráfico 7</b> - Percepção de sentimentos expressos na versão original de <b>Itsuki no Komori-uta</b> , em que os intervalos de terça são predominantemente maiores. ....	74
<b>Gráfico 8</b> - Percepção de sentimentos expressos na versão alterada de <i>Itsuki-no Komori-uta</i> , em que os intervalos de terça são predominantemente menores. ....	76
<b>Gráfico 9</b> - Percepção de sentimentos expressos na versão original de <i>Lala, mtoto lala</i> , em que há uma proporção igual de intervalos de terça maior e de terça menor.....	77
<b>Gráfico 10</b> - Percepção de sentimentos expressos na versão alterada de <i>Lala, mtoto lala</i> , em que todos os intervalos de terça menores.....	79
<b>Gráfico 11</b> - Percepção de sentimentos expressos na versão original de <i>Tuu, tuu, tupakarulla</i> , em que os intervalos de terça são predominantemente maiores. ....	80
<b>Gráfico 12</b> - Percepção de sentimentos expressos na versão original de <i>Tuu, tuu, tupakarulla</i> , em que os intervalos de terça são predominantemente maiores. ....	81
<b>Gráfico 13</b> - Relação entre variação de f0 e média do nível de musicalidade percebido em amostras de fala com vozes femininas .....	108
<b>Gráfico 14</b> - Relação entre variação de f0 e média do nível de musicalidade percebido em amostras de fala com vozes masculinas .....	108
<b>Gráfico 15</b> - O que faz uma língua soar melodiosa e/ou musical?.....	118
<b>Gráfico 16</b> - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua alemã.....	119
<b>Gráfico 17</b> - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua armênia.....	120
<b>Gráfico 18</b> - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua hindi .....	121
<b>Gráfico 19</b> - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua húngara .....	121
<b>Gráfico 20</b> - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua islandesa.....	122
<b>Gráfico 21</b> - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua norueguesa .....	122

## LISTA DE QUADROS

<b>Quadro 1</b> - Exemplo de tabela de contingência .....	66
<b>Quadro 2</b> - Tessitura das cantigas de ninar.....	95
<b>Quadro 3</b> - Análise silábica de <i>Der Mond is aufgegangen</i> .....	97
<b>Quadro 4</b> - Análise silábica de <i>Aludjal el, csucsuljal el</i> .....	98
<b>Quadro 5</b> - Análise silábica de <i>Gjendines badnlat</i> .....	99
<b>Quadro 6</b> - Análise silábica de <i>Akna Oror</i> .....	100
<b>Quadro 7</b> - Análise silábica de <i>Jasoda Hari palne</i> .....	101
<b>Quadro 8</b> - Análise silábica de <i>Sofðu unga ástin mín</i> .....	102

## LISTA DE TABELAS

<b>Tabela 1</b> - Frequência de intervalos de terça maior e menor em melodias eruditas e folclóricas.....	83
<b>Tabela 2</b> - Dados numéricos acerca da percepção da musicalidade de seis idiomas por parte dos informantes. ....	105
<b>Tabela 3</b> - Diferença nos valores relativos à percepção de musicalidade da língua alemã por parte de falantes e de não-falantes do idioma.....	110

## SUMÁRIO

<b>INTRODUÇÃO.....</b>	<b>12</b>
<b>1 REVISÃO BIBLIOGRÁFICA: TRABALHOS PRECURSORES E INSPIRAÇÕES..</b>	<b>18</b>
1.1 Música e Língua: Limites e intersecções .....	18
1.2 Linguagem verbal: Prosódia .....	21
1.3 Reconhecimento de emoções.....	25
1.4 A relação entre língua e música.....	29
1.5 Considerações finais .....	34
<b>2. EXPERIMENTO 1: OS INTERVALOS DA FALA.....</b>	<b>36</b>
2.1 Introdução .....	36
2.2 Conceitos prévios.....	36
2.3 Metodologia .....	38
2.4 Resultados .....	47
2.5 Discussão .....	49
2.5 Considerações finais .....	53
<b>3. EXPERIMENTO 2: QUAL EMOÇÃO ESSA MÚSICA EXPRESSA? .....</b>	<b>54</b>
3.1 Introdução .....	54
3.2 Conceitos prévios.....	54
3.3 Metodologia .....	60
3.3.1 Informantes .....	62
3.3.2 Estatística .....	65
3.4 Descrição dos estímulos, com resultados .....	68
3.5 Discussão .....	82
3.6 Considerações finais .....	84
<b>4. EXPERIMENTO 3: QUÃO MELODIOSA ESSA LÍNGUA SOA?.....</b>	<b>86</b>
4.1 Introdução .....	86
4.2 Conceitos prévios.....	91
4.3 Metodologia .....	93
4.3.1 Estímulos .....	93
4.3.2 Informantes .....	102
4.3.1 Estatística .....	103
4.4 Resultados .....	104
4.5 O que faz uma língua soar mais musical e melodiosa?.....	111
4.6 Que língua você acha que é essa?.....	119
4.7 Discussão .....	123
4.8 Considerações finais .....	126
<b>CONCLUSÕES .....</b>	<b>127</b>
<b>REFERÊNCIAS .....</b>	<b>129</b>
<b>7. APÊNDICES.....</b>	<b>148</b>

## INTRODUÇÃO

Esta tese é uma investigação acerca da relação entre a língua e a música.

A aproximação entre as duas linguagens começa já em sua constituição mais elementar. A música tem ritmo e melodia. A língua também tem ritmo, e também tem melodia (entoação). Logo, a língua tem música (prosódia). Não se sabe com certeza, pelo menos por enquanto, se a música se origina da fala, se a fala surgiu da música, ou se ambas tiveram um ancestral comum. Mas sabemos, com certeza, que a fala tem sua música, e que a música pode incluir a fala (no caso da música que incorpora a linguagem verbal na forma de letra) e frequentemente divide com ela a função expressivo-comunicativa<sup>1</sup>.

Na função de alicerce e de ponto inicial desta investigação, estão uma conjectura e uma observação anedótica. A conjectura: sendo ambas a entoação da fala e a melodia da música fenômenos relacionados ao *pitch*, a altura melódica, é possível que os mesmos padrões acústicos sejam usados em ambos os domínios para expressar emoções (FUJISAWA; COOK, 2004, p. 1333). A observação: o senso comum atribui maior ou menor valor estético a sistemas fonológicos diferentes, e esse valor estético é expresso, recorrentemente, na forma de comparação da fala com a música (ANTOLIK, 2020; SEELINGER, 2018), como se uma fala agradável aos ouvidos não soasse mais como fala, mas como uma melodia.

Há quem afirme que a entoação, parte constituinte da prosódia, é o elemento mais evasivo da linguagem e o mais difícil de se delimitar enquanto objeto de investigação científica (SOUZA, 2014, p. 10 e 108). Há quem pareça pensar o contrário. Cagliari (2007), por exemplo, argumenta na direção contrária, e atribui (pelo menos em parte) a dificuldade da investigação e posterior sistematização de conclusões acerca da prosódia à falta de uma formação linguística sólida. O autor lembra, inclusive, que a gramática mais antiga não é senão uma gramática da prosódia, visto que as reflexões mais antigas sobre a linguagem foram feitas sobre o material poético, para explicitar e prescrever padrões poéticos de metrificação (CAGLIARI, 2007, p. 16). A prosódia é o principal recurso linguístico nas fases iniciais do desenvolvimento infantil, permitindo à criança a assimilação da voz e das posições dos interlocutores, ao mesmo tempo em que permite aos adultos sistematizar critérios linguísticos para atribuir sentido ao choro infantil (VASCONCELOS; VIEIRA; SCARPA, 2021, p. 42 e 43). É a prosódia, aliás, a responsável por ser a ponte entre a organização formal da fala e o potencial significativo da

---

<sup>1</sup> O que não significa dizer que o tipo de expressão e comunicação engendradas pela fala e pela música sejam exatamente os mesmos. Esta questão será retomada mais adiante.

linguagem durante a fase de aquisição da linguagem (LAZZARI; PETRY, 2020). Chega a ser irônico que, num primeiro momento, estruturalistas e funcionalistas tenham encarado a prosódia de forma negativa, como fatos de fala que não entram no quadro fonemático (BARBOSA, 2012, p. 13). Hoje, a perspectiva é outra:

[...] a prosódia está, no cenário de pesquisa atual, associada a fatores linguísticos como acento, fronteira de constituinte, ênfase, entoação e ritmo, a fatores paralinguísticos como marcadores discursivos (e.g., “né”, “entendo”, “an-han”) e atitudes proposicionais (e.g., “confiante” e “duvidoso”) e sociais (e.g., “hostil” e “solidário”), além de tratar de fatores extralinguísticos como as emoções. Todos esses fatores se combinam com aspectos sociais e biológicos indiciais como gênero, faixa etária, classe social, nível de escolaridade, entre outros. (BARBOSA, 2012, p. 13)

Ao abordar a limitação da análise automática da prosódia por meio de algoritmos, Souza (2014, p. 14) afirma que o calcanhar de Aquiles deste tipo de análise é o fato de a prosódia revelar, em última instância, a intenção, um reflexo da consciência do falante que guia o enunciado – mas que, até o momento, não sabemos definir. Esta possível dificuldade metodológica, no entanto, não constitui um impedimento para que o autor desenvolva uma defesa da prosódia como o ponto de partida dos enunciados, e não algo acessório aplicado a posteriori (SOUZA, 2014, p. 11). Cagliari (2007), por outro lado, adverte contra definições pouco precisas. Ao mencionar os estudos voltados ao que se refere como “atitudes do falante”, o autor lembra que tal noção (a de “atitudes do falante”)

[...] tem uma definição precisa. Trata-se do uso de elementos prosódicos para acrescentar ao enunciado básico um elemento emocional que o falante quer que constitua um contexto psicológico no qual o enunciado deve ser interpretado. Atitude do falante não é pressuposto nem conotação, fenômenos diretamente ligados ao discurso e à semântica do enunciado, os quais não precisam necessariamente dos recursos da prosódia para se realizarem. Como os elementos prosódicos, esses elementos semânticos e discursivos devem sempre estar presentes na fala, a todo instante, mas não se confundem com as atitudes do falante. Dizer que as atitudes do falante são manifestações emocionais do falante (ou que o ouvinte assim interpreta) é descaracterizar o sentido mais exato e preciso que a tradição fonética atribuiu a essa expressão. [...] há um sistema que controla os limites do que é possível e do que já não pertence mais ao fenômeno. (CAGLIARI, 2007, p. 33-34)

Como vimos, uma das informações que a prosódia carrega é, como coloca Cagliari (2007), o elemento emocional que constitui um contexto psicológico dentro do qual o enunciado (e, portanto, o conteúdo semântico), deve ser interpretado. A forma como esse elemento emocional é expresso na língua é uma das questões mais importantes que abordaremos neste

trabalho e o primeiro (suposto) elo entre música e linguística que nos propomos a verificar. Aliás, o próprio autor menciona, no mesmo texto que acabamos de citar, a importância de trabalhos que aproximem a língua e a música:

Infelizmente, não faz parte dos interesses de grande parte dos foneticistas uma preocupação com os estudos do canto e da música, em geral. A comparação entre a fala e a música é muito relevante para os estudos prosódicos. Abercrombie costumava dizer que os foneticistas precisam, de vez em quando, olhar também as teorias musicais. Pike também tinha um interesse na comparação entre os estudos de música, sobretudo do canto, e da fonética (CAGLIARI, 2007, p. 17)

O simples fato de nomes como Abercrombie, Pike e Cagliari advogarem pela aproximação entre as duas áreas já é, de nossa perspectiva, motivo suficiente para justificar uma empreitada como a nossa. A seguir, descreveremos a estrutura desta tese, já adiantando algumas das inquietações que nos guiaram ao longo do caminho.

O primeiro capítulo desta tese apresenta uma revisão bibliográfica geral dos conceitos que serão mobilizados ao longo de toda a tese, bem como trabalhos anteriores que nos serviram de inspiração, e reflexões que, a nosso ver, devem ser consideradas ao longo das seções posteriores. Abordamos, primeiramente, os limites e as interseções de música e língua, inclusive a possibilidade de que as duas modalidades de comunicação sejam polos opostos em um espectro, sem uma divisão clara entre as duas. Passamos, a seguir, para uma seção na qual falamos um pouco sobre o conceito de prosódia, que nos será caro ao longo do trabalho. Seguimos com menções a trabalhos acerca do reconhecimento de emoções na fala e na música, ou sobre a forma como música e fala se relacionam e se influenciam mutuamente. Feitas as apresentações, seguimos para o cerne desta tese, que consiste em três experimentos.

Ainda que tratem, em linhas gerais, do mesmo tema – coincidências na forma, no uso e na percepção da língua e da música – e apesar do primeiro e do segundo experimento em especial, como veremos mais adiante, complementarem um ao outro, cada operação mobilizou conceitos diferentes e partiu de uma perspectiva levemente distinta. Assim, optamos por dedicar uma seção a cada experimento. Cada seção primária está subdividida em: uma “Introdução”, na qual expomos a problemática a ser tratada; uma subseção “Conceitos prévios”, na qual mobilizamos definições e conceitos teóricos relevantes para o experimento em questão; subseções “Metodologia” e “Resultados”, descrevendo os procedimentos utilizados e os resultados obtidos em cada experimento; segue-se a “Discussão”, na qual oferecemos nossas interpretações desses resultados, e as “Considerações finais” acerca do experimento em questão e seus respectivos resultados.

Os dois primeiros experimentos se apoiam sobre os conceitos de intervalos musicais e de tonalidade. Já ao longo de alguns anos, alguns trabalhos procuraram aproximar padrões formais da expressão da fala emotiva (entendida aqui como atitude do falante, assim como na definição mencionada anteriormente na citação de Cagliari, 2007, p. 33-34) com padrões musicais. Entre outros aspectos, estas aproximações frequentemente alegam descobrir uma correspondência entre intervalos musicais específicos que seriam usados tanto na música, quanto na fala, para expressar certos estados emocionais. Um trabalho deste tipo, por exemplo, alega que o padrão entoacional correspondente ao intervalo de terça menor, na fala de língua inglesa, está associado à expressão de tristeza, o que espelharia o uso desse intervalo na música (CURTIS; BHARUCHA, 2010). As aproximações entre as duas linguagens, porém, nem sempre observam uma definição teórica rígida, como pode ser observado no seguinte trecho:

*In music the difference between sad and cheerful melodies is often indicated as a difference between a **minor and a major key**. Our main objective is the identification of analogous **interval differences** in the pitch contours of emotional speech in Dutch. It is obvious that the range in the pitch contour of sad speech is much smaller than the range in cheerful speech, but do we also **speak in a minor key when we are sad and in a major key when we are happy?**<sup>2</sup> (SCHREUDER, 2006, p. 151, grifos nossos)*

No trecho citado, a autora cita a tonalidade musical como sendo um dos fatores indicativos da expressão de alegria e tristeza em uma melodia. Porém, no intuito de descobrir a “tonalidade da fala” quando o falante expressa uma atitude triste ou alegre, a autora recorre a intervalos, sendo que a mera presença de intervalos musicais específicos é um critério bastante problemático para determinar a tonalidade de uma melodia, e mais ainda para associar uma melodia a uma atitude/emoção específica. Estas duas problemáticas serão abordadas nas seções dois e três, nas quais apresentamos os dois primeiros experimentos desta tese.

A seção dois traz um experimento que buscou, grosso modo, replicar (em algum grau) os resultados obtidos por Curtis e Bharucha (2010). Analisamos amostras bissilábicas de falantes de Português Brasileiro expressando emoções diferentes, na expectativa de poder associar intervalos musicais definidos a emoções específicas. Não apenas nossos dados não corroboraram os resultados dos pesquisadores norte-americanos, como nos levaram a uma avaliação posterior que, apesar de não deslegitimar, problematiza a metodologia e as conclusões do trabalho.

---

<sup>2</sup> Em tradução livre: “Na música, a diferença entre melodias tristes e alegres é frequentemente indicada como uma diferença entre um tom menor e um tom maior. Nosso objetivo principal é a identificação de diferenças de intervalo análogas nos contornos de altura melódica da fala emocional em holandês. É óbvio que a tessitura no contorno da altura melódica da fala triste é muito menor do que a tessitura no discurso alegre, mas será que também falamos em tom menor quando estamos tristes e em tom maior quando estamos felizes?”

A seção três traz o segundo experimento. Buscamos verificar se é mesmo verdade que o intervalo de terça menor está mesmo associado à tristeza e o intervalo de terça maior à alegria. Se é verdade que a) o intervalo de terça menor é um marcador de expressão de tristeza na música e possivelmente também na fala (CURTIS; BHARUCHA, 2010); e b) que nossa fala tem um tom determinado e/ou que falamos usando “tom maior” quando estamos felizes e “tom menor” quando estamos tristes (SCHREUDER; VAN EERTEN; GILBERS, 2004, SCHREUDER, 2006, COOK; FUJISAWA; TAKAMI, 2004); decidimos, então, verificar se a (maior) presença de intervalos de terça maior ou menor em uma melodia cantada leva os ouvintes a interpretá-la como expressando tristeza ou alegria.

Nossa motivação para usar um *corpus* musical no segundo experimento segue a seguinte linha de raciocínio, que pretendemos investigar:

- Premissa um: seres humanos são capazes de reconhecer a atitude do falante e a carga emocional de um enunciado, mesmo em uma língua que não falam, com um índice de acerto acima do esperado em uma situação de acaso (ver, por exemplo, Peres, 2009);

- Premissa dois: existem evidências sólidas de correspondências na expressão de emoções na música e na fala;

- Premissa três: é possível que intervalos de terça menor e terça maior sejam indicadores de tristeza e alegria, respectivamente, tanto na música quanto na fala (CURTIS; BHARUCHA, 2010, SCHREUDER; VAN EERTEN; GILBERS, 2004)<sup>3</sup>;

- Hipótese: se as premissas anteriores estão corretas, a presença de intervalos de terça maior e de terça menor pode guiar a interpretação de um enunciado melódico cantado, na direção da alegria ou da tristeza.

Como veremos, de fato os intervalos guiam a percepção dos ouvintes, mas não de uma forma tão simples e direta como alguns trabalhos da literatura parecem sugerir.

O terceiro experimento é apresentado na seção quatro. A esta altura, voltamo-nos para outro aspecto da relação entre língua e música: a percepção, no senso comum, de que algumas línguas soam como mais melodiosas, musicais e/ou agradáveis do que outras. Procuramos verificar se há algo concreto nesta avaliação evidentemente subjetiva. Apesar da escassez de trabalhos que abordem o sistema fonético-fonológico das línguas sob essa perspectiva, acreditamos trazer evidências que não devem ser desconsideradas, especialmente no que se refere à importância da variação da altura melódica na percepção de “musicalidade” de um idioma.

---

<sup>3</sup> E, igualmente importante, não havendo (até onde pudemos averiguar) trabalhos na literatura que neguem veemente ou totalmente as conclusões apresentadas pelos trabalhos elencados.

Nossas breves reflexões acerca do trabalho em geral – e dos três experimentos em particular – fecham esta tese na seção “Conclusões”, a qual é seguida pelas referências e pelos apêndices. Os resultados obtidos neste trabalho podem ser resumidos em três pontos: 1) a expressão das emoções por parte dos falantes de português brasileiro não apresentou correspondências ao uso de intervalos musicais; 2) a percepção da carga emocional de um enunciado melódico não foi correlacionada com a presença de intervalos específicos, mas com a percepção de tonalidade expressa pela melodia; 3) há evidências de que um menor número de consoantes e, principalmente, maior variação da altura melódica, são fatores ligados à percepção de uma língua como mais musical e melodiosa. Para além destes resultados, argumentamos que vários trabalhos anteriores comparando a expressão na música e na fala não apresentam a clareza necessária na definição de conceitos, e apontamos a investigação da percepção estética das características fonéticas e fonológicas dos idiomas como um campo potencialmente contencioso, mas promissor.

## 1 REVISÃO BIBLIOGRÁFICA: TRABALHOS PRECURSORES E INSPIRAÇÕES

Nesta seção, faremos uma apreciação de alguns trabalhos que, quando não diretamente relacionados à pesquisa aqui apresentada, no mínimo apresentam temas e elucubrações interessantes, que serviram de inspiração para a concepção da pesquisa e de base para as discussões desenvolvidas.

### 1.1 Música e Língua: Limites e intersecções

É natural que linguistas em geral, fonólogos e foneticistas em particular, demonstrem grande curiosidade e sejam atraídos para o domínio da música. Afinal, a música é uma linguagem artística com alta penetração social, e está presente dos cultos religiosos às *raves*, demandando atenção em salas de concerto e como a trilha sonora que torna as infinitas tarefas domésticas mais suportáveis. A música, de uma forma ou de outra, é um fenômeno quase onipresente e, em que pese a existência da música instrumental e do *scat singing*<sup>4</sup> de vozes como a americana Ella Fitzgerald, frequentemente apresenta um componente linguístico marcante, na forma de um texto sendo cantado, ao qual nos referimos pelo termo *letra*.

Mas não é só o elemento indubitavelmente linguístico que pode ser de interesse ao linguista. Afinal, por mais que o nome dos cursos de graduação em *Letras* possa nos induzir ao erro, a maioria esmagadora das línguas naturais do mundo<sup>5</sup> é essencialmente oral – e, assim, formada da mesma materialidade da música e do canto. Como veremos ao longo de diversas passagens desta tese, não apenas fala e música lançam mão de recursos com a mesma materialidade, mas também podem apresentar intersecções notáveis. De fato, há que se questionar se é realmente possível estabelecer um limite rígido entre fala (e, por extensão, língua oral) e canto (música). Esta seção aborda alguns dos trabalhos que se ocuparam de comparar fala e música, ou verificar as dinâmicas apresentadas por essas modalidades.

Apesar da existência de enunciados falados que podem soar como se fossem cantados, humanos em geral distinguem fala de canto de forma bastante intuitiva (MEDEIROS; CABRAL, 2018, p. 542). Infelizmente, de acordo com Medeiros e Cummins (2014), comparações entre fala e canto nos estudos linguísticos são relativamente raras, e o entendimento acerca do que, *exatamente*, torna um enunciado musical ainda não está totalmente

---

<sup>4</sup> Tipo de improvisação vocal típica do jazz, que usa sílabas e fonemas desconexos ou até mesmo palavras inventadas, de modo que o intérprete parece estar usando a voz mais como um instrumento musical "puro" do que como um veículo para um texto (BRITANNICA, 2012, n.p.).

<sup>5</sup> A maioria, mas não a totalidade, haja vista a existência de diversas línguas de sinais. Em consulta realizada em junho de 2021, verificou-se que a enciclopédia *Ethnologue* (EBERHARD; SIMONS; FENNIG, 2021) editada pelo SIL International, lista 7.139 línguas orais em seu catálogo, mas “apenas” 150 línguas de sinais.

estabelecido - o que não significa que estudos que abordem fenômenos linguísticos com *corpus* musical, por exemplo, não constituam um campo prolífico (como exemplo, ver Massini-Cagliari, 2015).

Magdics (1963), ao comparar o que chama de "Melodia da Fala" e a "Melodia da Música", estabelece as *popular dirges* (cantos fúnebres) e os recitativos como linha divisória entre música e fala, sendo os critérios para tal o fato de a "melodia da música" ter alturas melódicas definidas com precisão, regularidade rítmica precisa e ausência de nuances de significado definido associado com seus intervalos, progressão melódica, andamento e dinâmica (MAGDICS, 1963, p. 338). Quatro décadas depois, Cummins (2013) volta a colocar fala e música como polos em um contínuo; no entanto, para isso, abandona o que seria uma visão linguística tradicional para adotar o que chama de "coordinative view of speech" (CUMMINS, 2013, p. 19), levando em conta a interação e a coordenação das atividades dos seres humanos. Deutsch, Henthorn e Lapidis (2011) postulam que, em geral, a diferenciação entre a fala e o canto depende de características acústicas, em especial a altura melódica. Coerentemente com Deutsch, Henthorn e Lapidis (2011), Merrill e Larrouy-Maestri (2017) também buscam nos elementos da fonação os indícios acústicos que levam os ouvintes a classificarem um excerto como "mais cantado" ou "mais falado". De fato, o entendimento do canto como uma variante específica da fala (MEDEIROS; CUMMINS, 2014, p. 749), ou de canto e fala como polos de um contínuo (CUMMINS, 2013, p. 19), pode parecer natural.

É importante ressaltar, mais uma vez, que, apesar das evidências de que características acústicas de um enunciado levam os ouvintes a classificá-lo como fala, música ou outra modalidade intermediária, ainda estamos longe de tirar conclusões categóricas. Assumir que a diferença entre (um enunciado ser categorizado como) fala ou canto reside apenas em suas propriedades acústicas também não explica a ilusão reportada por Deutsch, Henthorn e Lapidis (2011), que mostraram que um trecho de um enunciado falado, se repetido em *looping* várias vezes, passa a ser percebido como um segmento melódico. Apesar de mencionar que em geral se assume que o canto apresenta maior estabilidade de  $f_0$  do que a fala, a hipótese desse último trabalho é a de que a repetição aumenta a saliência perceptual das alturas melódicas da sequência, que passa a ser percebida como uma melodia tonal (DEUTSCH; HENTHORN; LAPIDIS, 2011, p. 2246).

Cabe lembrar que medidas acústicas não falam por si só, sendo necessário sensibilidade e perspicácia dos pesquisadores para interpretá-las. Os resultados de um primeiro trabalho de Medeiros e Cabral (2018), por exemplo, surpreenderam ao apontar que o canto, pelo menos no *corpus* analisado por eles, não era mais estável do que a fala, ou seja, o canto não apresentou

maior estabilidade de altura melódica quando comparado à fala, o que ia na contramão dos resultados gerais reportados na literatura<sup>6</sup>. Posteriormente, com uma segunda análise dos mesmos dados (desta vez considerando o efeito que mudanças abruptas de f0 que ocorrem em glissandos ou vibratos, por exemplo), confirmou-se a hipótese de maior estabilidade de f0 no canto (MEDEIROS; CABRAL; MEIRELES; BACETI, 2021).

Outros estudos buscaram desvelar a natureza da relação entre fala e canto/música a partir do ponto de vista neurológico e cognitivo. Kolinsky et al (2009), por exemplo, buscaram verificar se, em uma canção, a dimensão fonológica da letra e a dimensão melódica da composição são processadas de forma integrada ou independente. Entre outras conclusões, chegaram ao fato de que há maior interação entre o processamento de vogais e das melodias do que entre consoantes e melodias (quando são usadas *nonwords*, sequências de sons e sílabas que não constituem palavras), e que a presença (ou não) de conteúdo lexical influencia tais interações. Thompson, Marin e Stewart (2012) sugerem que a maior dificuldade que pacientes com amusia apresentam para decodificar a prosódia na língua indica que música e língua compartilham mecanismos de processamento. Este último trabalho citado é especialmente interessante, pois os pesquisadores se voltaram para um aspecto particular da prosódia, a expressão de emoções na fala, tomando aspectos como duração, intensidade e altura melódica como pistas acústicas. A menção ao trabalho de Thompson, Marin e Stewart (2012), assim, converte-se num ensejo para abordar outra perspectiva da interface canto/música e língua/linguística: a *expressividade*.

Em inglês, idioma de grande parte da bibliografia dessa pesquisa, vários textos se lançam a discutir se a música seria (um tipo de) *language* (ver, por exemplo, Bright, 1963; Levman, 1992; Tolbert, 2001 e Zbikowski, 2012). Neles, apesar de a música aparecer em oposição a *language*, por vezes é difícil saber se os autores usam a palavra *language* no sentido de “língua” ou de “linguagem”, uma diferenciação feita com facilidade no português, mas problemática em inglês. Frequentemente, nessas discussões, levanta-se o questionamento acerca da (in)capacidade da música de, por si só, expressar significados. Uma discussão realmente aprofundada acerca da natureza de música enquanto sistema, seja ela uma língua ou uma linguagem, foge ao escopo deste texto, mas é bastante claro que a música instrumental é incapaz, por exemplo, de passar uma informação direta e precisa, como a posologia de um

---

<sup>6</sup> Cabe mencionar a existência de estudos (MEDEIROS; CABRAL, 2018, MEIRELES; MEDEIROS; CABRAL, 2020) que analisaram o canto no estilo MPB. Este detalhe é significativo, uma vez que diferentes estilos podem se aproximar mais ou menos da fala. De modo geral, o canto na MPB apresenta uma estrutura harmônica mais bem definida, se comparado com a fala. Ainda assim, o canto na MPB ainda é mais próximo da fala do que o canto lírico, por exemplo.

medicamento, por exemplo – informações como esta são expressas facilmente através da linguagem verbal.

## 1.2 Linguagem verbal: Prosódia

Uma vez que a linguagem verbal possibilita a expressão mais direta e precisa de conteúdos proposicionais, e pensando a comunicação a priori como produção e transmissão de significados, talvez não seja exagero dizer que a linguagem verbal é veículo de comunicação por excelência. Em que pese a importância de outras formas de linguagem não verbal, como a visual, é difícil argumentar contra a importância – por vezes, o protagonismo – da língua na vida em sociedade. A linguagem verbal se presta à veiculação dos conteúdos semânticos, mas também à expressão de outros conteúdos, como o estado emocional e as atitudes do falante em relação a si mesmo, ao interlocutor e/ou ao conteúdo veiculado. Seja através da pontuação (no caso da escrita), de aspectos prosódicos (na fala), da escolha lexical ou do uso de estruturas gramaticais específicas, as possibilidades expressivas são inúmeras e exploradas não apenas no uso artístico da língua, como na poesia, mas por também pelos falantes na vida cotidiana, em maior ou menor grau e em níveis mais ou menos conscientes. É na veiculação desses conteúdos “paralinguísticos” que reside um dos interesses dessa pesquisa, mais especificamente, na forma como emoções podem ser expressas através de aspectos fonético-fonológicos do sistema linguístico, e a possível correspondência entre esses aspectos fonético-fonológicos e aspectos da linguagem musical também associados à expressão de emoções específicas – tópicos que serão abordados em mais detalhe mais adiante. Não menos importante, outros aspectos sonoros da enunciação (que, à primeira vista, parecem não carregar significado em si mesmos) são essenciais para outro uso da língua, a manifestação das identidades dos falantes. É comum ouvir referências ao “falar cantado” do nordestino, à “tendência de engolir sílabas” no falar de Minas Gerais, e ao “ritmo diferente” da fala do gaúcho. Esses aspectos, comumente chamados de “a melodia e o ritmo da língua” até mesmo por leigos, são fenômenos no domínio da *prosódia*.

Historicamente, o termo *prosódia* era usado pelos gregos para se referirem aos traços da fala que não podiam ser indicados na ortografia (COUPER-KUHLEN, 1986, p. 1). Massini-Cagliari (2015, p. 20) considera que a fala tem melodia (entoação, tons) e pulsação (acento, ritmo e duração) e toma esses fatores como prosódicos. Em uma definição sucinta, pode-se dizer que a prosódia inclui todos os aspectos acústicos da fala não limitados ao segmento e que possuem funções linguísticas (CÓRDULA, 2013, p. 26). Entre esses aspectos, pode-se citar a altura melódica, a intensidade e o timbre. Para além dos conteúdos paralinguísticos

mencionados anteriormente, a manipulação destes aspectos tem uma função comunicativa capital: através dela, podemos diferenciar uma pergunta de uma afirmação, definir o tópico frasal e expressar atitudes do falante, por exemplo, seu estado emocional, ou seja, se o falante está feliz, triste, com raiva, etc. (MASSINI-CAGLIARI; CAGLIARI, 2001, p. 118-119). Em outras palavras, a prosódia molda a enunciação, imprimindo àquilo que se fala um modo de falar que é dirigido, intencionalmente ou não, ao ouvinte (BARBOSA, 2012, p. 13-14).

Um exemplo da importância da prosódia na expressão de atitudes do falante na comunicação é o fato de que costumamos nos apoiar justamente na entoação para interpretar um enunciado como irônico ou sarcástico. O papel da entoação na veiculação dessa informação também pode ser percebido facilmente se pensarmos que um enunciado como “Abra a porta” não carrega, em estrutura, implicatura semântica ou mesmo no léxico, qualquer informação acerca de estados emocionais, mas, enquanto falantes, sabemos que uma pessoa alegre provavelmente diria essa frase de forma bastante distinta de uma pessoa que estivesse em um acesso de fúria. Da mesma forma, somos capazes de identificar com razoável confiabilidade a presença de uma alteração no estado emocional do falante mesmo sem ver sua expressão facial ou corporal, apenas ouvindo sua fala.<sup>7</sup>

Podemos observar, então, que a prosódia é uma fonte riquíssima de informações: para além de seu papel linguístico organizando o sinal sonoro no tempo (definindo fronteiras de constituintes, ênfases, abarcando entoação e ritmo), carrega também indícios de fatores extralinguísticos, como as emoções (BARBOSA, 2012, p. 13)

Neste trabalho, apesar de não desconsiderarmos a intensidade, o timbre e o ritmo, tratamos predominantemente da entoação, devido ao possível paralelo a ser desenhado entre linguagem verbal e expressão musical, a ser desenvolvido mais adiante no texto. Assim, é necessário adotar uma definição simples que norteie nosso raciocínio. Segundo Barbosa (2012, p. 19), há certo nível de ambiguidade entre os termos prosódia, ritmo e entoação. Segundo o autor, existem basicamente dois recortes possíveis:

Autores como Hirst e Di Cristo [...] tomam a prosódia como o termo mais geral, compreendendo tanto o domínio propriamente lexical, [...] quanto o domínio não-lexical ou pós-lexical, ao qual referem a variação no enunciado da frequência fundamental, da duração e a marcação de fronteiras prosódicas. [...] Sendo assim, para esses autores (entre outros), entoação não concerne apenas a percepção do *pitch* (altura) ao longo do tempo. A entoação também abarcaria o estudo da duração e seu correlato físico e perceptivo ao longo do enunciado, o ritmo. (BARBOSA, 2012, p. 19)

---

<sup>7</sup> Isso não significa que aspectos fora do domínio da fonologia, como a escolha lexical, não possam ser relevantes na expressão de atitudes do falante, mas que, em se tratando da língua falada, aspectos prosódicos são extremamente importantes no processo.

Outro recorte possível, ainda segundo Barbosa (2012), seria restringir o termo "entoação" à análise das variações de altura ao longo do enunciado. Sob essa perspectiva, "o ritmo faz parte da prosódia, mas é independente da entoação, uma vez que compreende as variações de duração percebida de unidades do tamanho da sílaba ao longo do enunciado" (BARBOSA, 2012, p. 20). Esta é a perspectiva que adotamos neste trabalho. Em outras palavras, entendemos que o termo entoação se refere às flutuações da altura melódica da fala e aos padrões dessas flutuações (PIKE, 1945, p. 20) sendo a altura melódica representada pela frequência fundamental ( $f_0$ ) da voz durante a fonação. Tal interpretação também fica evidente em Lehiste (1970), quando a autora afirma que a entoação é análoga à taxa de elocução, mas não se confunde com ela. Na definição da própria pesquisadora, enquanto a entoação consiste no uso de características tonais para carregar informações linguísticas e significados não-linguísticos, a taxa de elocução (à qual ela se refere usando o termo *tempo*) consiste no uso de características de duração, a nível de sentença, para refletir atitudes do falante (LEHISTE, 1970, p. 95). A entoação carrega informações linguísticas, essenciais para a configuração de um enunciado em determinado sistema linguístico, e também informações paralinguísticas, que não interferem na gramaticalidade do enunciado, mas guiam a interpretação deste.

Como vimos, a entoação está longe de ser “apenas” um elemento organizador dos segmentos sonoros no tempo, mas, segundo Vassoler e Martins (2013), foi só a partir de meados da década de 70 que a entonação passou a ser mais bem compreendida em suas funções linguísticas e paralinguísticas. Nesse período, surgem várias publicações importantes sobre o tema, entre as quais se pode citar Libermann (1975), Maeda (1976) e Pierrehumbert (1980). Sobre esses estudos, Vassoler e Martins (2013, p. 10, grifo nosso) afirmam que:

Esta geração de pesquisadores preocupou-se, principalmente, em definir e caracterizar linguisticamente o fenômeno da entoação. Para este fim, assumiram um pressuposto teórico-metodológico inicial: a entoação deve ser compreendida, apenas como a variação de  $F_0$ . Tal definição, embora operacional, incluía ainda os tons lexicais, como aqueles que ocorrem nas línguas da família *bantu* ou no chinês, por exemplo. Por esta razão, uma segunda definição teve que ser assumida: a entoação é pós-lexical, não sendo intrínseca aos itens do léxico. Ainda que se note, em alguns casos, tons cristalizados em itens lexicais, como no marcador conversacional “né”, ao qual se atribui uma curva ascendente. [...] A terceira definição assume que pode haver dois tipos de informações veiculadas pela entoação, sendo uma linguística e outra paralinguística. A informação linguística, diz respeito a aspectos especificamente linguísticos, como a asserção ou a interrogação, ao passo que **a informação paralinguística transmitiria informações suplementares que contribuem para a compreensão da linguagem de modo geral [...]**.

Desde então, foram vários os pesquisadores que se debruçaram sobre a prosódia, e uma tentativa de fazer um levantamento exaustivo desses nomes está fadada a cometer injustiças. Porém, considerando o contexto acadêmico no qual a presente tese de doutorado foi desenvolvida, bem como os profissionais que contribuíram, de alguma forma, para sua conclusão, é justo mencionar ao menos alguns nomes importantes que deixaram sua marca nos estudos da prosódia, no Brasil.

Cagliari, além de trabalhos que são referência na área de alfabetização (ver Cagliari, 1997 e 2008), produziu algumas das obras mais marcantes sobre fonética e fonologia do Português Brasileiro. Já no começo da década de 80, sua tese de livre-docência (CAGLIARI, 1981) apresentava uma descrição fonética riquíssima do Português Brasileiro, a qual era encerrada com um estudo sobre o ritmo da fala e um estudo sistemático dos padrões entoacionais do Português Brasileiro. Uma das linhas de pesquisa mais interessantes desenvolvidas pelo autor nos anos seguintes é a investigação do que ele chama de "marcadores prosódicos", recursos gráficos e lexicais que permitem ao leitor recuperar em um texto escrito elementos de natureza prosódica, como ritmo e qualidade de voz (CAGLIARI, 1989, 2003)<sup>8</sup>.

Cabe destacar também os trabalhos de Massini-Cagliari, que desde o começo da década de 90 tem apresentado uma produção bastante diversificada na área dos estudos linguísticos<sup>9</sup>. Porém, o grande destaque da carreira da autora é, certamente, seu trabalho com prosódia, unindo fonética, fonologia e linguística histórica. Sua tese de doutorado (MASSINI-CAGLIARI, 1995) inaugura, no Brasil, uma metodologia para o estudo da prosódia de línguas "mortas" ou de períodos passados da língua utilizando um *corpus* poético-musical – no caso em questão, cantigas trovadorescas medievais. Esta proposta ousada, posteriormente revista e republicada (MASSINI-CAGLIARI, 1999), continua a render frutos quase três décadas depois, tanto em trabalhos de autoria própria como em coautorias e orientações. Dentre eles, podemos citar: Costa (2006 e 2010); Barreto, Costa, Mielo e Massini-Cagliari (2019); Favaro, Cangemi e Massini-Cagliari (2020); Fonte e Massini-Cagliari (2021); e Massini-Cagliari (2015 e 2021).

Partindo de um tipo de trabalho mais "tradicional", por assim dizer, de descrição fonético-fonológica (FERREIRA NETTO, 2007), Ferreira Netto e seus parceiros de pesquisa têm desenvolvido trabalhos bastante sofisticados envolvendo descrição automática da prosódia no Português Brasileiro por meio de um aplicativo patenteado (FERREIRA NETTO, 2010). As

---

<sup>8</sup> Sobre este tema, uma leitura bastante didática e clara é o texto de Pacheco (2007).

<sup>9</sup> Tal versatilidade pode ser comprovada se compararmos trabalhos tão distintos quanto Massini-Cagliari (2001) e O'Neill e Massini-Cagliari (2021).

análises encabeçadas pelo pesquisador vão desde uma descrição mais “geral” da língua (ver Ferreira Netto e Consoni, 2009; bem como Ferreira Netto e Peres, 2009) até uma busca por padrões observáveis na fala emotiva (ver, por exemplo, Ferreira Netto, Peres, Martins e Vieira, 2014; e Martins e Ferreira Netto, 2017).

Um direcionamento diferente na abordagem dos fenômenos prosódicos pode ser encontrado nos trabalhos de Scarpa, envolvendo aquisição da linguagem (ver Scarpa, 1995, 1999 e 2005; e, em coautoria, Bailey, Plunkett e Scarpa, 1999). Em outra direção, Fernandes, também assinando como Fernandes-Svartman, traz uma aproximação entre prosódia e sintaxe, frequentemente comparando diferentes variedades do Português (ver Fernandes, 2007; e Fernandes-Svartman, 2009; para uma comparação entre as variedades brasileira e europeia do Português, remetemos a Frota *et al.*, 2014, já para uma comparação entre as variedades do Português Brasileiro e o de Guiné Bissau, remetemos a Santos e Fernandes-Svartman, 2014). Por fim, podemos citar também Barbosa (2012), Medeiros (2002 e 2010), cujos trabalhos são citados em outros momentos, nesta tese.

Em suma, fica evidente que os estudos de prosódia constituem um campo bastante vasto e incrivelmente fértil, dentro do qual os mais diversos interesses e as mais distintas perspectivas podem ser acomodados.

### **1.3 Reconhecimento de emoções**

As evidências de que os seres humanos são capazes de apreender a carga emocional de enunciados mesmo em contextos interculturais e/ou sem ter acesso ao conteúdo semântico são contundentes e robustas (SCHERER; BANSE; WALLBOT, 2001; THOMPSON; BALKWILL, 2006; PERES, 2016; SILVA; BARBOSA; ABELIN; 2016). A precisão desse reconhecimento parece variar de acordo com a emoção apresentada e com a distância cultural entre os envolvidos (SCHERER; BANSE; WALLBOT, 2001, p. 87 e 89). Porém, isso não muda o fato de que os trabalhos são consistentes ao apontarem probabilidades que embasam os resultados e afastam a possibilidade do mero acaso. É cada vez mais difícil negar que, não obstante aspectos culturais específicos, parece haver um amplo território comum para todos os seres humanos no tocante à codificação e à decodificação de emoções.

A expressão de estados emocionais é outro ponto de interesse desse trabalho, ao qual retornaremos com frequência ao longo deste texto. Há um grande número de estudos que investigaram os correlatos vocais de expressões de estados emocionais. Juslin e Laukka (2003, p. 770), por exemplo, revisaram 104 estudos de expressão vocal e 41 estudos de performance

musical e verificaram semelhanças entre as duas modalidades de expressões vocais (chamados pelos autores de "canais") no que se refere à precisão com a qual emoções distintas foram comunicadas aos ouvintes e aos padrões acústicos específicos usados para expressar cada emoção. Um resultado em comum nesses estudos é que a carga emocional do enunciado influencia a frequência fundamental (f0) em aspectos como f0 média, nível de f0 e tessitura de f0 – ao mesmo tempo, resenhas posteriores desses trabalhos também mostraram influências da carga emocional em outros elementos da entoação, em especial a intensidade acústica do contorno entoacional e medidas relacionadas à duração relativa de segmentos de fala (BÄNZIGER; SCHERER, 2005, p. 254).

No entanto, o fato de já se acumularem estudos sobre o assunto e de haver elementos em comum em suas conclusões não deve nos levar a acreditar que se trata de um caso encerrado. Um ponto a ser levantado é o de que esses estudos, em sua maioria esmagadora, relatam experiências envolvendo línguas indo-europeias ocidentais (principalmente o inglês), e sabemos que línguas diferentes fazem usos distintos dos parâmetros prosódicos. Em línguas tonais, por exemplo, variações de altura melódica (padrões entoacionais) agem sobre um item lexical, fazendo com que uma mesma sequência de fonemas tenha significados distintos; por outro lado, em línguas entoacionais, a variação da altura melódica age sobre a frase, mantendo o significado lexical e morfológico, mas alterando a função sintática e semântica/pragmática (CÓRDULA, 2013, p. 26). Se reconhecermos a diferença do componente entoacional em um nível organizacional do sistema linguístico, é razoável questionar se haveria diferenças na veiculação de informações paralinguísticas, como a emoção, em diferentes línguas (ou em diferentes dialetos, ou em diferentes sociedades que falem o mesmo idioma). Até o momento, apesar de outras famílias linguísticas que não o indo-europeu estarem subrepresentadas nos estudos analisados, há evidências de que certos aspectos sejam universais, especialmente no que se refere justamente à expressão de emoções:

*[...] certain aspects of emotional expression in speech prosody appear to be universal [...] and convincing evidence points to similar prosodic codes used across cultures to convey similar emotions [...]. This is apparent, for instance, in our capacity to decode emotional meaning even in unfamiliar languages.*<sup>10</sup>  
(COUTINHO; DIBBEN, 2013, p. 659)

A “capacidade de decodificar significados emocionais mesmo em línguas desconhecidas” mencionada por Coutinho e Dikken (2013) foi o objeto de estudo de Peres

---

<sup>10</sup> “[...] certos aspectos da expressão emocional na prosódia parecem ser universais [...] e evidências convincentes apontam para códigos prosódicos semelhantes usados entre culturas para expressar emoções semelhantes [...]. Isso é aparente, por exemplo, na nossa capacidade de decodificar significados emocionais mesmo em línguas desconhecidas.” (Tradução nossa.)

(2016). Nesse trabalho, foram conduzidos três experimentos de percepção de fala envolvendo quatro emoções básicas (raiva, medo, tristeza e alegria), a partir de trechos de fala emotiva espontânea em português brasileiro. O autor averiguou que tanto participantes nativos quanto participantes não nativos de português brasileiro foram capazes de julgar as emoções com êxito, mas com a ressalva de que os informantes não nativos não apresentaram resultado significativo em um dos cenários avaliados pela pesquisa. No exterior, podemos mencionar outros exemplos de estudos sobre prosódia, expressão de emoções e percepção dos ouvintes: Banse e Scherer (1996) usaram gravações de atores expressando diferentes emoções para observar como alguns parâmetros acústicos, como perturbação de f<sub>0</sub>, ruído espectral e precisão de formantes, se relacionam na criação de perfis acústicos de expressão vocal que podem ser utilizados para diferenciar a intensidade e a qualidade de emoções distintas. Nesse mesmo âmbito, também merecem menção os trabalhos de Ferreira Netto et al (2013), Peres (2016), Scherer (1984, 2000, 2006), Scherer, Ladd e Silverman (1984). Tais estudos, entre outras contribuições interessantes, aduzem evidências da possibilidade de prever qual será a percepção, por parte dos ouvintes, da carga emocional expressa em um determinado estímulo acústico.

While expressions of anger, fear, and happiness, for example, are accompanied by increased vocal intensity including an enhanced intensity variation, expressions of sadness are usually accompanied by low intensity. Similarly, angry, happy, and fearful voices show higher levels of pitch as mainly determined by the F<sub>0</sub> as well as pitch variability compared to sad voices.<sup>11</sup> (FRÜHHOLZ; TROST; GRANDJEAN, 2014, p. 10)

Paralelamente, ou talvez não tão paralelamente assim, outra linguagem que não a verbal parece ter seu próprio padrão sonoro para expressar, se não todas, pelo menos uma emoção. Em música, a noção de que uma sequência de duas alturas melódicas cuja distância corresponda à medida chamada de “terça menor” tem uma carga afetiva triste e melancólica é bastante difundida. Porém, explicações para essa noção parecem estar calcadas na força da tradição e da convenção.

A ideia de que determinados elementos musicais podem provocar e despertar estados emocionais<sup>12</sup> específicos realmente não é nova. No período barroco, por exemplo, a Doutrina dos Afetos, também chamada de Teoria dos Afetos ou *Affektenlehre* em alemão, foi

---

<sup>11</sup> “Enquanto expressões de raiva, medo e felicidade, por exemplo, são acompanhadas por intensidade vocal aumentada, incluindo uma variação de intensidade aumentada, expressões de tristeza são geralmente acompanhadas por baixa intensidade. Da mesma forma, as vozes irritadas, felizes e temerosas mostram níveis mais altos de altura melódica, como determinado principalmente pela f<sub>0</sub> e também pela variabilidade de altura melódica em comparação com vozes tristes.” (Tradução nossa.)

<sup>12</sup> Ou paixões, outro termo usado para se referir a esses estados (BUELOW, 2001).

amplamente aceita por teóricos e compositores. Tal doutrina se constrói sobre a crença de que o uso adequado de certos recursos ou procedimentos musicais permite ao compositor criar uma obra capaz de produzir uma resposta emocional involuntária no público (BRITANNICA, 2014). Apesar da popularidade da ideia, uma teoria unificada dos afetos não foi estabelecida no período barroco. Porém, a partir do século XVII, muitos teóricos dedicaram parte de seus tratados à categorização e à descrição não só de tipos de afetos, mas também das conotações afetivas de escalas, ritmos, instrumentos, formas e estilos (BUELOW, 2001):

*These devices and their affective counterparts were rigorously cataloged and described by such 17th- and 18th-century theorists as Athanasius Kircher, Andreas Werckmeister, Johann David Heinichen, and Johann Mattheson. Mattheson is especially comprehensive in his treatment of the affections in music. In *Der vollkommene Capellmeister* (1739; “The Perfect Chapelmaster”), he notes that joy is elicited by large intervals, sadness by small intervals; fury may be aroused by a roughness of harmony coupled with a rapid melody; obstinacy is evoked by the contrapuntal combination of highly independent (obstinate) melodies. Carl Philipp Emanuel Bach (1714–88) and the Mannheim school were exponents of the doctrine.<sup>13</sup> (BRITANNICA, 2014 n.p.)*

Cabe mencionar aqui que a classificação dos recursos que despertavam cada um dos afetos não seguia necessariamente uma motivação (puramente) acústica. Por exemplo, o compositor e acadêmico alemão Johann Mattheson postula que a alegria não é senão a expansão de nosso espírito vital, do que decorre ser natural e apropriado usar de intervalos grandes e expandidos para a expressão desse afeto (MATTHESON; LENNEBERG, 1958, p. 50-51). A tristeza, por outro lado, consistiria na contração de partes de nossos corpos, sendo, portanto, os intervalos mais "estreitos" os mais adequados à expressão dessa emoção (MATTHESON; LENNEBERG, 1958, p. 51). A própria noção do compositor controlando e direcionando os afetos do ouvinte é herdeira das doutrinas de retórica e oratória greco-latinas, que se ocupavam dos discursos verbais (BUELOW, 2001):

*Until fairly late in the history of Western civilization, music was predominantly vocal and thus bound to words. Composers have therefore generally been influenced to some degree by rhetorical doctrines governing the setting of texts to music, and even after the growth of independent instrumental music, rhetorical principles continued for some time to be used not only for vocal music but for instrumental works too.<sup>14</sup> (WILSON; BELOW; HOYT, 2001, n.p.)*

---

<sup>13</sup> Em tradução livre: Estes recursos e suas contrapartes afetivas foram rigorosamente catalogados e descritos por teóricos dos séculos XVII e XVIII como Athanasius Kircher, Andreas Werckmeister, Johann David Heinichen e Johann Mattheson. Mattheson é particularmente abrangente em sua abordagem dos afetos na música. Em *Der vollkommene Capellmeister* (1739; “O Mestre de Capela Perfeito”), ele observa que a alegria é provocada por intervalos grandes, a tristeza por intervalos pequenos; a fúria pode ser despertada por uma severidade harmônica combinada com uma melodia rápida; a obstinação é evocada pela combinação contrapontística de melodias altamente independentes (obstinadas). Carl Philipp Emanuel Bach (1714-1788) e a escola de Mannheim foram expoentes da doutrina.

Por outro lado, subjacente a boa parte das evidências que se referem ao reconhecimento de emoções na voz, está o argumento de que a expressão vocal das emoções reflete aspectos fisiológicos associados a esses estados emocionais sendo expressados (SCHERER; BANSE; WALLBOT, 2001; SALOMÃO, 2016).

#### 1.4 A relação entre língua e música

Em outra frente, há uma discussão duradoura acerca da gênese da fala e da música, frequentemente representada pelo canto. Pesquisadores indagam se a fala gerou o canto, se o canto gerou a fala, ou se ambos compartilham um ancestral comum (ver, por exemplo, PINKER, 1997). Ao mesmo tempo, os próprios limites entre a fala e o canto passaram a ser questionados (CUMMINS, 2013). Se a própria diferenciação rígida entre as duas linguagens passa a ser questionada, talvez não seja tão surpreendente que trabalhos na última década tenham encontrado semelhanças na estruturação e na expressão dessas duas linguagens.

Voltando às comparações entre fala e música, é importante lembrar, de forma explícita, que a materialidade da fala é o som, a mesma materialidade da linguagem musical. Assim, música e linguagem verbal são duas das formas mais importantes através das quais os seres humanos usam o som, de modo que esses dois "departamentos" da cultura humana apresentam semelhanças importantes e pontos de contato (BRIGHT, 1963, p. 26). Acreditamos ter elencado, anteriormente, argumentos suficientes para afirmar com segurança que a natureza específica da relação entre música e língua, sua possível hierarquia e dinâmica, ainda não é *completamente* compreendida, mas que passos estão sendo dados na direção dessa compreensão.

Alguns estudos, por exemplo, buscaram verificar se e como a língua falada pode influenciar a *composição* musical. O trabalho de Patel, Iversen e Rosenberg (2006) mostrou que composições instrumentais refletem a prosódia da língua falada pelo compositor. Uma análise do contraste na duração de vogais sucessivas na fala e notas na música, bem como padrões de variação de intervalos melódicos na fala e na música, revelou que os padrões

---

<sup>14</sup> Em tradução livre: Até pouco tempo na história da civilização ocidental, a música era predominantemente vocal e, portanto, vinculada às palavras. Portanto, os compositores em geral foram influenciados até certo ponto pelas doutrinas retóricas que governam adaptação dos textos em música e, mesmo após o crescimento da música instrumental independente, os princípios retóricos continuaram por algum tempo a ser usados não apenas para música vocal, mas também para obras instrumentais.

observados em amostras de frases faladas em inglês britânico e em francês também foram notados nos temas de compositores da Inglaterra e da França na virada do século XX<sup>15</sup>. Mais especificamente: no âmbito melódico, aferiu-se que, da mesma forma que a mudança da altura melódica conforme a voz se "move" de uma vogal para outra é mais uniforme na fala em francês do que na fala em inglês, também na música as mudanças de altura melódica de uma nota para outra foram mais uniformes nas composições francesas do que nas inglesas. No âmbito rítmico, o contraste de duração (*durational contrastiveness*) da fala, maior em inglês do que em francês, também foi observado em relação às notas nas composições melódicas.

Jekiel (2014) utiliza a mesma metodologia – ainda que com modificações na seleção do *corpus* - na análise comparativa entre o ritmo na fala e na música inglesa e polonesa. A aplicação do método na comparação do Polonês, uma língua que não se encaixa facilmente na classificação bipolar de ritmo silábico ou acentual (JEKIEL, 2014, p. 61), com o inglês, considerada um exemplo prototípico do ritmo acentual, levou a resultados mais ambíguos, não corroborando a relação entre ritmo de fala e o ritmo na música erudita, mas ainda reconhecendo um elo aparente entre o ritmo da fala e o ritmo da música folclórica, de forma que a hipótese não pôde ser descartada.

Curiosamente, o trabalho de Schellenberg (2012), analisando melodias com letras escritas em línguas tonais, aponta uma direção levemente distinta, ainda que não contrária. Em sua revisão dos trabalhos publicados sobre o tema, Schellenberg conclui que há variação na taxa de concordância entre tom linguístico e melodia, não apenas entre línguas, mas entre diferentes gêneros situados em um contínuo fala-canto em um mesmo idioma (SCHELLENBERG, 2012, p. 269-271). Em outras palavras, a taxa de concordância será distinta se tomarmos duas línguas tonais diferentes, como o zulu e o cantonês, e mesmo dentro de uma mesma língua, a concordância varia dependendo da composição analisada (em chinês, recitativos na dramaturgia tradicional e canções infantis apresentam maior concordância do que canções populares contemporâneas, por exemplo). É importante salientar, no entanto, que todas as taxas de concordância superaram os 33,3% que poderiam indicar mero acaso, e a significância dos testes foi confirmada por meio de testes estatísticos (SCHELLENBERG, 2012, p. 2070).

Se há cada vez mais evidências de influência de aspectos fonológicos da língua materna na composição musical, também é possível que compositores se baseiem diretamente nos padrões entoacionais da língua em suas composições musicais. Um excelente exemplo disso é

---

<sup>15</sup> A análise de Patel, Iversen e Rosenberg (2006) tomou como *corpus* as partituras de 137 temas ingleses e 181 temas franceses, bem como 20 frases faladas em inglês e 20 frases faladas em francês.

o compositor austríaco Arnold Schönberg, cuja *Sprechmelodie* tinha a intenção de ser um meio-termo entre fala e canto, baseado nos padrões entoacionais da língua alemã (MEIRELES *et al.*, 2017, p. 334). O alemão Johann Sebastian Bach foi um dos pioneiros daquilo que recebeu o nome de *recitativo secco*, incorporando a fala na partitura e permitindo que o intérprete adotasse os ritmos da fala cotidiana (MEIRELES *et al.*, 2017, p. 334). Também se pode mencionar o tcheco Leoš Janáček, que buscava representar de forma fiel a fala em suas composições (MEIRELES *et al.*, 2017, p. 335).

A representação da fala (especialmente da entoação) em uma notação musical, aliás, é um campo ainda sendo desbravado (ver, por exemplo, Simões e Meireles, 2016, e a sequência do trabalho em Meireles *et al.*, 2017). Alguns pesquisadores acreditam que o desenvolvimento de um sistema de representação da fala em notação musical pode ajudar a resolver padrões imprevisíveis da prosódia da fala e, assim, levar a um aperfeiçoamento dos modelos de prosódia atuais (SIMÕES; MEIRELES, 2016, p. 212). De fato, o *script* de estilização de contorno entoacional Prosogram (sobre o qual falaremos mais adiante) adotou a medida musical *semitons* para medir variações na entoação – não porque o objetivo da empreitada fosse exatamente verificar pontos de aproximação entre música e língua, mas devido à observação de que, mais do que as alturas melódicas absolutas, são as proporções (intervalos) entre as sequências de alturas melódicas que diferenciam os padrões entoacionais (MERTENS, 2020, p. 49).

Como se vê, as correlações entre aspectos da linguagem verbal e da música existem e não devem ser ignoradas.

Outros trabalhos vão pela direção contrária, abordando as influências da música sobre o texto no que se refere à interpretação, ou seja, tratam do que ocorre quando o texto é cantado, na forma de melodia, por um cantor. Como exemplo, podemos citar Medeiros (2002) que, após uma análise fonético-acústica de um *corpus* de canto erudito, conclui que é necessário haver uma negociação entre a fala e o canto para que seja mantida uma inteligibilidade mínima do texto. Mais especificamente, no canto erudito, não há coarticulação e as consoantes são encurtadas<sup>16</sup>.

A relação entre música e língua não se limita, no entanto, a questões de organização do sistema e de influência mútua de uma linguagem sobre a construção de enunciados na outra. Há também o aspecto *expressivo*, ou seja, *se e até que ponto* os mecanismos utilizados para

---

<sup>16</sup> Santos (2017) chegou a um resultado um tanto diferente, ainda que não contraditório, ao analisar um *corpus* de música popular. Santos conclui, entre outras coisas, que as vogais cantadas são mais longas do que as vogais faladas, mas que o padrão formântico das vogais cantadas não difere das faladas.

expressar conteúdos específicos são os mesmos em ambas as linguagens. Esse questionamento não é descabido, uma vez que:

*To the best of our knowledge, language and music are among the things that are unique to humans. It follows, then, that the solutions to the problems of creating linguistic or musical expressions, originating as they do with a single species, will exhibit broad commonalities of strategy and design, and share certain features.*<sup>17</sup> (ZBIKOWSKI, 2012, p. 128)

Nesse sentido, por exemplo, Coutinho e Dibben (2013) se basearam em trechos de música presente em filmes e em amostras de fala natural para criarem um modelo que reúne informações relevantes dos estímulos acústicos e faz previsões sobre a expressividade da fala e da música, previsões essas que se aproximaram bastante das avaliações feitas por sujeitos humanos. Com isso, os autores mostram que grande parte da percepção de emoções por parte dos ouvintes se deve a sete características psicoacústicas, entre elas intensidade, taxa de elocução e melodia/contorno prosódico.

As amostras de fala utilizadas no estudo conduzido por Coutinho e Dibben (2013) estavam em um idioma do qual os informantes não tinham conhecimento, fato este que, além de aduzir mais evidências de que a decodificação de emoções na fala é, pelo menos até certo ponto, independente da compreensão semântica, também mostra que a percepção de emoções expressas em uma composição musical também não depende de compreensão de conteúdo semântico e que mesmo composições instrumentais apresentam características que podem ser interpretadas como expressivas de emoção. Neste ponto, deve-se mencionar que, assim como a fala apresenta uma altura melódica,  $f_0$ , a melodia em uma composição musical nada mais é do que uma sequência de alturas melódicas distribuídas ao longo do tempo. A melodia, aliás, costuma ser a parte mais reconhecível de uma composição musical: uma mesma melodia pode ser executada com o acompanhamento em diversos ritmos, e na maior parte do tempo é a melodia que cantamos espontaneamente quando nos lembramos de uma composição. Não parece estranho supor, portanto, que pelo menos parte da expressividade emocional da música se encontre na melodia.

O trabalho de Curtis e Bharucha (2010) defende essa ideia. Os autores afirmam que, ainda que normalmente se assuma que a fala, ao contrário da música, não utiliza relações

---

<sup>17</sup> “Até onde sabemos, língua e música estão entre os aspectos exclusivos dos seres humanos. Segue-se, então, que as soluções para os problemas de criações de expressões linguísticas ou musicais, originadas de uma única espécie como elas são, exibirão amplas semelhanças de estratégia e design e compartilharão certas características.” (Tradução nossa.)

específicas entre alturas de  $f_0$  para expressar emoção, a análise das alturas melódicas em amostras bissilábicas mostrou que a relação entre as alturas tendia a se aproximar do que se chamaria de um intervalo de terça menor, quando a intenção era expressar tristeza. Em outras palavras, emergiu das amostras de fala coletadas pelos pesquisadores um padrão sonoro que foi utilizado da mesma forma que a música o utiliza. No intuito de validar as conclusões obtidas a partir dos resultados, outros dois experimentos foram realizados: participantes avaliaram as emoções percebidas em amostras de fala, e foi observado que, dentre vários parâmetros acústicos, o intervalo de terça menor se mostrou como o mais confiável para identificação de expressão de tristeza. Paralelamente, voluntários também avaliaram as emoções percebidas em intervalos musicais, corroborando a associação histórica entre a terça menor e a tristeza. Os autores concluem, portanto, que a terça menor comunica tristeza na fala, espelhando o uso desse intervalo na música, e que os resultados desses experimentos dão força à teoria de que as expressões vocais e musicais humanas compartilham um mesmo código acústico para comunicar tristeza.

Em consonância com Curtis e Bharucha (2010) está o trabalho de Schreuder, van Eerten e Gilbers (2004), que analisou a entoação da fala de leitores lendo passagens de ficção infantil que expressavam tristeza ou alegria. Partindo dos mesmos pressupostos, os pesquisadores utilizaram análises de agrupamentos para verificar quais frequências fundamentais estavam mais presentes nas gravações. Os contornos melódicos da gravação também foram convertidos em partituras musicais em uma análise paralela. Verificou-se, na fala de todos os leitores envolvidos no estudo, que intervalos de terça menor estavam presentes ao longo das passagens tristes, enquanto intervalos de terça maior (equivalentes a quatro semitons) estavam presentes nas passagens alegres, do que se conclui que a fala de carga emotiva, tal qual a melodia musical, envolve modalidades maiores e menores. Tais resultados vão um passo além das pesquisas anteriores, ao proporem uma correlação não só na expressão de tristeza, mas também na expressão da alegria<sup>18</sup>. Note-se, entretanto, que as análises acústicas dos trabalhos mencionados se voltaram para amostras de fala, e não de canto. Até onde se pôde verificar, essa é uma lacuna nos estudos relacionando emoção e expressão vocal.

À primeira vista, seria possível argumentar que o canto implica a entonação de uma sequência de notas – e, portanto, intervalos – musicais específica e a interpretação de intervalos musicais não parece ser diferente mesmo em contextos culturais, linguísticos e musicais distintos (MORTON; TREHUB, 2001), logo, um intervalo de terça menor seria interpretado da

---

<sup>18</sup> O que não significa que as evidências sejam robustas o suficiente para se tirar conclusões mais categóricas. Os resultados destas pesquisas serão retomados e discutidos em mais detalhes em uma seção posterior.

mesma forma quando executado em um piano ou cantado. Entretanto, o canto frequentemente envolve uma letra que acompanha a melodia, e o conteúdo semântico das palavras não deve ser ignorado, pois pode ser um fator desestabilizador da percepção “puramente musical” que se obteria de uma composição instrumental.

O trabalho de mestrado de Machado (2017), ponto de partida para este projeto de doutorado, buscou verificar se o aspecto dominante na percepção de emoções seria o lexical-linguístico ou o musical. Para tanto, comparou como um mesmo texto era percebido, do ponto de vista emocional, quando declamado e cantado. Os resultados mostraram que, ainda que os intervalos de terça maior ou menor não sejam proeminentes na melodia, as alterações rítmicas e entoacionais acarretadas pela transposição do texto para a melodia interferem de forma significativa na percepção dos ouvintes, sendo que as maiores diferenças foram quando os voluntários avaliaram se o enunciador estava alegre ou triste, excitado ou deprimido e falando a sério ou brincando. Por extensão, é possível deduzir que a aproximação entre música e linguística também pode servir como um meio para alcançar um melhor entendimento da cognição humana.

Jackendoff e Lerdahl (1980 e 1981) desenvolveram uma teoria gerativa da música tonal, baseados na gramática gerativa de Noam Chomsky, advogando que tal paralelismo entre música e língua ocorreria não devido à natureza específica dessas duas linguagens, mas pela estrutura cognitiva do ser humano – os trabalhos de Steinbeins e Koelsch (2008) e de Frühholz, Trost e Grandjean (2014), indicando que o processamento de música e da língua utiliza recursos cerebrais em comum, talvez seja um indício de que Jackendoff e Lerdahl (1980, 1981) estavam certos.

Em suma, podemos resumir os fatos elencados acima nesta citação de Meireles et al. (2017, p. 337): “In the light of what has been exposed so far, it may be said that speech and music - although with different and separate functions - appear to “meet again” in different aspects of their production/perception”<sup>19</sup>.

## 1.5 Considerações finais

Em suma, os trabalhos mencionados anteriormente mostram que música e linguagem verbal estão intimamente ligadas e compartilham muitas características expressivas, bem como

---

<sup>19</sup> Em tradução livre: À luz do que foi exposto até agora, pode-se dizer que fala e música - ainda que com funções distintas e separadas - parecem se “reencontrar” em diferentes aspectos de sua produção/percepção.

meios de produção e percepção. No entanto, até onde, exatamente, chega essa profunda ligação, e qual sua verdadeira natureza, são questões que ainda não foram totalmente respondidas. Espera-se que este trabalho contribua para este campo de estudos. Para tanto, 1-) conduzimos um experimento inspirado em um estudo previamente realizado em outro contexto, com falantes de outra língua (CURTIS; BHARUCHA, 2010), 2-) adaptamos e expandimos uma metodologia já utilizada anteriormente por Peres (2016) a um *corpus* de natureza distinta (fala lexicalizada e canto), e 3-) consideramos outro fenômeno de percepção, a “musicalidade da fala”.

## 2. EXPERIMENTO 1: OS INTERVALOS DA FALA

### 2.1 Introdução

Como vimos na seção 1.2, somos capazes de depreender significados não apenas das palavras que usamos, mas dos padrões prosódicos que imprimimos à fala. Esses padrões prosódicos estão previstos no sistema linguístico e podem ser descritos de forma objetiva. Ao contrário da linguagem verbal, os significados expressos na linguagem musical podem ser mais elusivos para a maioria das pessoas, mas isso não nos impede de depreender (ou atribuir, dependendo do ponto de vista) significados quando ouvimos uma peça musical, mesmo que nenhuma palavra seja dita. As tentativas de explicar como e por quê os seres humanos percebem a música como expressiva e comunicativa continuam surgindo e sendo postas à prova, o que garante um *crescendo*<sup>20</sup> de conhecimento, ainda que estejamos bastante longe de uma *barra dupla*<sup>21</sup>. Isto posto, alguns elementos, padrões e estruturas musicais podem carregar associações, sejam elas semânticas ou expressivas<sup>22</sup>.

Uma dessas associações é a ideia de que o intervalo musical de três semitons, chamado de terça menor, estaria associado a afetos negativos, em especial a tristeza. Como já mencionado anteriormente, os pesquisadores norte-americanos Curtis e Bharucha (2010) não apenas verificaram a associação entre esse intervalo e o sentimento de tristeza, como verificaram que o intervalo de terça menor também emerge da fala de atrizes, quando estas declamavam um enunciado de duas sílabas expressando tristeza. Neste primeiro experimento, procuramos descobrir se o mesmo padrão de intervalo seria verificado na fala de atores e músicos falantes de português.

### 2.2 Conceitos prévios

Quando as pregas vocais batem uma na outra, libertam a cada intervalo uma pequena porção de ar, resultando em uma excitação acústica periódica de ar em uma frequência igual ao

---

<sup>20</sup> O termo *crescendo*, em música, indica um crescimento gradual de volume. Em uma partitura, é indicado pelo símbolo <.

<sup>21</sup> Em uma partitura musical, a barra dupla é a linha que marca o fim de uma composição.

<sup>22</sup> Evidentemente, até que ponto se pode apreender os significados transmitidos por um código depende do conhecimento que se tem do código em questão, e em música isto não é diferente. Assim como não conseguimos entender o conteúdo semântico expresso em um idioma que não dominamos, também a nossa apreensão de significados e afetos na música dependerá do quão familiarizados estamos com a tradição musical em questão. Sendo os seres humanos capazes, desde muito cedo, de perceber padrões e fazer associações, é natural que a maioria das pessoas inserida em uma cultura dotada de uma tradição musical possua algum nível de compreensão intuitiva sobre os códigos utilizados nesta tradição – o que não implica dizer que um processo de sensibilização, formal ou não, não seja necessário para que se domine o código musical em questão.

número de batidas das pregas vocais por segundo (CAGLIARI, 1981, p.14). Essa frequência, chamada de frequência fundamental ( $f_0$ ), é a responsável por interpretarmos um som periódico como mais agudo ou mais grave. Nem todos os sons do nosso cotidiano possuem uma frequência fundamental definida. Mas, no caso do som oriundo da vibração das pregas vocais humanas (bem como naqueles produzidos por instrumentos musicais como o piano, o violino e a flauta), essa frequência fundamental – chamada em inglês de *pitch* – é manipulada para os mais distintos fins comunicativos e artísticos.<sup>23</sup>

Na música – ou, mais precisamente, na música de tradição ocidental europeia – a manipulação dessa frequência, também chamada de altura melódica, é um dos elementos basilares da estrutura musical. Na fala, a variação da altura melódica pode distinguir uma palavra de outra, em línguas tonais como o chinês (MASSINI-CAGLIARI; CAGLIARI, 2001, p. 117), pode distinguir afirmações de perguntas em línguas entoacionais como o português (MASSINI-CAGLIARI; CAGLIARI, 2001, p. 117), e expressar outros fatos semânticos e paralinguísticos, como atitudes de falantes (CAGLIARI, 1981, p. 155). Assim como diferentes línguas apresentam padrões de variação de altura melódica (padrões entoacionais) distintos, também na música o uso de alturas melódicas não é universal. Dizer que um som vibrando a 440 Hz é uma nota Lá não é senão aderir a um padrão, uma convenção de uniformização de alturas melódicas, sendo que diversas convenções são e foram usadas, tanto na Europa quanto fora dela, em vários períodos históricos (HAYNES; COOKE, 2001). As discussões apresentadas neste trabalho aderem à tradição tonal europeia, cujo padrão de afinação, que vinha se acomodando desde meados do século XIX e foi estabelecido em 1939 (HAYNES; COOKE, 2001), é certamente o mais conhecido e influente no ocidente, incluindo o Brasil, embora certamente não o único.

Cabe aqui, no entanto, fazer a ressalva de que não há uma equivalência *total, de 1 para 1*, entre a frequência fundamental ( $f_0$ ) e a percepção de altura melódica. Tan, Pfordresher e Harré (2010, p. 74) apontam que a percepção de altura melódica é multidimensional. Uma dimensão, comum aos sons tanto dentro quanto fora de contexto musical, é a frequência de vibração, que pode ser chamada de *pitch height*. No entanto, uma dimensão mais importante, do ponto de vista musical, seria a denominada *chroma*, a qual se refere à categoria (ou "classe") representada por determinada altura melódica:

---

<sup>23</sup> Generalização à parte, nem todas as vibrações das pregas vocais geram altura melódica definida. É o caso do registro vocal que se convencionou chamar, em inglês, de *vocal fry*, produzido com uma oclusão glotal frouxa que gera um som de baixa frequência. Por outro lado, é possível haver excesso de ruído acústico, dificultando a detecção da altura melódica. Esse é o caso dos registros *breathy voice*. Sobre esses registros, ver Cagliari e Massini-Cagliari (2001, p. 110-112) e Crystal (2008, p. 62 e 121).

*Chromas are identical when separated by an octave (a 2:1 ratio of a tone's fundamental frequency) and thus constitute a distinct dimension from pitch height. The fact that tones separated by an octave inhabit the same chroma category is called octave equivalence. Whithin an octave, however, changes in chroma are also changes in pitch height; the dimensions are therefore distinct but not independent.*<sup>24</sup> (TAN; PFORDRESHER; HARRÉ; 2010, p. 74)

À distância entre duas alturas melódicas se dá o nome de intervalo, sendo os intervalos tradicionalmente classificados de acordo com o número de graus que abarcam na escala diatônica (LINDLEY, 2001). Assim, um intervalo de Dó para Fá é um intervalo de quarta, enquanto um intervalo de Mi para Sol constitui uma terça. Os intervalos podem ser qualificados de forma ainda mais precisa, através do uso de termos como "maior", "menor", "justo" ou "aumentado", a depender do número de tons e semitons abarcados pela distância entre as duas alturas melódicas - ou seja, entre as duas notas (LINDLEY, 2001). Ressalte-se, aqui, que o intervalo se refere à *distância* entre duas alturas melódicas, e não ao *percurso específico de uma altura melódica específica a outra*:

*Consider the fact that most individuals, even from an early age [...], perceive a melody to be the same when it is played in a different key (a 'transposition'). If 'Over the Rainbow' is transposed from the key of C to D major, the opening phrase changes from C-C'-B-G-A-B-C' do D-D'-C#-A-B-C#-D', which changes every pitch in the sequence. Yet most people have no trouble recognizing these melodies as the 'same'. In transposed melodies, even though specific pitches may be changed, the relationships (i.e., intervals) between the pitches are retained, and it is these relationships that people primarily use when recognizing a melody.*<sup>25</sup> (TAN; PFORDRESHER; HARRÉ; 2010, p. 75)

## 2.3 Metodologia

O trabalho de Curtis e Bharucha (2010), que inspirou esse primeiro experimento, verificou uma coincidência na fala de atrizes expressando tristeza com um padrão musical, o intervalo de terça menor. No trabalho original, nove atrizes enunciaram frases bissilábicas, sendo que cada frase foi gravada quatro vezes, cada vez expressando uma de um conjunto de

<sup>24</sup> Em tradução livre: "Os cromas são idênticos quando separados por uma oitava (uma proporção de 2: 1 da frequência fundamental de um tom) e, portanto, constituem uma dimensão distinta da altura de f0. O fato de alturas melódicas separadas por uma oitava habitarem a mesma categoria de cromas é chamado de *equivalência de oitava*. Dentro de uma oitava, porém, as mudanças no cromas também são mudanças na altura de altura melódica; as dimensões são, portanto, distintas, mas não independentes."

<sup>25</sup> Em tradução livre: "Considere-se o fato de que a maioria dos indivíduos, mesmo desde a tenra idade (...), percebe que uma melodia é a mesma quando ela é tocada em um tom diferente (uma 'transposição'). Se 'Over the Rainbow' for transposta da tonalidade de Dó para Ré maior, a frase de abertura muda de Dó-Dó'-Si-Sol-Lá-Si-Dó 'para Ré-Ré'-Dó#-Lá-Si-Dó#-Ré', mudando cada altura melódica na sequência. No entanto, a maioria das pessoas não tem dificuldade em reconhecer essas melodias como as "mesmas". Em melodias transpostas, embora as alturas específicas possam ser alteradas, as *relações* (ou seja, os *intervalos*) entre as alturas são retidas e são essas relações que as pessoas usam, principalmente, ao reconhecer uma melodia."

quatro emoções básicas (raiva, felicidade, simpatia e tristeza). O uso de atores pode parecer duvidoso a princípio, mas Williams e Stevens (1972) mostraram que as expressões vocais de emoção produzidas por atores são semelhantes às expressões vocais produzidas por não-atores em situações reais, e diversos – quiçá a maioria dos – estudos ao longo dos anos utilizaram falas gravadas por atores como estímulo e objeto de análise acústica.

[...] *although acted emotional expressions may be deprived of certain characteristics pertaining to spontaneous expressions—such as subtle modifications of voice quality related to physiological modifications that actors might not succeed in producing deliberately—and may in some instances sound “theatrical”, they still represent close approximations of genuine emotional expressions.*<sup>26</sup> (BÄNZIGER; SCHERER, 2005, p. 257-258)

Durante o desenvolvimento desta tese de doutorado, surgida a oportunidade, ocorreram a ideia de verificar se o mesmo ocorreria não em atores, mas em músicos (instrumentistas e/ou vocalistas), já que estes profissionais também se utilizam desses padrões musicais de forma expressiva. Também consideramos que seria interessante verificar se os padrões percebidos na fala dos atores seriam os mesmos da fala dos músicos. Assim sendo, foram recrutados quatro atores (dois do sexo masculino, duas do sexo feminino) e três músicos (todos do sexo masculino).

As gravações foram realizadas remotamente pelos próprios artistas voluntários. Foi solicitado a eles que se concentrassem e gravassem em um local silencioso. Foram descartadas as gravações com ruídos no fundo e/ou chiados fortes, ou aquelas em que o voluntário tivesse encerrado a gravação antes do final da fonação. A pedido do pesquisador, dois dos voluntários aceitaram repetir pelo menos algumas das gravações em uma data posterior, a fim de verificar se o padrão acústico das gravações originais se repetiria. A repetição indicaria que o padrão entoacional seria um aspecto sistemático da fala. Esses fatores explicam a leve diferença no total de amostras para cada emoção. No total, foram coletadas 120 amostras bissilábicas: 30 para alegria, 29 para raiva, 31 para simpatia e 30 para tristeza.

Cabe, neste momento, um esclarecimento acerca das emoções básicas selecionadas. Alguns trabalhos, como Scherer et al. (1991) e mesmo Peres (2016), tomam alegria, raiva, tristeza e medo como emoções básicas. A opção de Curtis e Bharucha (2010) de substituir a emoção medo por simpatia (*pleasantness* no original, que também poderia ser traduzido como

---

<sup>26</sup> “Embora expressões emocionais atuadas possam não apresentar certas características das expressões espontâneas – como modificações sutis de qualidade de voz, relacionadas a modificações fisiológicas que os atores podem não conseguir produzir deliberadamente – e possam em algumas situações soar “teatrais”, elas ainda representam aproximações próximas de expressões emocionais genuínas.” (Tradução nossa.)

“agradabilidade”) segue por uma via distinta. Ao justificarem a escolha dessas quatro emoções, os autores alegam que

*These four emotions were selected because they map onto different dimensions of valence and arousal (which we hoped would maximize the acoustic differences between emotions) and because they are among the most common expressive states (which should lead to accurate encoding and decoding).<sup>27</sup>* (CURTIS; BHARUCHA, 2010, p. 336).

As dimensões a que os autores se referem são:

- Valência: agradável/positivo e não-agradável/negativo;
- Ativação: agitado e não agitado<sup>28</sup>.

A opção dos autores certamente poderia ser problematizada, e apenas esse aspecto do trabalho já poderia engendrar uma discussão mais aprofundada. No entanto, estabelecer exatamente quais são (ou se realmente existem) as emoções básicas não está entre os objetivos desta tese. Assim, considerando que a pesquisa de Curtis e Bharucha (2010) serviu como base para o trabalho de Mestrado do autor e continua sendo uma das maiores inspirações para esta pesquisa de Doutorado, optamos por aderir à seleção feita por esses autores.

As amostras de fala deveriam apresentar duas características principais, a princípio:

a-) serem bissilábicas, de modo a ser possível extrair uma altura melódica saliente de cada núcleo silábico, o que posteriormente possibilitaria estabelecer a relação entre essas duas alturas melódicas e verificar a qual intervalo musical essa relação corresponderia;

b-) não apresentarem, do ponto de vista semântico, qualquer indicação de estado emocional.

Com isso em mente, foram estabelecidas as quatro frases utilizadas no primeiro experimento, sendo elas: "Sou eu", "Tá bom", "Vem cá", "Eu sei".

Uma primeira sessão de gravação se mostrou infrutífera. Os atores relataram dificuldade em conceber uma situação em que enunciassem as frases indicadas ao mesmo tempo em que expressassem a emoção correspondente. A pedido dos próprios atores, mas refletindo uma prática utilizada em estudos anteriores (como em Banse e Scherer, 1996, e o próprio estudo de Curtis e Bharucha, de 2010), foram criados cenários fictícios, “cenas embrionárias” que facilitariam o ato de incorporar um personagem em um contexto que evocasse a emoção desejada e no qual o personagem enunciasse a frase-estímulo. As instruções e os cenários

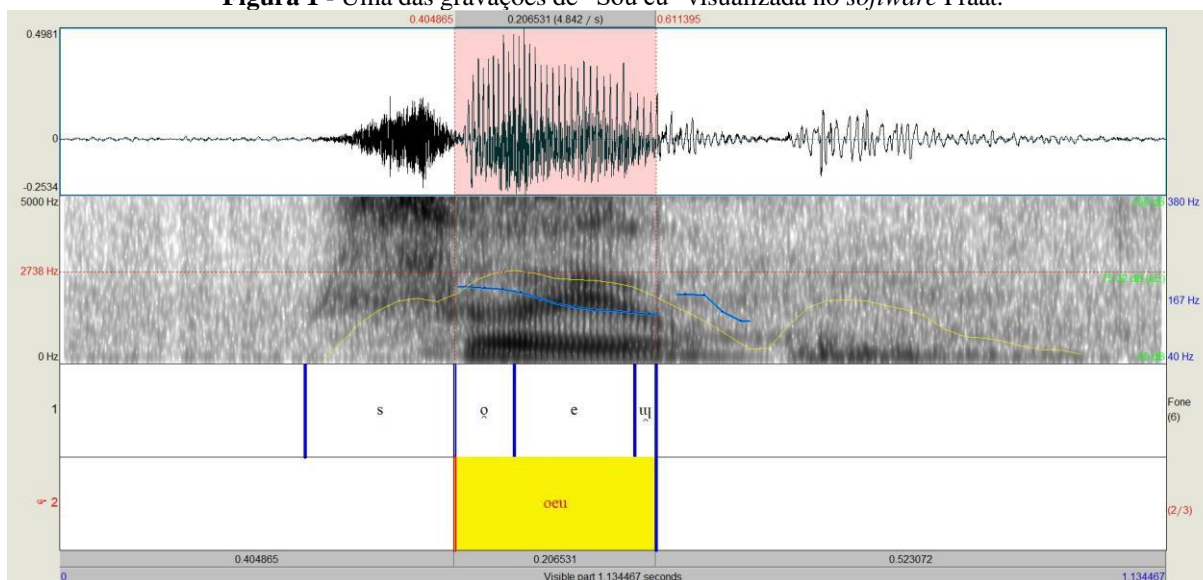
<sup>27</sup> “Essas quatro emoções foram selecionadas porque mapeiam diferentes dimensões de valência e excitação (o que, esperávamos, maximizaria as diferenças acústicas entre emoções) e porque estão entre os estados expressivos mais comuns (que devem levar a codificação e decodificação precisas).” (Tradução nossa)

<sup>28</sup> Alegria – agradável/agitado, Simpatia – agradável/não-agitado, Raiva – desagradável/agitado, Tristeza – desagradável/não-agitado.

fictícios entregues aos atores encontram-se no apêndice D. A criação de cenários se mostrou importante não apenas para facilitar a participação dos atores e músicos voluntários, mas também por uma questão metodológica. Pensando em cada falante como um codificador (*encoder*) e cada emoção como um rótulo (*label*), não se pode negligenciar a possibilidade de que cada codificador atribuísse uma nuance emocional diferente a cada rótulo, e, por conseguinte, pensasse em situações-gatilho diferentes para evocar as emoções. O uso de cenários pré-definidos junto aos rótulos minimiza esse risco (SCHERER; BANSE; WALLBOT, 2001).

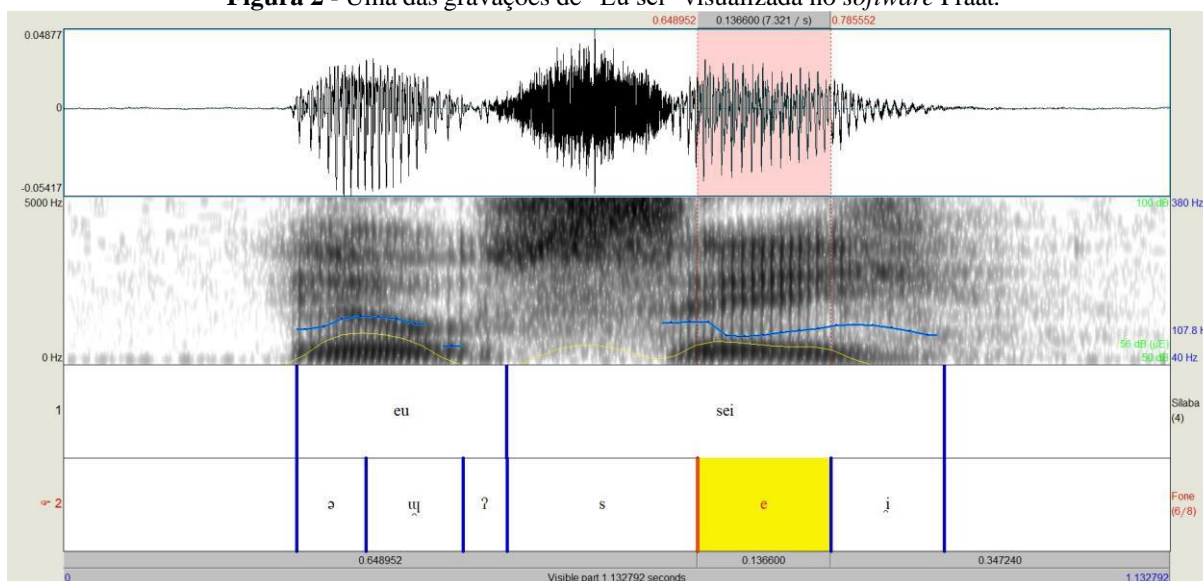
Também foi verificada a inadequação de uma das frases: na maioria esmagadora das gravações de “Sou eu”, foi verificada a ocorrência de apenas uma sílaba fonética, resultando em realizações como [su̯ eu̯] [sɔ eu̯] e [sɔ eũ̯]. Esse foi o caso da gravação cujo espectrograma pode ser visto na figura 1. Nesses casos, não foi possível verificar correspondência com intervalos musicais, pois a sequência de vogais em posição de rima e ataque silábico em “Sou eu” inviabilizou o estabelecimento de dois núcleos vocálicos distintos, ou seja, os atores enunciaram as duas sílabas fonológicas como apenas uma única sílaba fonética. Compare-se a realização de “Sou eu” na figura 1 com a realização de “Eu sei” mostrada na figura 2, na qual podemos observar claramente duas sílabas separadas por uma consoante desvozeada.

**Figura 1** - Uma das gravações de “Sou eu” visualizada no *software* Praat.



A intensidade do sinal sonoro é indicada pela linha amarela. É possível perceber que o pico silábico ocorre justamente no momento de transição entre [ɔ] e [e]. Ademais, há apenas um pico, seguido de um trecho em que a intensidade se mantém estável, antes de cair. Em azul, podemos verificar dois segmentos com  $f_0$  definida. O segundo trecho, porém, não consiste em fonação, mas em uma expiração de ar forçada pela atriz.

**Figura 2** - Uma das gravações de “Eu sei” visualizada no *software* Praat.



A linha amarela mostra a existência de mais de um pico de intensidade, e a existência de duas sílabas é corroborada por dois segmentos distintos da linha azul, que indica  $f_0$ . Neste caso, todo o sinal sonoro é proveniente da fonação do ator, não havendo ocorrência de som de expiração, por exemplo.

Optamos por não forçar uma separação artificial entre núcleos silábicos com base na percepção fonológica do pesquisador, descartando as gravações da frase “Sou eu”, e substituindo-a nas gravações posteriores por “Aqui”.

Para conduzir a análise das amostras de fala obtidas, optou-se, a princípio, por aderir ao procedimento do estudo original de Curtis e Bharucha (2010) e utilizar o modelo computacional Prosogram, proposto por Mertens (2004), o qual já foi utilizado por outros estudos que se debruçaram sobre os padrões de altura melódica observados na fala (PATEL, 2005; PATEL; IVERSEN; ROSENBERG, 2006). O Prosogram, além de um modelo de estilização de contorno de altura melódica, encontra-se disponível como um script em licença Creative Commons para o software Praat (BOERSMA; WEENINK, 2020), amplamente utilizado em pesquisas nas áreas de fonética e fonologia. O script fornece uma estimativa do contorno melódico percebido pelo ouvinte médio, computado como uma estilização dos dados de  $f_0$ .

O desenvolvimento do modelo do Prosogram se baseia na (possível) existência de diferenças entre a percepção do contorno da altura melódica pelos ouvintes e o que se observa em uma análise acústica padrão. Segundo o autor do modelo, quando a comunicação ocorre através da língua falada, há diferença no processamento do sinal realizado pelo sistema auditivo em comparação com o processamento digital, uma vez que os contornos da altura melódica são "processados" por ouvintes humanos e não por máquinas (MERTENS, 2004, p. 2). Mertens defende, por exemplo, que uma variação na frequência fundamental deve ter um tamanho mínimo, sendo que esse tamanho varia como uma função da frequência de início e a duração

do estímulo (MERTENS, 2004, p. 3). Nesse sentido, em última análise, o modelo de transcrição tem o objetivo de distinguir variações audíveis de não-audíveis, seja em sílabas individuais ou em sequências de sílabas – o que, tecnicamente, pode ocasionar uma diferença entre a entoação efetivamente produzida e a percepção da entoação, algo que não é necessariamente um problema, se considerarmos que na própria distinção entre fonética e fonologia está a diferença entre aquilo que é efetivamente produzido e a forma como o estímulo produzido é interpretado e categorizado pelos falantes. Outro aspecto importante destacado pelo autor é o de que a transcrição seja neutra, ou seja, independente de modelos teóricos, de modo a possibilitar seu uso por pesquisadores com diferentes referenciais teóricos. Em suma:

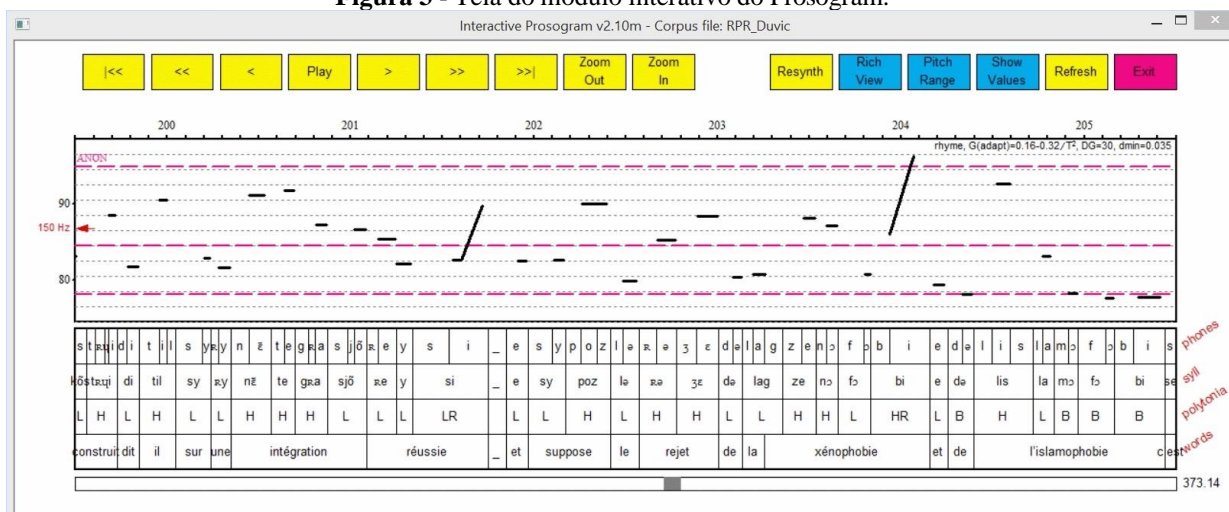
*The approach for prosody transcription described in this paper has the following properties. It's a stylization of the F0 curve of vowel nuclei, aiming at the reconstruction of the perceived pitch contour, based on a psycho-acoustic model of tonal perception. The transcription preserves temporal structure of the acoustic signal, includes annotations of text and phonetic transcription. The latter is used in the identification of vocalic nuclei. Pitch variations are shown on a pitch scale in semitones, for readability. The temporal alignment enables duration measurements of sounds and syllables, the localisation of pauses, and the study of speech rate and rhythm.*<sup>29</sup>  
(MERTENS, 2004, p. 4)

O processamento de segmentos de áudio através do script gera perfis prosódicos expressos em uma série de gráficos com medidas em Herz e Semitons. O script também gera um gráfico mostrando o contorno prosódico estilizado, como pode ser observado na figura 3, em que os traços em preto representam um núcleo silábico cada, sendo que a sequência de traços ao longo do eixo horizontal forma o contorno da altura melódica do segmento analisado. No caso dos traços diagonais, o núcleo silábico é constituído por um glissando e o script não julga possível estabelecer uma f0 única. Na interface interativa do programa, o usuário pode navegar pelo sinal de fala e sua estilização, executar o áudio e resintetizar o sinal com a altura melódica estilizada.

---

<sup>29</sup> A abordagem de transcrição prosódica descrita neste trabalho tem as seguintes propriedades. É uma estilização da curva da f0 de núcleos vocálicos, com o objetivo de reconstruir o contorno de altura melódica percebido, baseado em um modelo psicoacústico de percepção tonal. A transcrição preserva a estrutura temporal do sinal acústico, inclui anotações de texto e transcrição fonética. Esta última é baseada na identificação dos núcleos vocálicos. Variações de altura melódica são mostradas em uma escala de altura melódica em semitons, para fins de legibilidade. O alinhamento temporal permite a medida da duração de sons e sílabas, a localização de pausas, e o estudo da taxa de elocução e ritmo. (Tradução nossa).

**Figura 3 - Tela do módulo interativo do Prosogram.**



Fonte: <https://sites.google.com/site/prosogram/home>, acesso em 29 de maio de 2019.

Cabe mencionar que um dos parâmetros de análise do *script* é o "limite de glissando", que determina se o núcleo vocálico seria percebido por ouvintes como tendo uma altura melódica estável ou se seria percebido como um *glide* (caso em que a altura melódica sobe ou desce ao longo da duração do núcleo). Esse limite pode ser alterado pelo usuário, mas existe um valor padrão (indicado pela fórmula  $G = 0.32/T^2$ ) que, segundo o autor do *script*, apresenta um alto grau de correspondência em relação às avaliações e transcrições feitas manualmente. Na fórmula,  $G$  indica o *glissando threshold*, o limite matemático para que a variação de frequência fundamental seja audível, e  $T$  indica a duração da variação. Esta fórmula, apresentada por Mertens (2004), remonta a trabalhos anteriores (HART, 1976, D'ALESSANDRO; ROSSET; ROSSI, 1998). Caso o núcleo silábico seja interpretado como tendo uma altura melódica estável<sup>30</sup>, sua estilização será materializada em um traço horizontal, enquanto um núcleo silábico constituído por *glide* será representado com uma linha diagonal mais ou menos inclinada, a depender da duração do núcleo vocálico.

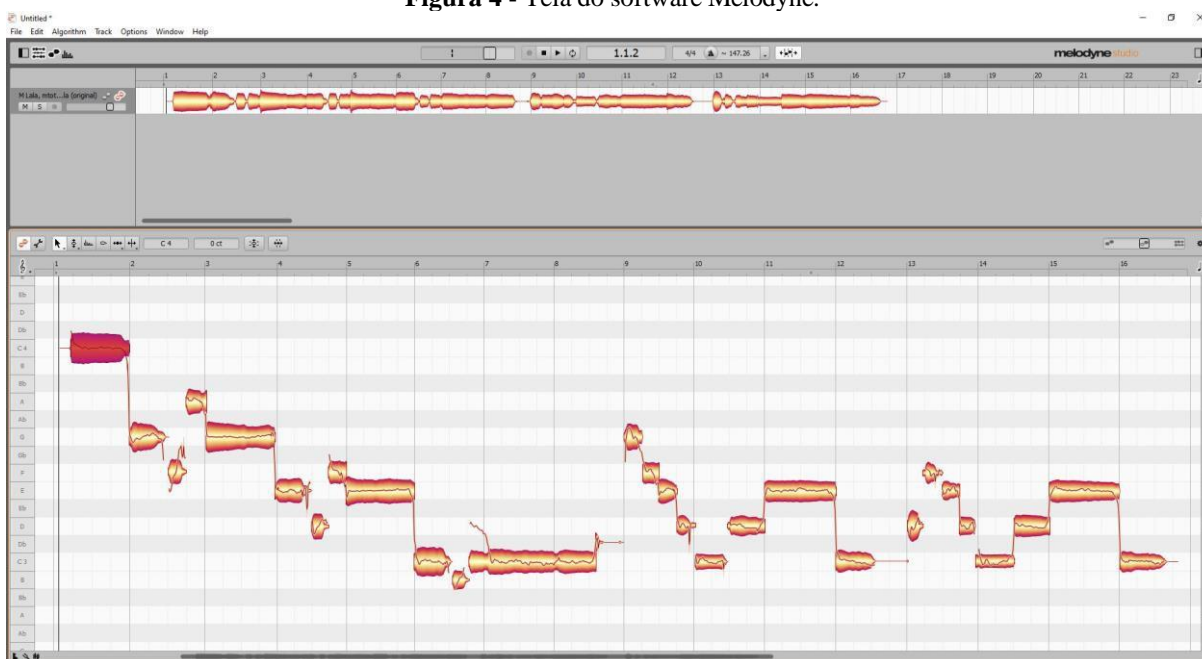
A intenção da estilização da altura melódica das amostras bissilábicas é chegar a duas alturas melódicas distintas por amostra, cada altura correspondendo a um núcleo vocálico. Além da unidade Hertz, a plotagem gráfica na versão atual do *script* Prosogram inclui as medidas em semitons nos gráficos e diagramas, de modo que se torna bastante fácil medir a distância entre

<sup>30</sup> Cantores, pessoas com ouvido absoluto e músicos em geral certamente são muito mais sensíveis às variações de  $f_0$  na música do que leigos. No entanto, o modelo de estilização ao qual nos referimos aqui tem como objetivo primeiro a análise da *fala*, e não da música (apesar de lançar mão da metalinguagem e do arcabouço teórico musical). Assim sendo, a palavra *estável* aqui não deve ser entendida no mesmo sentido usado para se referir a uma nota musical estável (sem vibrato ou outras oscilações), mas no sentido de uma frequência fundamental que se mantém estável por tempo o suficiente para ser saliente no processamento da fala.

as frequências de  $f_0$  (ou, já incorporando a analogia com a música, da “nota musical” do núcleo vocálico) dos núcleos vocálicos. Em outras palavras, se a cada núcleo vocálico corresponde uma nota musical, através do processo acima descrito é possível verificar qual o intervalo musical correspondente a cada amostra.

Uma segunda ferramenta para verificar o intervalo formado entre as alturas dos dois picos silábicos, que também foi utilizada neste trabalho, foi o *software* Melodyne, da empresa Celemony. O Melodyne é um editor de áudio voltado para produção musical, frequentemente usado para corrigir a afinação de gravações vocais. Ao importar uma gravação de áudio para o programa, os algoritmos reconhecem as variações da frequência fundamental e dos harmônicos do sinal sonoro, e as ondas sonoras são exibidas em uma interface dividida em notas musicais. A figura 4 mostra a tela principal do programa quando um arquivo de áudio é aberto. As manchas amarelas com bordas na cor magenta representam o sinal sonoro como um todo, enquanto as linhas vermelhas representam a frequência fundamental do som. Como se pode imaginar, ao longo do eixo Y, à esquerda, há uma divisão em linhas, cada linha representando uma nota musical distinta. As manchas, bem como as linhas vermelhas, estão alinhadas à nota musical correspondente ao sinal sonoro identificado pelo *software*. Assim, é possível verificar qual é a nota musical correspondente à frequência fundamental em determinado momento da fonação, bem como as flutuações nessa frequência fundamental.

**Figura 4 -** Tela do software Melodyne.



O Melodyne foi lançado comercialmente em 2009. Assim, na época da coleta de dados do estudo de Curtis e Bharucha (2010), o software talvez ainda não estivesse disponível e, se estivesse, possivelmente não teria sido considerado como uma opção de ferramenta de análise. Ao longo dos anos, o software foi incorporado ao processo de registro de estúdio dos mais variados artistas, como o pianista norte-americano Herbie Hancock, a cantora e compositora islandesa Björk, a banda britânica Coldplay, o cantor e compositor inglês Peter Gabriel e o compositor de trilhas sonoras norte-americano Thomas Newman. O software pode ser – e foi – utilizado não apenas para a correção de afinação de vocais e instrumentos, mas como ferramenta de criação de arranjos (WILSON, 2017). A identificação das notas correspondentes às amostras bissilábicas também é muito mais rápida e fácil em comparação com o script Prosogram. Por isso, após algumas análises, verificamos a possibilidade de utilizar o Melodyne como substituto do Prosogram.

Curtis e Bharucha disponibilizaram uma pequena seleção das amostras de fala utilizadas no estudo em um site pessoal cujo endereço está indicado no artigo de 2010. Originalmente hospedado no servidor da Tufts University, o site foi acessado pelo pesquisador, que colheu quatro das (menos de dez) amostras disponibilizadas e que haviam sido analisadas pelos autores. A seguir, foi feita a verificação acerca da interpretação do sinal acústico pelo script Prosogram e pelo software Melodyne. Ambos acusaram o mesmo intervalo correspondente a uma terça menor (as amostras de fala em questão eram gravações das atrizes enunciando a frase “Come here” expressando tristeza). Depois de nos certificarmos que as notas atribuídas pelo Melodyne seriam correspondentes aos intervalos acusados pelo script Prosogram, optou-se por utilizar apenas o software Melodyne nas análises subsequentes, de modo a agilizar o processo. O site indicado pelos pesquisadores no artigo publicado, infelizmente, já não está no ar. Duas tentativas posteriores de acessar a página (em outubro de 2020 e fevereiro de 2022) se mostraram infrutíferas, e o endereço agora redireciona para a página inicial do Departamento de Psicologia da Tufts. Já não encontramos, pelo sistema de busca no site do departamento, qualquer menção a Curtis, Bharucha ou à pesquisa de 2010. Porém, como evidência perene de que as amostras de fato estavam disponíveis *online*, remetemos a Day-O’Connel (2013), que também discute as amostras utilizadas no estudo de 2010, e também ao relato de Subramanian (2010), que reporta os resultados do trabalho no portal de notícias Mint..

Comparações entre a verificação do Prosogram e do *software* Melodyne com uma seleção de 5 amostras do nosso próprio *corpus* nos deram confiança de que os resultados seriam os mesmos, de modo que para as verificações subsequentes utilizamos apenas o Melodyne.

Em suma, além do *script* Prosogram, a análise das amostras foi complementada verificando cada uma das amostras bissilábicas com o Melodyne, verificando se as variações de frequência formavam intervalos musicais definidos. Anotados os valores de referência, os intervalos foram verificados através das ferramentas de cálculo de intervalos musicais dos sites Music Theory (INTERVAL..., 2022) e Omni Calculator (RAIN, 2021).

## 2.4 Resultados

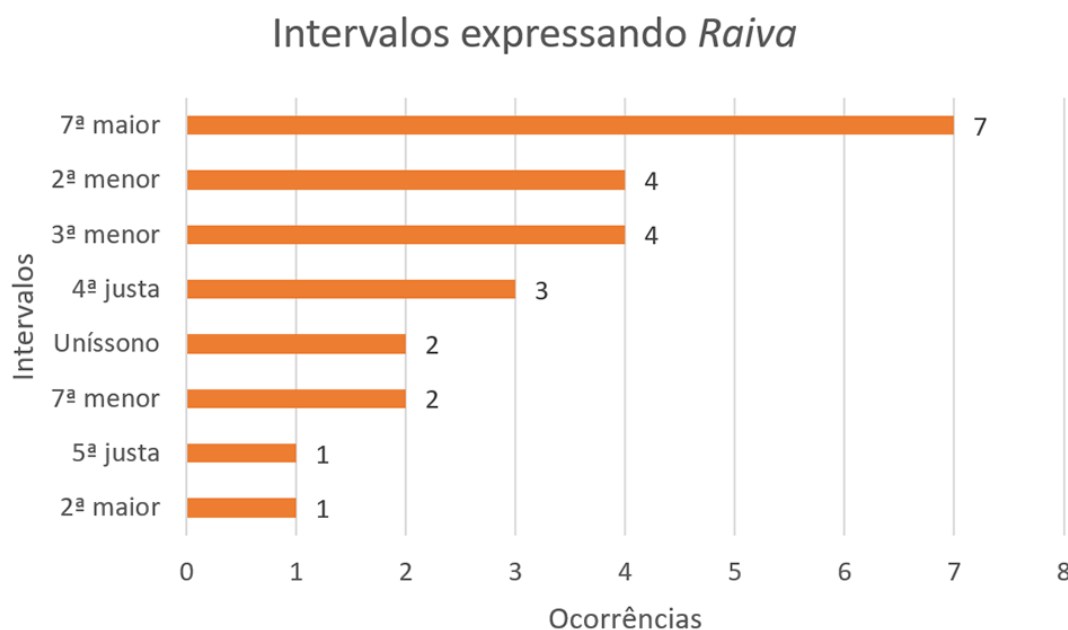
Os resultados obtidos no primeiro experimento são bastante diferentes dos reportados por Curtis e Bharucha (2010). Começando pelo intervalo mais comum nas amostras expressando tristeza, o grande chamariz do estudo e a origem do título do trabalho publicado, Curtis e Bharucha reportaram um pico bastante acentuado de ocorrências correspondentes ao intervalo de terça menor descendente. Ou seja, a grande maioria das amostras bissilábicas expressando tristeza correspondia ao intervalo de terça menor. O mesmo não foi verificado em nosso experimento. O intervalo mais comum entre os falantes que participaram de nosso experimento foi, com uma larga margem, o de quinta justa (todas as ocorrências com padrão descendente), com mais do que o dobro de ocorrências da quarta justa, o segundo intervalo mais comum quando os falantes expressaram tristeza (também descendente em todas as ocorrências observadas).

**Gráfico 1** - Intervalos observados na fala de atores e músicos expressando tristeza.



A segunda emoção para a qual os autores do estudo de 2010 verificaram picos de ocorrência de intervalos específicos foi a raiva. Os intervalos característicos dessa emoção foram, no estudo original, o de segunda menor ascendente e o de quinta justa ascendente. Nossos resultados não foram os mesmos. Apesar de também verificarmos que a segunda menor ascendente foi (empatado com o de terça menor) o segundo intervalo mais utilizado pelos falantes nas amostras de raiva analisadas, a diferença entre as ocorrências não foi tão acentuada. Além disso, duas das quatro ocorrências do intervalo de segunda menor foram ascendentes, e não descendentes – logo, essas duas amostras vão na direção contrária à do estudo de Curtis e Bharucha (2010). O intervalo observado com maior frequência nas amostras de raiva foi o de sétima maior (descendente em todas as ocorrências).

**Gráfico 2** - Intervalos observados na fala de atores músicos expressando raiva.



As emoções agradabilidade (*pleasantness*) e alegria, no estudo de 2010, não retornaram picos de ocorrência. Em outras palavras, não foram intervalos que emergissem com frequência nas amostras bissilábicas expressando essas duas emoções. Tal observação talvez corresponda, no que se refere aos nossos resultados, ao fato de que a quantidade de intervalos utilizada para expressar simpatia e alegria ter sido um pouco maior do que a quantidade utilizada para expressar raiva (oito intervalos usados para raiva, nove para tristeza, dez para simpatia, onze para alegria, . Ademais, a distribuição de intervalos foi menos desigual: a diferença entre a frequência do intervalo mais usado e dos intervalos menos usados não foi tão acentuada quanto, por exemplo, no caso da tristeza.

Para simpatia, os intervalos de sétima menor descendente e de quarta aumentada (descendente em quatro das cinco ocorrências) foram os mais utilizados. A presença do intervalo de quarta aumentada na expressão de uma emoção de valência positiva é surpreendente, dado o caráter bastante dissonante do intervalo. Já para alegria, a distribuição dos intervalos foi ainda menos desigual, com o intervalo de segunda menor ascendente sendo o mais comum na amostra observada.

**Gráfico 3** - Intervalos observados na fala de atores expressando simpatia.



Para todas as emoções, diferenças entre sexo, educação e envolvimento com música (atores que também são músicos em comparação com atores que não são músicos) foram negligenciáveis – isto é, não se mostraram relevantes na distribuição dos intervalos.

**Gráfico 4 - Intervalos observados na fala de atores expressando alegria.**



Considerando os resultados como um todo, parece-nos que as únicas associações que podem ser inferidas são o intervalo de quinta justa na expressão de tristeza e de oitava diminuta na expressão de raiva. E, mesmo nesses dois casos, há de se reconhecer que as evidências não são particularmente robustas, considerando o número de amostras totais das duas emoções (26 para tristeza, 24 para raiva). Portanto, é necessário que novas pesquisas busquem replicar os resultados, tanto para português brasileiro quanto para o inglês, a fim de acumular mais evidências acerca do tema, seja para corroborar algumas associações (que, aparentemente, podem ser específicas para cada comunidade linguística) ou para descartá-las.

## 2.5 Discussão

Os resultados reportados por Curtis e Bharucha (2010), no que se refere à emergência de intervalos de 3ª menor na fala expressando tristeza, não puderam ser replicados. Ao contrário, o intervalo mais utilizado pelos falantes para expressar tristeza foi a quinta justa.

Tampouco os outros intervalos, utilizados para expressar outras emoções, corroboraram os resultados do estudo de 2010. Isso não significa, necessariamente, que os resultados dos dois pesquisadores norte-americanos devam ser descartados de imediato. É necessário que mais pesquisadores tentem replicar o experimento, principalmente com línguas que não sejam o inglês, para verificar se o negativo aqui apresentado foge ao padrão, se a recorrência de um intervalo pode ser observada apenas em algumas línguas, ou se há outras correlações que possam ser verificadas.

As discordâncias entre os resultados do experimento desta tese de doutorado e do estudo de 2010 ensejaram uma volta à literatura e uma reavaliação, tanto do estudo original, quanto de sua repercussão. Os fatos observados levam a uma discussão pertinente à prática científica, discussão essa que desenvolveremos brevemente a seguir.

Uma consulta ao sistema de busca de literatura acadêmica Google Scholar, realizada em março de 2022, apontou 133 artigos indexados citando o trabalho de Curtis e Bharucha até o momento. Evidentemente, verificar cada um dos textos que citam o trabalho original é uma tarefa inviável para um único pesquisador. Ainda assim, selecionamos uma amostra de 15 artigos (correspondentes a mais de 10% do número de citações) e verificamos de que maneira, exatamente, os resultados de Curtis e Bharucha eram incorporados. Foram eles: Bowling *et al* (2012), Cespedes-Guevara e Eerola (2018), Day-O’Connell (2013), Escoffier *et al* (2013), Gosselin, Paquette e Peretz (2015), Lima e Castro (2011); Ma e Thompson (2015); Paquette *et al* (2018), Park *et al* (2015), Piazza *et al* (2013), Quinto, Thompson e Keating (2013), Stevens (2012), Wöllner (2012), Mantell e Pfordresher (2013), Mathur *et al* (2015) e Slevc (2012). A maioria dos trabalhos elencados incorporavam os resultados de Curtis e Bharucha apenas como evidência e suporte para afirmar vários vínculos possíveis entre música e língua. Três deles, Ma e Thompson (2015), Paquette *et al* (2018) e Slevc (2012) sequer citaram o artigo textualmente. Porém, em três casos, verificamos discussões sobre o estudo que são pertinentes para as considerações acerca do nosso primeiro experimento.

Cespedes-Guevara e Eerola (2018) afirmam que os resultados de Curtis e Bharucha (2010) teriam sido replicados por Bowling *et al.* (2012). Apontam, porém, que o resultado teria sido replicado apenas com um grupo de falantes de inglês, e que mais replicações com amostras maiores de línguas eram necessárias antes de aceitar a hipótese do trabalho de 2010. A leitura do texto integral de Bowling *et al.*, no entanto, revela uma história consideravelmente diferente. A correspondência encontrada consiste no fato de que tanto melodias ocidentais, quanto melodias clássicas do sul da Índia, quanto amostras de fala de pessoas anglófonas, tendem a utilizar *intervalos menores* na expressão de emoções negativas e quietas. Falar em intervalos

menores é muito diferente de destacar um único intervalo, a terça menor. Ora, o que temos são resultados que coincidem em partes, mas não resultados replicados. Desta forma, não houve, até onde pudemos verificar, nenhuma instância em que os resultados de Curtis e Bharucha (2010) tenham sido replicados.

O trabalho experimental de Day-O'Connell (2013), aliás, vai na direção oposta. O escrutínio do que o autor chama de *stylized interjections* o leva a concluir que o intervalo de terça menor é um indicativo de *playfulness* (jocosidade) por parte do falante. No que parece ser uma tentativa um tanto quanto forçada de conciliar dois resultados discrepantes, Day-O'Connell faz a ressalva de que fatores metodológicos tornam temerária a comparação dos resultados de seu trabalho com o estudo de Curtis e Bharucha (2010). Porém, em seguida, pondera que enquanto *tristeza* seria uma emoção genuína, *playfulness* se referiria a um modo de enunciar, um “aspecto perlocucionário do ato de enunciação” (DAY-O'CONNELL, 2013, p. 457). O autor postula ainda que, apesar da aparente incongruidade entre os resultados, é plausível que a terça menor seja um aspecto essencial tanto da expressão de tristeza quando de jocosidade, pois ambos os conceitos estariam unidos em sua atitude “não-agressiva” (DAY-O'CONNELL, 2013, p. 457). Em outras palavras, como tanto tristeza quanto jocosidade expressam estados de espírito e/ou atitudes não-agressivas, o intervalo de terça menor seria comum a essas duas expressões, com outros fatores, como a altura melódica média e a taxa de elocução, contribuindo para diferenciar os dois estados de espírito.

O que se pode observar, enfim, é que os resultados apresentados por Curtis e Bharucha (2010), apesar de constantemente citados na literatura interdisciplinar, não foram devidamente corroborados por tentativas de replicação, nem entre falantes de inglês, nem entre falantes de outras línguas. O primeiro experimento desta tese de doutorado, na tentativa de suprir essa lacuna, tentou replicar os resultados dos dois pesquisadores com falantes de Português Brasileiro, porém, sem sucesso. Por um lado, a impossibilidade de corroborar os resultados de um artigo publicado há 12 anos consiste quase em um convite para novas empreitadas futuras, uma vez que as evidências da hipótese acerca da correspondência do uso de terça menor na fala e na música agora são controversas. Por outro lado, julgamos alarmante a falta de tentativas de replicar os resultados, ao longo de mais de uma década, ao mesmo tempo em que o trabalho em questão continua a ser citado em pesquisas da área.

O título do artigo de Curtis e Bharucha (2010) vende uma ideia sem dúvida chamativa e atraente. A correspondência exata de um intervalo musical específico sendo usado na fala, precisamente por sua inusitada exatidão, pode parecer muito mais impactante do que uma correspondência entre intensidade ou frequência fundamental média, por exemplo. E é

justamente por isso, também, que é necessário colocar uma hipótese como essa à prova, repetidamente, nos mais diversos contextos possíveis, se o intuito for o de chegar a generalizações como “os intervalos da fala refletem os intervalos da música”. A lacuna, que nos esforçamos para preencher, ainda que parcialmente, é preocupante. Quais outras afirmações e frases de efeito baseadas em estudos não-replicados estão sendo repetidas *ad nauseam* na literatura?

Há de se manter os olhos atentos para reconhecê-las quando aparecerem, e colocá-las à prova antes de espalhá-las.

## **2.5 Considerações finais**

No primeiro experimento relatado nesta tese, buscamos verificar se a expressão de emoções na fala apresentaria algum tipo de correspondência musical. Uma eventual correspondência poderia ocorrer na forma de um mesmo padrão acústico sendo utilizado na fala e na música para expressar uma mesma emoção, ou no emprego recorrente de um padrão musical específico na expressão de um conteúdo emocional específico. Como vimos, nenhuma correspondência foi encontrada, contrariando nossas hipóteses iniciais e indo na direção contrária dos estudos que nos serviram de inspiração – estudos estes, porém, que também apresentam fragilidades, como pudemos perceber no decorrer de nossa tentativa de corroborá-los.

Passemos, agora, para o próximo experimento.

### 3. EXPERIMENTO 2: QUAL EMOÇÃO ESSA MÚSICA EXPRESSA?

#### 3.1 Introdução

Na seção 2.2, vimos que os intervalos de terça menor costumam ser associados à tristeza. O segundo experimento deste trabalho utilizou um *corpus* musical para verificar se, mantendo todos os demais elementos intactos, a maior ou menor presença de intervalos de terça menor pode acarretar mudanças categóricas na percepção de emoções em um enunciado cantado em uma língua desconhecida.

Nossa hipótese é a de que, se intervalos de terça menor são associados à tristeza, a maior presença destes intervalos acarretará maior percepção de expressão de tristeza, e que uma menor quantidade de intervalos de terça menor fará com que uma melodia tenda a ser percebida como menos triste.

#### 3.2 Conceitos prévios

Já abordamos, na seção 2.1, as noções de altura melódica (ou *pitch*) e de intervalos, bem como a associação dos intervalos de terça menor à tristeza, na seção 2.2. Antes de passarmos para o próximo experimento desta pesquisa, é necessário visitar os conceitos de *escala*, *modo* e de *tonalidade*.

De acordo com Drabkin (2001c), uma escala pode ser definida, no nível mais básico, como uma sequência de notas em ordem ascendente ou descendente de alturas melódicas. Porém, enquanto conceito musicológico, uma escala é uma sequência longa o suficiente para definir, de maneira não-ambígua, um modo, tonalidade, ou alguma forma particular de construção linear, começando e terminando na nota fundamental do modo ou da escala. Metaforicamente, pode-se pensar na escala como a paleta de cores selecionada por um pintor para servirem de base para a pintura: as cores (notas musicais), organizadas de determinada maneira em uma determinada dimensão (espacial no caso da pintura, temporal no caso da música), formam uma obra cuja estrutura emerge das proporções e relações desses elementos constituintes.

O uso do conceito *modo* variou ao longo da história, mas sempre foi usado para designar classes de melodias, sendo que do século XX em diante, vem sendo usado para designar certos tipos de normas ou modelos para composição ou improvisação (POWERS et al., 2001). De acordo com os autores, uma definição resumida que inclui tanto a definição histórica ocidental

e os aspectos mais recentes propostos por estudiosos de música asiática e do oriente médio pode ser encontrada em Winnington-Ingram (1936):

*Mode is essentially a question of the internal relationships of notes within a scale, especially of the predominance of one of them over the others as a tonic, its predominance being established in any or all of a number of ways: e.g., frequent recurrence, its appearance in a prominent position as the first note or the last, the delaying of its expected occurrence by some kind of embellishment.*<sup>31</sup> (WINNINGTON-INGRAM, 1936, p.2, apud POWERS et al., 2001)

Finalmente, temos a noção de tonalidade. De acordo com Hyer (2001), na música tonal, o tom se refere ao arranjo abstrato de fenômenos como melodias, harmonias e cadências ao redor de um ponto de referência, ou nota tônica<sup>32</sup>. Em uma outra definição, Cooke (1959, p. 40) afirma que a tonalidade é a base expressiva da linguagem musical da Europa ocidental consiste no sistema intrincado de relações tensionais entre notas, o qual se pode chamar de sistema tonal. Há uma relação íntima entre a noção de tonalidade e a de escala: o tom se refere ao tipo de estrutura de escala implicado por uma melodia: melodias tonais são aquelas em que as alturas melódicas implicam um tom específico, enquanto para melodias atonais tal implicação não existe (TAN, PFORDRESHER, HARRÉ, 2010, p. 76). Também de acordo com Hyer (2001), a nota tônica, que dá nome ao tom, unifica e coordena os fenômenos musicais dentro de seus domínios: uma composição no tom de Dó maior, por exemplo, possuiria uma essência de "Dó". Notas "dentro" de um tom soam como se "pertencessem" a uma melodia composta nesse tom, enquanto notas que não pertencem a essa escala, se usadas em uma melodia, são altamente salientes e podem criar momentos de alta tensão (TAN; PFORDRESHER; HARRÉ, 2010, p. 76)

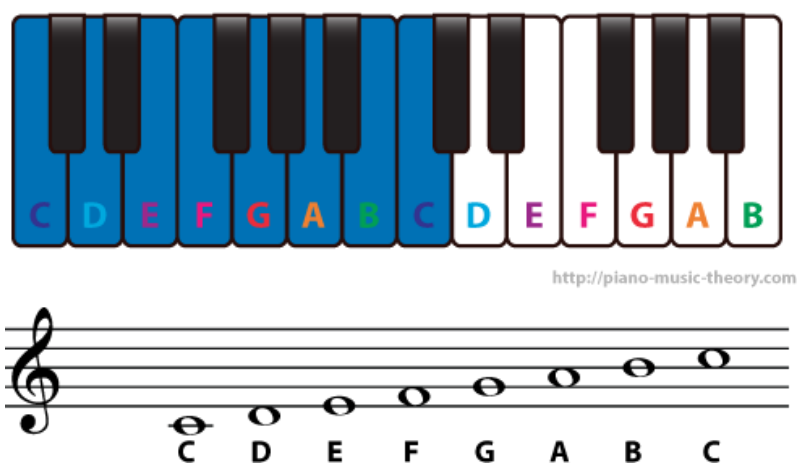
Ainda no que se refere à noção de tonalidade, também é considerada crucial a existência de dois gêneros modais básicos, maior e menor, cada um com características musicais diferentes oriundas, basicamente, da disposição de tons e semitons em suas respectivas escalas (HYER, 2001). Esses dois modos são dois polos "gêmeos" de expressão, dois sistemas interdependentes, baseados em tríades maiores e menores (COOKE, 1959, p 50). Uma das diferenças entre uma escala maior (por exemplo, Dó maior, figura 5) e uma escala menor natural (por exemplo, Dó

<sup>31</sup> Em tradução livre: "O modo é, essencialmente, uma questão das relações internas das notas dentro de uma escala, especialmente da predominância de uma delas sobre as outras enquanto nota tônica, sendo sua predominância estabelecida de uma de todas dentre várias maneiras: por exemplo, recorrência frequente, sua presença em posição de destaque como a primeira ou última nota, o adiamento de sua ocorrência esperada por algum tipo de embelezamento."

<sup>32</sup> Segundo o mesmo autor, o termo em inglês, *key*, tem uma dimensão metafórica, em que a chave (*key*) de uma música "destranca" ou clarifica o arranjo de relações entre alturas melódicas subjacentes à peça. A noção de uma peça ou passagem estar em um determinado tom envolve uma conceituação cultural, segundo a qual o tom envolve certas tendências melódicas e relações harmônicas que mantêm a nota tônica como centro de atenção (HYER, 2001).

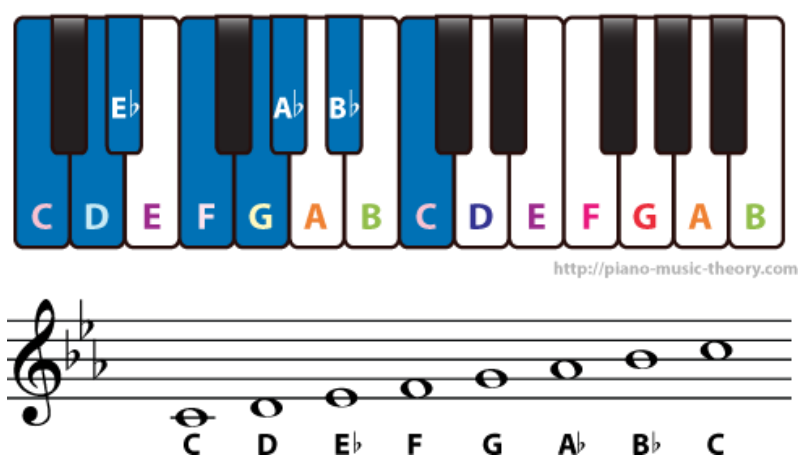
menor, figura 6) está justamente na terceira nota (terceiro grau) da escala. Entre o primeiro grau/nota e o terceiro grau/nota de uma escala menor, que por sua vez evoca um modo/tonalidade menor, há uma distância de três semitons, correspondente a um intervalo de terça menor, enquanto em uma escala maior, a distância entre o primeiro e o terceiro graus é de quatro semitons ou dois tons, correspondendo a um intervalo de terça maior<sup>33</sup>.

**Figura 5** - Escala de Dó maior, representada em teclas de piano e em partitura.  
**C Major Scale**



Fonte: Piano Music Theory<sup>34</sup>.

**Figura 6** - Escala de Dó menor, representada em teclas de piano e em partitura.  
**C Minor Scale**



Fonte: Piano Music Theory<sup>35</sup>.

<sup>33</sup> Ver seção 3.1.

<sup>34</sup> Disponível em < <https://piano-music-theory.com/2016/06/02/c-natural-minor-scale/>>, acesso em 17 de junho de 2021.

<sup>35</sup> Disponível em < <https://piano-music-theory.com/2016/05/31/c-major-scale/>>, acesso em 17 de junho de 2021.

As figuras 5 e 6 ilustram a diferença entre as escalas de Dó Maior e de Dó Menor. Nelas, as sete notas musicais (Dó, Ré, Mi, Fá, Sol, Lá e Si) aparecem anotadas como letras (C, D, E, F, G, A e B, respectivamente), e as notas pertencentes a cada escala respectiva aparecem coloridas de azul na representação do piano. A distância entre uma tecla de piano e a tecla adjacente, seja ela preta ou branca, é de um semitom. Como se pode ver, na escala de Dó Maior, chamada em inglês de *C Major Scale*, a distância entre a primeira nota da escala, Dó/C, e a terceira nota pertencente à escala, Mi/E, é de 2 tons ou quatro semitons<sup>36</sup> (um semitom entre C e a tecla preta adjacente correspondente a C#, um semitom entre C# e D, um semitom entre D e a tecla preta correspondente, nesse contexto, a D#, e mais um semitom entre D# e E). Temos, assim, um intervalo de terça maior. Já na escala de Dó Menor (*C minor*), a distância entre a primeira nota da escala, Dó/C, e a terceira nota da escala, Mi<sup>b</sup>/E<sup>b</sup>, é de 1 tom e meio ou três semitons (um semitom entre C e a tecla preta adjacente correspondente a C#, um semitom entre C# e D, e um semitom entre D e a tecla preta correspondente a E<sup>b</sup>).

Uma vez esclarecido o que entendemos aqui por “maior” e “menor”, podemos nos voltar para uma noção bastante difundida acerca desses dois conceitos:

*It is commonplace (though a much-disputed one) that the positive emotions (joy, confidence, love, serenity, triumph, etc.) are expressed by major music, and the negative emotions (sorrow, fear, hate, disquiet, despair, etc.) by minor music. Using Freud's basic categories of human emotion - pleasure and pain - we may say that there is a tendency to equate the major system with pleasure, the minor system with pain.*<sup>37</sup> (COOKE, 1959, p. 50-51)

A passagem acima, escrita pelo músico e musicologista britânico Deryck Cooke, postula de forma sucinta que, em música, a noção de “menor” estaria associada a afetos negativos e melancólicos, enquanto a noção de “maior” estaria associada a estados de espírito mais positivos e alegres.

Tal associação se encontra expressa de forma elegante na letra do standard de jazz *Ev'ry Time We Say Goodbye*, escrita pelo compositor norte-americano Cole Porter. Nela, a mudança no estado de espírito ocasionada pela despedida da pessoa amada é retratada metafórica e

<sup>36</sup> Um tom equivale a dois semitons.

<sup>37</sup> Em tradução livre: “É lugar comum (embora muito contestado) que as emoções positivas (alegria, confiança, amor, serenidade, triunfo, etc.) são expressas pela música maior, e as emoções negativas (pesar, medo, ódio, inquietação, desespero, etc.) pela música em tom menor. Usando as categorias básicas de Freud da emoção humana - prazer e dor - podemos dizer que há uma tendência de equiparar o sistema maior ao prazer, o sistema menor à dor.”

literalmente com uma mudança “de maior para menor”<sup>38</sup>. Essa característica foi muito usada no Barroco pela Teoria dos Afetos e propagou-se na música erudita.

*For most Western listeners, the minor mode suggests a negative emotional tone while the major mode has a more positive connotation, other things equal. Whether the different connotations of the two modes result from acoustic, spectral differences in the major and minor triads [...], or whether they are culturally transmitted, is an unresolved question [...]*<sup>39</sup>. (KASTNER, CROWDER, 1990, p.189, destaques nossos)

Uma explicação é que a terça menor, sendo mais baixa do que a terça maior, tem um som "deprimido", e o fato de não ser parte da série harmônica base a faz uma "depressão não-natural" no estado "naturalmente feliz" das coisas, de acordo com ideias ocidentais (COOKE, 1959, p. 57).

É interessante notar, entretanto, que essa associação não é um consenso. De fato, por mais disseminada que a associação entre “maior-menor” e “alegria-tristeza” seja, ela nem sempre se encontra “validada” em manuais e dicionários de música. De fato, os verbetes respectivos a “Maior”, “Menor” e “Terça” (DRABKIN, 2001a, 2001b, 2001d) no dicionário de música publicado pela universidade de Oxford não apresentam qualquer menção a essa suposta melancolia característica dos intervalos, escalas e modos menores - há apenas a limitação técnica e quase matemática. No verbete relativo a “Menor”, por exemplo, lemos que menor é o intervalo entre quaisquer duas notas com uma distância de dois graus da escala diatônica entre si, por exemplo, Dó-Mi (DRABKIN, 2001c).

Há também aqueles que, quando não se opõem totalmente à associação, mostram-se (no mínimo) céticos em relação às generalizações. Em um texto para o site *The Atlantic*, comentando o estudo<sup>40</sup> de Meagan E. Curtis e Jamshed J Bharucha (2010), o jornalista Daniel Wattenberg (2010) lança as perguntas retóricas: “*Where does sad music get its sadness from? And whom should you ask — a composer or a cognitive psychologist?*”<sup>41</sup> Em um texto que disfarça pouco apreço pela pesquisa em questão, Wattenberg (2010) admite que pode existir uma correlação vaga, reforçada pela nossa tradição musical, entre escalas menores e "tristeza", mas que é um erro pensar que os humores evocados pela música possam ser reduzidos à

<sup>38</sup> Como diz a letra da canção, “[...] but how strange the change from major to minor every time we say goodbye”.

<sup>39</sup> Em tradução livre: “Para a maioria dos ouvintes ocidentais, o modo menor sugere um tom emocional negativo, enquanto o modo maior tem uma conotação mais positiva, se outras coisas são mantidas iguais. Se as diferentes conotações dos dois modos resultam de diferenças acústicas, espectrais nas tríades maiores e menores [...], ou se são transmitidas culturalmente, é uma questão em aberto.”

<sup>40</sup> Mais precisamente, comentando o texto do blog da revista *Scientific American* sobre o estudo.

<sup>41</sup> Em tradução livre: De onde vem a tristeza da música triste? E a quem você deve perguntar – a um compositor ou a uma psicóloga cognitiva? (WATTENBERG, 2010).

tonalidade em si. Depois de apresentar vários exemplos que, segundo ele, desafiam essa ideia<sup>42</sup>, Wattenberg (2010, não paginado) conclui:

*If a cognitive psychologist tries to tell you the minor third interval is intrinsically, universally "sad," it's not true. If, on the other hand, she says that, well, in the right musical setting and cultural context it can help evoke an ultimately elusive range of sad or mysteriously unresolved emotions—then it's not exactly new.*<sup>43</sup>

Na mesma linha, o texto *Is Major Really Happy and Minor Sad?* (EWELLUL, 2020?), publicado no site *School of Composition*, também tenta relativizar a associação “menor-triste” elencando músicas que fogem a essa regra: *Nothing Compares to You*, composta por Prince e famosa na voz da cantora irlandesa Sinead O’Connor, seria um exemplo de canção triste em Fá maior, enquanto *Happy*, de Pharrell Williams, é uma música alegre escrita em Fá menor. O texto conclui:

*[...] major doesn't automatically mean 'happy music' and minor doesn't automatically suggest 'sad music'. If major is 'happy' and minor is 'sad', it is true only to a very limited extent. [...] Tempo, timbre, rhythm, melody and other musical attributes have a greater impact on the overall effect of a piece. Think about it like this: The color red has different meanings to us depending on when, where and how it is used. On Valentine's day it symbolizes passion and love but in a war movie it means blood and pain. [...] It's all about the context, the subtleties and the nuances. Everything in music and what it means to people is determined by the subtle details within the notes. It's the context that makes it what it is.*<sup>44</sup> (EWELLUL, 2020, não paginado).

Em que pesem a ocasional controvérsia, eventuais queixas de compositores e a possível frustração advinda da necessidade de reconhecer que, assim como em inúmeros outros fenômenos da vivência humana, estamos diante de uma **tendência e não de uma regra determinista**, o que se tem é uma ampla gama de evidência experimental documentando que os indivíduos no ocidente associam os modos maiores e menores com emoções específicas (KASTNER; CROWDER, 1990, p. 190). Mais uma vez, voltamos às palavras de Deryck Cooke:

---

<sup>42</sup> Entre os exemplos citados pelo autor, estão *Eleanor Rigby*, do grupo The Beatles - segundo o Wattenberg, mantenha a melodia, altere qualquer outro elemento, letra, andamento, ou arranjo, para conseguir um resultado não-triste - *We Are the Champions*, da banda Queen, e *My Favorite Things*, do filme *A Noviça Rebelde*.

<sup>43</sup> Em tradução livre: “Se uma psicóloga cognitiva tentar te dizer que o intervalo de terça menor é intrinsecamente, universalmente, “triste”, não é verdade. Se, por outro lado, ela disser que, bem, no ambiente musical e no contexto cultural certos ele pode ajudar a evocar uma, em última instância, elusiva gama de emoções tristes ou misteriosamente não-resolvidas - então não é exatamente novidade.”

<sup>44</sup> Em tradução livre: “(modo) maior não necessariamente significa “música feliz” e menor não sugere automaticamente “música triste”. Se maior é “feliz” e menor é “triste”, isso é verdade apenas até certo ponto. [...] Andamento, timbre, ritmo, melodia e outros atributos musicais têm impacto maior no efeito geral de uma peça. Pense desse jeito: a cor vermelha tem significados diferentes para nós dependendo de quando, onde e como é usada. No Dia dos Namorados, ela simboliza paixão e amor, mas em um filme de guerra significa sangue e dor. [...] Tudo depende do contexto, das sutilezas e das nuances. Tudo no que se refere à música e ao que ela significa para as pessoas é determinado pelos detalhes sutis entre as notas. É o contexto que faz dela o que é.”

*People often single out a lively minor piece from the eighteenth century and say, 'Here is proof, if proof were needed, that minor music need not be sad'. The fact that proof does seem to be needed shows the inherent connection between the minor system and painful emotions; exceptions will have to be dealt with when the time arises.*<sup>45</sup> (COOKE, 1959, p. 51)

### 3.3 Metodologia

Para este segundo experimento, foram selecionadas cantigas de ninar a partir da coletânea *Lullabies of the World* (THE VOICE PROJECT, 2013). No site oficial do projeto na internet, estão disponíveis interpretações artísticas dessas cantigas e, mais importante neste momento, as partituras destas músicas. O processo de seleção e preparação das músicas para o experimento é descrito a seguir.

O objetivo é verificar se, como e até que ponto a presença de intervalos de terça maior e de terça menor no contorno melódico altera a percepção da emoção expressa, mas sem incluir o conteúdo semântico como uma das variáveis. Desta forma, era importante minimizar as chances de que os informantes conseguissem assimilar esse conteúdo semântico expresso pelas palavras cantadas. Assim, logo de início, foram desconsideradas as gravações em português e espanhol, pois os informantes (se não todos, com certeza a maioria esmagadora deles) são falantes de português como língua materna, e o espanhol apresenta notável nível de inteligibilidade para falantes de português. Também foram descartadas canções em inglês (por ser uma língua amplamente difundida e estudada), em francês e em italiano (nestes últimos dois casos, por serem idiomas estudados com relativa frequência no Brasil, além de serem idiomas com alto prestígio e carregados de associações positivas, pelo menos no ocidente).

As interpretações disponíveis no site do projeto incluem acompanhamento instrumental, um aspecto negativo para o propósito desse experimento. Nosso objetivo é verificar na melodia um fenômeno já observado na fala (deslexicalizada ou não), e um acompanhamento instrumental introduz outras variáveis além da voz. Assim sendo, reconheceu-se a necessidade de realizar novas gravações para as canções selecionadas. Desta forma, foi necessário levar em consideração também a complexidade rítmica, a tessitura vocal da linha melódica e, não menos importante, a complexidade do sistema fonológico da letra da música. Deu-se preferência,

---

<sup>45</sup> Em tradução livre: “As pessoas frequentemente selecionam uma peça menor jovial do século dezoito e dizem, 'Eis a prova, se prova era necessária, de que música menor não necessita ser triste'. O fato de que uma prova parece ser necessária mostra a conexão inerente entre o sistema menor e emoções dolorosas; ter-se-á que lidar com as exceções quando a ocasião chegar”

portanto, a músicas cuja fórmula de compasso fosse comum na música popular ocidental, como 3/4 e 4/4, e foram excluídas canções como *Jasoda Hari palne*, da Índia, cuja partitura indica um compasso em 7/4, e *Oj Jano, Jano*, da Macedônia, cuja fórmula de compasso é indicada por 3+2/8+2.

Com todos os aspectos elencados acima em mente, foram selecionadas as seguintes composições: *Itsuki no Komori-uta* (em japonês), *Tuu, tuu, tupakkarulla* (em finlandês), *Lala, mtoto lala* (em suaíli) e *Aja tutaja* (em esloveno).

Sabe-se que a taxa de elocução é um dos aspectos que podem ser modificados de acordo com a carga emocional ou a atitude de um falante em um enunciado (ver, por exemplo, “taxa de elocução” em Silva, 2019). Para mitigar possíveis efeitos da diferença de andamento entre as composições, todas as canções foram gravadas com o mesmo andamento ( $\downarrow = 80$  batidas por minuto).

O teste de percepção foi conduzido de forma remota, através de um questionário *online* na plataforma Google Docs. O endereço do questionário foi divulgado pelo pesquisador e por pessoas de seu círculo social próximo em mídias sociais como Facebook, Instagram e também por WhatsApp, individualmente ou em grupos. A primeira página do questionário, além de apresentar uma breve explicação do experimento, pedia ao voluntário que escolhesse entre duas versões do questionário, a versão azul e a versão vermelha. Os informantes que escolheram a versão vermelha ouviram gravações alteradas digitalmente (ver subseção 3.4, mais adiante), enquanto os que escolheram a versão azul ouviram as gravações originais (ver apêndice A).

Quando expostos às gravações, foi pedido aos informantes que associassem o estímulo ouvido a uma dentre quatro imagens de expressões faciais médias (ver figura 7) das emoções da coleção *The Averaged Karolinska Directed Emotional Faces – AKDFE* (LUNDQVIST; LITTON, 1998) apresentadas na tela junto ao estímulo, de maneira semelhante ao último experimento descrito por Peres (2016). As AKDEF são "médias" obtidas através da mescla das fotos em preto e branco de diversos atores expressando a mesma emoção. Por exemplo, as fotos de atores do sexo masculino expressando tristeza são "empilhadas" (seguindo a analogia feita por Peres, 2016, p. 191) digitalmente, e a partir dessa pilha uma média é calculada, de modo que a imagem resultante apresenta as características em comum em cada um dos atores expressando essa determinada emoção. No estudo de Peres (2016), trechos de fala emotiva foram selecionados da plataforma de streaming de vídeo YouTube e divididos em quatro emoções básicas (raiva, medo, tristeza e alegria), e pediu-se a ouvintes leigos que associassem cada um desses trechos de fala emotiva a uma AKDFE. Neste estudo, assim como em Peres (2016), era obrigatório que o participante associasse cada gravação a uma, e apenas uma,

imagem. Seguindo o mesmo raciocínio exposto na seção 2.3, decidimos por não incluir a emoção *medo*, optando, assim, por mostrar um rosto com uma expressão neutra. Também se considerou que, sendo a escolha da emoção expressa na gravação uma escolha categórica, sem gradação, o rosto com a expressão neutra poderia ser escolhido por informantes que não reconhecessem uma emoção específica sendo expressa, ou tivessem dificuldades na categorização.

**Figura 7** - Exemplos de expressões faciais médias das emoções; cada imagem representa a expressão média de 35 atores para uma emoção específica. Da esquerda para a direita: raiva, alegria, neutralidade e tristeza.



Fonte: Lundqvist; Litton, 1998.

### 3.3.1 Informantes

Os dados foram coletados ao longo do mês de abril de 2021. Neste período, o Brasil e o mundo passavam por um período bastante conturbado, devido à pandemia de Covid-19. Esta situação, por si só, inviabilizaria que a coleta de dados para o Experimento 2 (bem como para o Experimento 3, a ser discutido mais adiante nesta tese) fosse realizada presencialmente. De todo modo, mesmo que a situação sanitária do país fosse mais favorável à época da coleta, há de se convir que a aplicação de um questionário para coleta de dados é muito mais facilmente conduzida *online*.

É claro que conduzir a aplicação de um questionário presencialmente permite maior controle ao pesquisador. Pela internet, não podemos controlar possíveis estímulos externos que distraiam o informante. Também não é possível controlar o tipo de equipamento usado para a reprodução do estímulo, e é bastante difícil aferir ou limitar a quantidade de vezes que esse estímulo é reproduzido ou o tempo que o informante demora para completar o questionário<sup>46</sup>.

<sup>46</sup> Atualmente, existem ferramentas e serviços que permitem um controle maior por parte do pesquisador, mas tais serviços são mais complicados de usar e, muitas vezes, os preços são pouco acessíveis.

Por outro lado, a aplicação *online* facilita a adesão à pesquisa, uma vez que o indivíduo não precisa se deslocar até um lugar determinado para participar, e não há limitação de horário, sendo assim uma opção mais conveniente para o informante.

Obviamente, há sempre o risco de algum “ponto cego” causar distorções no resultado, e é preciso ter em mente a forma como os dados foram coletados na análise e discussão dos resultados. Lembramos, porém, que a aplicação de um questionário presencialmente também não garante necessariamente precisão total dos dados (os informantes podem, por exemplo, mentir, ou não prestar atenção ao estímulo e assinalar uma resposta aleatória). Por outro lado, há de se considerar que, por se tratar de uma participação totalmente voluntária, os informantes teriam pouca ou nenhuma motivação para sabotar a pesquisa dando respostas erradas ou não prestando atenção aos estímulos. Isto posto, passamos às informações acerca do recrutamento e dos informantes em si.

A princípio, o questionário foi divulgado pelo próprio pesquisador em suas redes sociais e entre os participantes do grupo de pesquisa coordenado pela orientadora. Em um segundo momento, foi pedido que amigos e conhecidos do pesquisador repassassem o convite adiante, também em redes sociais e grupos de WhatsApp. Sabemos, por exemplo, que até pessoas de outros estados participaram da pesquisa: havia a possibilidade de os informantes interessados deixarem seu endereço de *email* para receberem mais informações acerca do estudo. Depois de o pesquisador enviar o *email* para os interessados explicando a origem das músicas e a quais línguas eles tinham sido expostos, alguns informantes se mostraram tão entusiasmados que compartilharam espontaneamente suas impressões e algumas informações sobre si mesmos, como sua área de formação.

Foi pedido aos participantes que informassem sua idade, quantas e quais línguas falavam, e se tinham envolvimento ativo com música (cantando, tocando um instrumento, ou exercendo alguma atividade relacionada). Seria inviável pedir que todos comprovassem seu conhecimento linguístico e seu envolvimento com música por meio de documentos ou gravações. Assim, confiamos na autodeclaração dos participantes.

Em relação às línguas, não foi pedido que os informantes declarassem de qual ou quais idiomas eram falantes nativos, tampouco que discriminassem o nível de conhecimento dos idiomas estrangeiros que afirmassem conhecer. Tomamos essa decisão pelos seguintes fatores:

- Considerando que a pesquisa foi conduzida no Brasil por pesquisador brasileiro sediado no estado de São Paulo, e que os participantes foram recrutados a partir dos contatos do pesquisador, são relativamente baixas as chances de que os informantes não fossem falantes

nativos de português brasileiro (já que a região não tem uma população massiva de imigrantes ou de indígenas, por exemplo);

- A compreensão do conteúdo semântico expresso no estímulo sonoro não era necessária para a participação na pesquisa, já que o objetivo era avaliar a percepção do sinal sonoro em si, em suas características físicas e fonético-fonológicas puras (contorno melódico, proporção de vogais e de consoantes, etc);

- Apesar de haver indícios que a familiaridade com o idioma influencia o julgamento (REITERER et al. 2020), familiaridade não implica necessariamente proficiência, e também não está claro se e como a proficiência em um idioma se correlaciona com uma avaliação mais positiva – da mesma forma, é possível gostar ou não de um idioma, achá-lo “melodioso” ou “áspero”, sem ter qualquer conhecimento.

Os informantes explicitaram seu conhecimento linguístico respondendo diretamente à pergunta "Quantas línguas você fala? Inclua a sua língua materna e qualquer idioma no qual você consiga se comunicar minimamente, mesmo que com alguma dificuldade" e "Quais línguas você fala?". No caso de um informante que afirmasse falar quatro línguas e que essas línguas eram português, inglês, espanhol e alemão, consideramos esse informante um falante dessas quatro línguas<sup>47</sup>. Em outras palavras, quando mencionamos “falantes de alemão”, por exemplo, estamos nos referindo a todos os informantes que afirmaram ser falantes de alemão, independente do *status* do idioma (língua materna ou língua estrangeira) e do nível de proficiência.

Mais relevante para o experimento em questão nesta seção é que, caso o participante conhecesse (ou acreditasse conhecer) o idioma da melodia sendo cantada, a orientação era não categorizar a imagem em um dos sentimentos, e sim assinalar a opção “Eu conheço essa língua”. Tomamos tal decisão para evitar que um possível nível de compreensão semântica da letra interferisse na categorização. Se desejassem, os informantes também podiam apontar qual idioma acreditavam estar ouvindo. Os erros e acertos no reconhecimento dos idiomas serão abordados mais adiante. Por enquanto, basta dizer que não era um problema o informante identificar erroneamente a língua do estímulo, ou mesmo achar que conhecia o idioma sem de fato conhecê-lo. Afinal, se uma pessoa tem um histórico de traumas e associações negativas com a língua italiana, por exemplo, não é necessário que a pessoa esteja de fato ouvindo italiano para que sua avaliação seja afetada. Basta que ela pense estar ouvindo italiano, e toda a rede de associações será mobilizada, potencialmente afetando os resultados.

---

<sup>47</sup> Não verificamos casos de incongruência nessas duas respostas (por exemplo, alguém que afirmasse falar três idiomas na primeira questão e dissesse ser falante apenas de português e espanhol na segunda questão.)

Por fim, deixemos claro que os informantes foram divididos entre dois grupos antes de começarem a ser expostos aos estímulos. Cada grupo foi exposto a *uma* das duas versões possíveis de cada melodia. Isto significa que nenhum informante ouviu as *duas* versões de uma mesma melodia.

### 3.3.2 Estatística

Para que se possa tirar qualquer conclusão a partir dos dados colhidos neste primeiro experimento, observar os números brutos não é o suficiente. Tomemos, por exemplo, a canção *Aja tutaja*, utilizada nesse estudo. Um grupo composto de 99 indivíduos ( $N = 99$ ) foi exposto à versão original da canção; destes, 13 julgaram que a canção expressava tristeza. Paralelamente, um segundo grupo, composto de 94 indivíduos ( $N = 94$ ) foi exposto à versão alterada da canção, e 18 pessoas desse segundo grupo consideraram que a música expressava tristeza. Porém, por mais tentador que nos possa parecer, não podemos nos apressar e concluir que as alterações da gravação fazem com que a melodia seja percebida como triste com mais frequência. Afinal, como se pode ter certeza que essa diferença de seis indivíduos é estatisticamente significativa, e não aleatória? Em outras palavras, qual a chance de que essa diferença observada ser fruto de mero acaso e que as alterações realizadas não tenham sido relevantes na percepção dos informantes? É preciso recorrer a testes estatísticos para esclarecer essa dúvida e, assim, tirar conclusões sólidas.

A escolha da prova estatística a ser utilizada depende do tipo de dado que se pretende analisar. Dados numéricos (quantitativos), como a altura e o peso de alguém, são chamados de dados paramétricos. Por outro lado, temos os dados categóricos, que não são medidos numericamente e a partir dos quais não se pode estabelecer um intervalo preciso entre categorias. Tomemos como exemplo uma pesquisa de satisfação de clientes, em que um cliente pode se declarar como tendo ficado “Muito satisfeito”, “Satisfeito”, “Neutro”, “Um pouco insatisfeito” e “Muito insatisfeito”. Obviamente, um cliente “Muito satisfeito” apresenta maior satisfação do que um cliente apenas “Satisfeito”, mas não é possível ter certeza que a variação entre “Muito satisfeito” e “Satisfeito” tem a mesma dimensão do que a variação entre “Neutro” e “Um pouco insatisfeito” ou entre “Um pouco insatisfeito” e “Muito insatisfeito”. O valor do intervalo entre as categorias não é conhecido ou determinável. Também podemos ter em mãos dados categóricos nominais, para os quais não há uma escala de grandezas, intensidade ou algo

do gênero, apenas a opção de pertencimento a duas (ou mais) categorias. Podem ser mencionados como exemplos de dados nominais respostas divididas entre “Sim” ou “Não”, bem como os dados de cor/raça da população colhidos pelo Instituto Brasileiro de Geografia e Estatística ("branca", "preta", "parda", "amarela" ou "indígena"). Os dados categóricos são também chamados de não-paramétricos.

Neste primeiro experimento, lidamos justamente com dados categóricos nominais, não-paramétricos. Aos informantes, foi pedido que *categorizassem* sua percepção de acordo com as emoções expressas nas imagens. Assim sendo, o teste estatístico deveria ser também não-paramétrico.

Há de se considerar também o tipo de amostras (isto é, a natureza dos casos observados) com as quais estamos lidando. Quando os elementos das amostras são oriundos dos mesmos indivíduos, temos amostras dependentes (é o caso, por exemplo, de um grupo de pacientes cujos dados são colhidos antes e depois de uma intervenção clínica). No caso deste experimento, as amostras são independentes, pois cada informante foi exposto a apenas uma ou outra versão da melodia (em outras palavras, cada indivíduo categorizou o estímulo apenas uma vez, e sem qualquer interferência de seus pares).

Temos, portanto, duas amostras (conjuntos de dados) X e Y independentes. Precisamos verificar se a proporção de categorizações (ocorrências) de cada uma das emoções é a mesma em ambas as amostras. Em outras palavras, há de se verificar se a proporção da categoria “Alegria” é a mesma em X e Y, se a proporção de “Tristeza” é a mesma em X e Y, e assim por diante.

**Quadro 1** - Exemplo de tabela de contingência

Ocorrências da categoria “Tristeza”	Amostra X (melodia original)	Amostra Y (melodia alterada)
“Tristeza”	13	18
Não-Tristeza	85	75

Nestas condições, podemos utilizar, por exemplo, o teste Exato de Fisher (figura 8). Para a execução do teste, os dados são dispostos em uma tabela de contingência 2 x 2 no *software* estatístico a ser utilizado. Aqui, utilizamos o programa de distribuição livre BioEstat, versão 5.3.

**Figura 8** - Captura de tela do *software* BioEstat, com os dados e resultados do teste Exato de Fisher.

Teste Exato de Fisher

Imprimir

**Dados**

		Colunas		Total / Linhas
		1	2	
Linhas	1	13	85	98
	2	18	75	93
Total / Colunas		31	160	191

Executar Sair

**Resultados**

Unilateral	Bilateral
$p = 0.1725$	$p = 0.3267$

A hipótese nula ( $H_0$ ) é a de que os escores do sentimento *Tristeza* independem das alterações realizadas digitalmente. A hipótese alternativa ( $H_1$ ) afirma que há diferença entre os grupos da pesquisa, ou seja, que as ocorrências de percepção da melodia como expressando *Tristeza* foram alteradas pelas modificações no estímulo. Os valores expressos por  $p$  se referem à probabilidade de se aceitar erroneamente uma hipótese alternativa como verdadeira. Normalmente, adota-se um valor de  $p$  inferior a 0,05 ou 0,01 para rejeitar a hipótese de nulidade. Tais valores significam aceitar a probabilidade de ocorrência de um erro em 20 (5%) ou um erro em 100 (1%), respectivamente. No caso em questão, é possível verificar na figura 8 que os dois valores de  $p$  fornecidos pelo teste são superiores a 0.05. Portanto, a decisão estatisticamente razoável é rejeitar a hipótese alternativa e aceitar a hipótese nula: os escores de *Tristeza*<sup>48</sup> observados não dependem das alterações no estímulo.

<sup>48</sup> É importante deixar claro que, no exemplo apresentado, reconhecer que as modificações do estímulo não influenciaram a percepção de tristeza não significa que estas modificações também não influenciaram a percepção de outras emoções. É possível que a diferença entre os estímulos não seja relevante na percepção de tristeza, mas ainda assim altere de forma significativa o número de pessoas que classificam a melodia como expressando raiva, por exemplo.

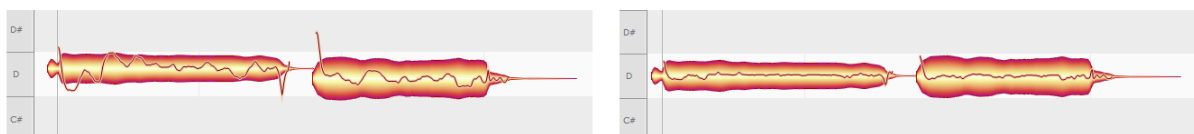
### 3.4 Descrição dos estímulos, com resultados

Antes de qualquer gravação, foram decididas quais notas musicais seriam alteradas digitalmente. Esse processo foi supervisionado pela professora de piano Natalia Gomes Sartori, formada no curso de música da Universidade de São Paulo (USP), campus de Ribeirão Preto. Tentou-se modificar o menor número de notas e de forma menos drástica possível, já que quanto mais drástica a alteração de altura melódica realizada digitalmente, maior a chance de se imprimir à voz uma qualidade notavelmente artificial, podendo até mesmo soar robótica, o que poderia distrair os informantes.

Cada uma das músicas foi gravada apenas uma vez, em sua versão original, por um vocalista amador, nascido e criado na região central do interior de São Paulo (entre as cidades de Itápolis e Araraquara), tinha 31 anos na época das gravações, seis anos de experiência cantando coral e três anos de estudo de piano. O vocalista teve à sua disposição as partituras e arquivos MIDI das melodias, mas não teve acesso nem às interpretações artísticas profissionais nem às traduções das letras, ambas disponíveis no site oficial do projeto.

Depois das gravações, os arquivos de áudio foram editados digitalmente no software Melodyne Studio, versão 4.2.3.001. Primeiramente, verificou-se a precisão da afinação do vocal. Apesar de não haver momentos em que o vocal correspondia a uma nota totalmente diferente da prevista na partitura, a voz apresentava pequenas modulações e micro-ondulações. Essas ondulações são naturais e esperadas, considerando que se tratava de um vocalista amador. Porém era conveniente removê-las para garantir que elas não estivessem interferindo no resultado final. A figura 9 mostra a comparação entre as duas notas Ré finais na gravação da música *Tupakkarulla*. Assim, obtivemos duas versões diferentes de cada música.

**Figura 9** - Correção de afinação no *software* Melodyne.



Capturas de tela do software Melodyne mostrando as duas últimas notas da melodia *Tupakkarulla*. À esquerda, antes da correção de afinação, e à direita, depois de a gravação ser editada digitalmente.

Na figura 9, imagem à esquerda, é possível observar que as manchas amarelas estão situadas na altura da nota Ré. Isso significa que, no contexto em questão, ambos os segmentos

seriam interpretados como a nota Ré. Porém, a linha vermelha representando a frequência fundamental, apesar de “orbitar” a frequência correspondente a Ré, apresenta oscilações, caracterizando uma afinação que não é perfeitamente estável. A imagem à direita mostra as mesmas duas notas após as correções de afinação conduzidas pelas ferramentas do programa. A linha vermelha se mostra muito mais estável. Houve o cuidado de não tornar a afinação artificialmente perfeita, com a linha vermelha totalmente reta, uma vez que isso possivelmente faria a voz soar robótica e sintética, potencialmente afetando os resultados. Esse procedimento foi feito com cada uma das gravações, em todas as notas que apresentavam oscilações consideráveis<sup>49</sup>. Após as correções, os arquivos de áudio resultantes foram submetidos à apreciação de dois músicos (um cantor e compositor de músicas infantis e um ator com histórico de atuação como cantor em musicais). A instrução dada aos dois músicos resumiu-se a pedir que verificassem se as gravações estavam boas e se havia algo estranho nelas. Uma vez que nenhum dos dois percebeu qualquer aspecto estranho e ambos afirmaram que as gravações estavam boas, consideramos esses arquivos finais as versões originais.

Posteriormente, estes arquivos considerados “versões originais” foram editados mais uma vez. Esta segunda edição compreendeu em modificar a altura das notas correspondentes aos intervalos de terça maior ou menor. Mais uma vez, as gravações foram submetidas à apreciação de dois músicos independentes (desta vez, dois professores) para garantir que o resultado final não estava artificial e que a manipulação não estava óbvia. Os pareceres foram positivos, e essa nova leva de gravações editadas foi tomada como as “versões modificadas” das músicas.

Evidentemente, é impossível garantir que, ainda que em nível subconsciente, a percepção emocional que o vocalista teve das melodias não tenha, em algum grau, guiado sua performance. Certamente, outros vocalistas cantando as mesmas melodias apresentariam outros nuances de intensidade, fraseado e até mesmo de qualidade de voz. No entanto, esses outros fatores não nos são relevantes por dois motivos. Em primeiro lugar, porque não pretendemos fazer uma análise extensiva de todos os parâmetros acústicos que exercem influência na percepção da carga emocional de uma melodia, e sim verificar se, todo o resto mantido igual, os intervalos de terça maior ou de terça menor exercem influência considerável nessa interpretação. Em segundo lugar, mesmo que a qualidade de voz por exemplo, tenha sido um fator mais importante na percepção dos informantes do que as notas musicais da melodia, a

---

<sup>49</sup> Sabemos que a definição de “considerável” pode ser bastante subjetiva. Como parâmetro, adotamos um indicador visual (a presença de picos e vales na linha da frequência fundamental) e um auditivo, este mais subjetivo (a percepção, por parte do pesquisador, de uma nota oscilando de forma perceptível em termos de afinação).

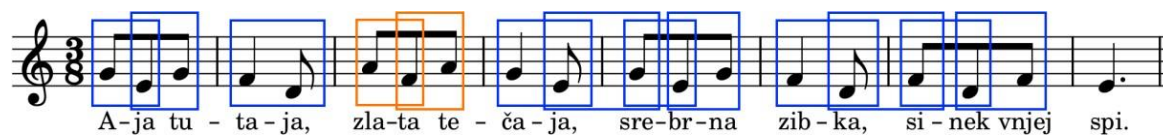
qualidade de voz apresentada nas duas gravações é a mesma, uma vez que ambas as versões são, a princípio, a mesma gravação do mesmo falante. Assim, parece razoável afirmar que as diferenças de percepção das duas versões de uma mesma música são acarretadas – se não totalmente, certamente principalmente – pela alteração digital das alturas melódica.

A seguir, abordaremos cada uma das canções utilizadas na pesquisa, bem como as alterações realizadas e as diferenças de percepção acarretadas por essas alterações.<sup>50</sup>

### *Aja Tutaja*

Letra em esloveno, melodia e texto originários da Eslovênia. A armadura de clave<sup>51</sup> sugere as tonalidades Dó maior ou Lá menor. É interessante notar que, na melodia, não há destaque para as possíveis notas tônicas<sup>52</sup>: não há qualquer nota Dó, e mesmo as notas Lá aparecem apenas duas vezes, rapidamente, no terceiro compasso<sup>53</sup>.

**Figura 10** - Partitura de *Aja tutaja*. Os intervalos de terça menor aparecem destacados em azul, e os intervalos de terça maior aparecem destacados em laranja.



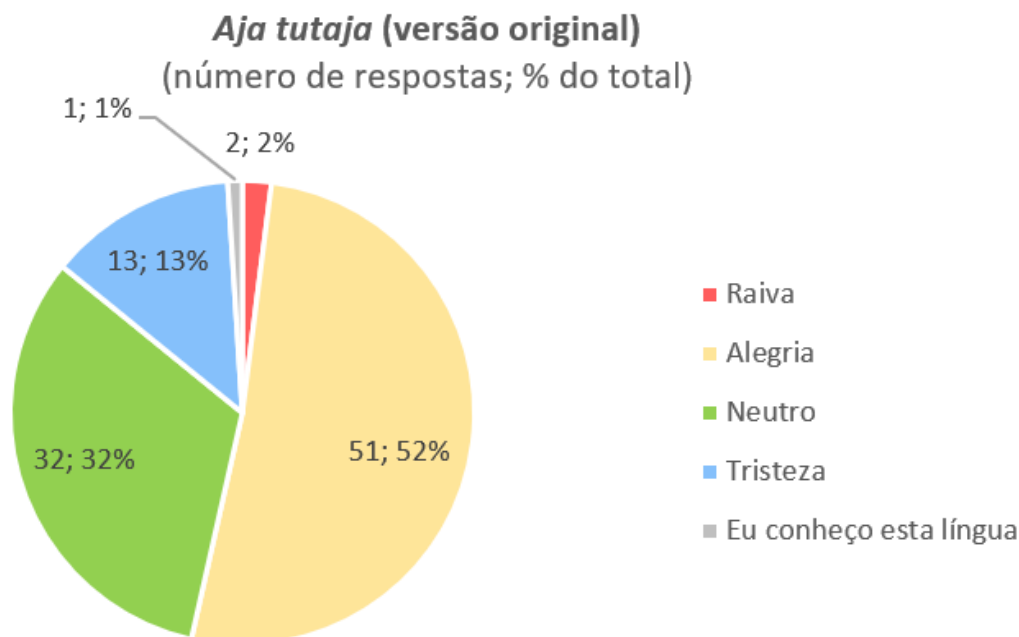
<sup>50</sup> A afirmação de que as diferenças de percepção foram acarretadas pelas alterações realizadas nas partituras (e, posteriormente, nas gravações) pode parecer temerária a alguns. De fato, a voz carrega muito mais informações do que “apenas” uma altura melódica e eventual conteúdo semântico. Informações como taxa de elocução, registro de voz e até mesmo diferenças na articulação dos fonemas podem ser relevantes. Porém, considerando que o experimento aqui abordado foi conduzido a partir de apenas uma gravação de cada melodia, a única variável era justamente a alteração da altura de algumas notas. Todas as demais características do sinal vocal foram mantidas intactas. Assim, parece-nos razoável creditar as diferenças de percepção às mudanças na altura das notas (obviamente, desde que os testes estatísticos apontem que as diferenças de escores são estatisticamente significativas).

<sup>51</sup> Na notação ocidental, armadura de clave é o sinal colocado no começo de uma pauta (conjunto de cinco linhas no qual as notas musicais são escritas) para estabelecer a altura melódica de uma das linhas, e, portanto, de todas as demais linhas (HILEY, 2001).

<sup>52</sup> Em música, chama-se *tônica* a primeira nota de uma escala, a qual configura o centro tonal harmônico melódico de uma tonalidade e que dá nome à escala. Assim, por exemplo, a nota Dó é a nota tônica das escalas de Dó (maior e menor).

<sup>53</sup> Tradução livre da letra, com base na tradução para o inglês disponível no site original: Aja tutaja, duas dobradiças douradas / um berço de prata / onde dorme o filhinho.

**Gráfico 5** - Percepção de sentimentos expressos na versão original de *Aja Tutaja*, em que os intervalos de terça são predominantemente menores.



Mesmo com 11 intervalos de terça menor e apenas um intervalo de terça maior, a maioria dos informantes considerou que a canção expressa Alegria. Tem-se, aqui, um indício de que os informantes interpretaram a melodia como estando no tom de Dó maior. De fato, as três primeiras notas da melodia (Mi-Sol-Mi) podem ser interpretadas como pertencentes a um acorde de Dó – nesse caso, seria esperado que a nota Dó fosse tocada em algum instrumento que estivesse fazendo o acompanhamento da melodia. O mesmo ocorre com a última nota, Mi, que também estaria presente num acorde de Dó maior.

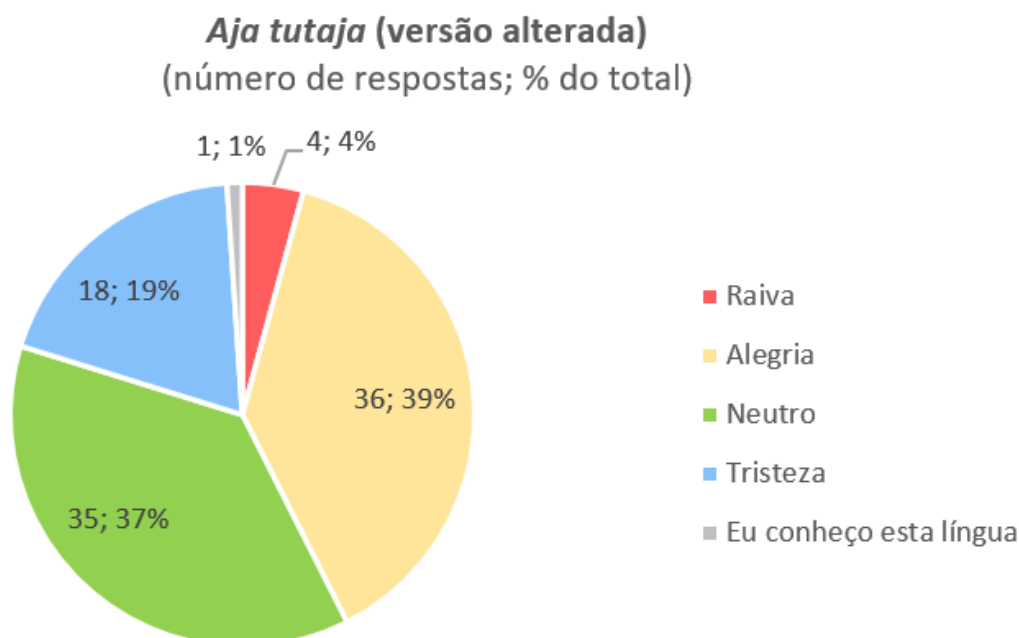
**Figura 11** - Partitura da versão alterada de *Aja tutaja*. Os intervalos de terça maior aparecem destacados em laranja.

A-ja tu - ta - ja, zla-ta te - ča - ja, sre-br-na zib-ka, si - nek vnjej spi.

As alterações nas alturas melódicas em *Aja Tutaja* transformaram todos os intervalos de terça na melodia em intervalos de terça maior. Para esta canção em especial, o foco se limitou

aos intervalos individuais, não havendo qualquer esforço no sentido de adequar a melodia a uma escala ou tonalidade específica.

**Gráfico 6** - Percepção de sentimentos expressos na versão alterada de *Aja Tutaja*, em que todos os intervalos de terça são maiores.



Ainda que o número de informantes interpretando a canção como expressando Tristeza ou Neutralidade tenha oscilado levemente em termos absolutos, o teste exato de Fisher mostrou que apenas a variação na percepção de expressão de Alegria foi estatisticamente relevante ( $p=0,0441$  para alegria, contra  $p=0,1725$  e  $p=0,2845$  para Tristeza e Neutralidade, respectivamente). Ainda assim, há uma queda de aproximadamente 29% na percepção de Alegria, configurando assim um resultado que não deve ser ignorado.

Quando se alteram os intervalos, muda-se o acorde inicial, o do meio e o do fim para Dó menor, o que pode ter contribuído para a diminuição da percepção de alegria. No entanto, a alteração nos demais acordes não gera uma cadência tonal (eles não pertencem ao campo harmônico de Dó menor). Este pode ser o motivo da estranheza por parte dos ouvintes, fazendo com que o número de informantes que marcaram “neutro” aumentasse, por não conseguirem identificar uma cadência em tonalidade menor<sup>54</sup>.

<sup>54</sup> Agradecemos ao Prof. Dr. Daniel Soares da Costa por chamar nossa atenção para esse aspecto.

### *Itsuki no Komori-uta*

A canção *Itsuki no Komori-uta* apresenta letra em japonês, sendo que tanto a letra quanto a melodia são originárias do Japão. O título significa, literalmente, "Canção de Ninar de Itsuki", Itsuki sendo uma vila na ilha de Kyushu<sup>55</sup>.

**Figura 12** - Partitura original de *Itsuki no Komori-uta* do site Lullabies of the World.

1. O - do - ma bon - giri bon - giri bon kara sa - kya o - ran - do bo -

n ga ha - yo ku - - rya ha - yo mo - do - ru.

A partitura sofreu uma leve adaptação no compasso seis, no qual uma sequência de quatro notas (Fá-Mi-Ré-Mi) foi simplificada para apenas duas (Fá-Mi), para facilitar a performance do vocalista, que não é um profissional. Também se optou por gravar o segundo verso da letra em vez do primeiro, já que a quantidade de sílabas no segundo verso se encaixa com maior facilidade nas notas da melodia (vide o segundo compasso de ambas as partituras). A melodia efetivamente gravada e utilizada no experimento apresenta 11 intervalos de terça maior e um intervalo de terça menor. Na figura 13, os intervalos de terça maior estão destacados em retângulos laranja, enquanto o intervalo de terça menor aparece dentro do retângulo azul, entre os compassos seis e sete.

A escala da melodia *Itsuki no Komori-uta*, com apenas um Si bemol, sugere as tonalidades de Fá maior e de Ré menor. A presença da nota Ré em posições de final de frase (nos compassos dois e seis) e, especialmente, no compasso oito, encerrando a melodia, levam o ouvinte a interpretar a melodia como estando no tom de Ré menor.

<sup>55</sup> Tradução livre das duas primeiras estrofes, com base na tradução para o inglês disponível no site original: Vai demorar até o verão / e então não estarei mais aqui / Se o verão chegasse mais cedo / Eu poderia voltar para casa antes / Eu sou como um mendigo, mendigo / Essas pessoas são os ricos / As pessoas ricas têm cinturões finos / e belos quimonos.

**Figura 13** - Partitura adaptada de *Itsuki-no Komori-uta*, com a letra do segundo verso Os intervalos de terça menor aparecem destacados em azul, e os de terça maior, em laranja.

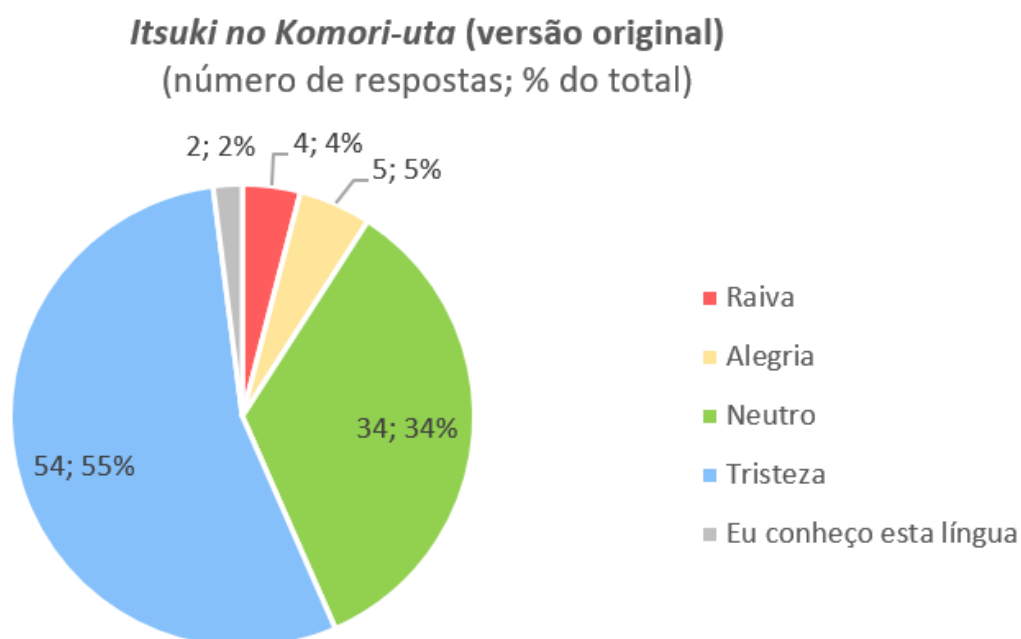
The musical score is in 3/4 time and B-flat major. The lyrics are: O-do-ma kan-jin an-hi-to - ta - cha-yo-ka - shi Yo - ka-sha yo - ka o bi Yo - ka ki - mon.

Intervals are highlighted as follows:
 

- Orange boxes (Major Third): O-do-ma, an-hi-to, ta - cha-yo-ka, shi Yo - ka-sha, yo - ka
- Blue box (Minor Third): bi

Mais da metade dos informantes (54%) interpretou a melodia como expressando Tristeza, uma quantidade que corresponde a quase onze vezes o número de informantes que interpretou a melodia como expressando Alegria (5%). Se pensássemos apenas do ponto de vista dos intervalos, tal resultado seria inesperado, já que há um predomínio esmagador de intervalos de terça maior, associados à alegria, sobre os de terça menor, associados à tristeza. Mas esta é apenas parte do quadro que se delinea à nossa frente, e não a imagem completa.

**Gráfico 7** - Percepção de sentimentos expressos na versão original de *Itsuki no Komori-uta*, em que os intervalos de terça são predominantemente maiores.



Como já mencionado, mesmo com a presença dos intervalos 3M, a melodia é interpretada como no tom de Ré menor e, como veremos melhor na seção de Discussão, mais adiante, a associação à tristeza se dá mais pelo tom e pela escala do que pelos intervalos isolados. Ainda que haja apenas uma ocorrência do intervalo de terça menor Ré-Fá, a melodia é constituída, basicamente, pelas notas Ré, Fá e Lá, que constituem o acorde de Ré menor. Assim, os intervalos de terça maior não evitam a percepção de tristeza, pois a melodia é basicamente o acorde de Ré menor tocado melodicamente (ou seja, uma nota depois da outra, e não todas ao mesmo tempo).

Voltemo-nos, agora, à gravação alterada digitalmente e a percepção dos informantes acerca dela.

**Figura 14** - Partitura da versão alterada digitalmente de *Itsuki-no Komori-uta*. Os intervalos de terça menor aparecem destacados em azul, e os de terça maior, em laranja.

The figure shows a musical score in treble clef, key of D major (one sharp), and 3/4 time. The melody is written on a single staff. The lyrics are: O - do - ma kan - jin an - hi - to - ta - cha - yo - ka - shi Yo - ka - sha yo - ka o bi Yo - ka ki - mon. The intervals between notes are highlighted with boxes: blue boxes highlight minor thirds (e.g., G4-A4, B4-C5, D5-E5, F5-G5, A5-B5, C6-D6), and an orange box highlights a major third (E5-F5). The score starts with a measure rest and a quarter note G4. The piece ends with a double bar line.

Uma vez que a melodia original apontava para o tom de Ré menor, alteramos a gravação digitalmente para simular uma melodia em Ré maior, mas com a proporção dos intervalos invertida, sendo agora de 11 de terça menor para apenas um de terça maior<sup>56</sup>.

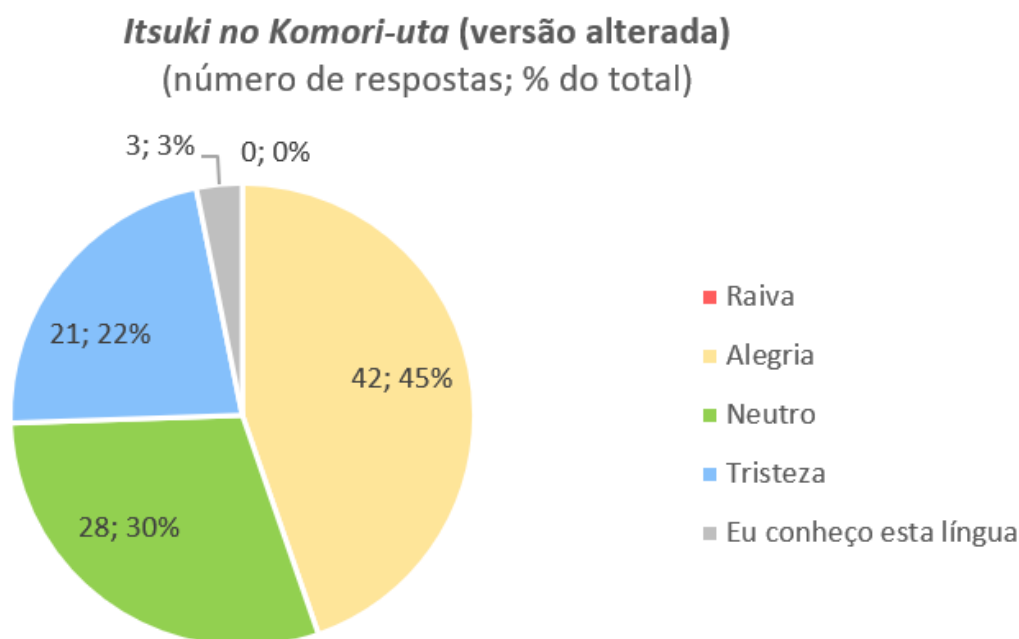
É possível notar que a predominância de intervalos de 3m não se converteu na predominância da percepção de expressão de tristeza. Como se pode observar, o número de informantes que considerou a gravação como expressando tristeza é expressivo, mas substancialmente menor quando comparado ao verificado em relação à gravação original (22% do total contra 54% na versão original), enquanto o número de participantes que reconheceu a emoção Alegria na gravação saltou de 5% para 42%. Através do teste exato de Fisher, verificou-se que a diferença na percepção dos informantes no que se refere à expressão de alegria e tristeza

<sup>56</sup> Cabe lembrar que o total de informantes foi dividido em dois grupos, e que cada grupo ouviu uma das versões de cada melodia, mas não ambas as versões.

foi extremamente significativa estatisticamente ( $p = 0,0000$ ), o que nos permite afirmar que a mudança dos intervalos na melodia acarretou grande variação na percepção – ainda que não necessariamente na direção imaginada: a percepção da tonalidade sobrepõe à percepção dos intervalos individuais de terça maior ou menor.

Isto acontece exatamente pela mesma razão que leva a versão original de *Itsuki-no Komoro uta* a ser percebida como expressando tristeza, apesar do grande número de intervalos melódicos de terça maior. Aqui, porém, a situação é invertida. O Fá sustenido, mesmo fazendo um intervalo de terça maior com um Ré adjacente apenas uma vez, configura um acorde de Ré maior quando combinado com Ré e Lá, as duas outras notas mais frequentes da melodia. Mais uma vez, tem-se a percepção de um acorde (agora, maior) tocado melodicamente, do que decorre a percepção de expressão de alegria.

**Gráfico 8** - Percepção de sentimentos expressos na versão alterada de *Itsuki-no Komori-uta*, em que os intervalos de terça são predominantemente menores.



### ***Lala, mtoto Lala***

*Lala, mtoto lala* é uma cantiga de ninar originária do leste da África, com letra em Suaíli. A figura 15 mostra a partitura da canção<sup>57</sup>.

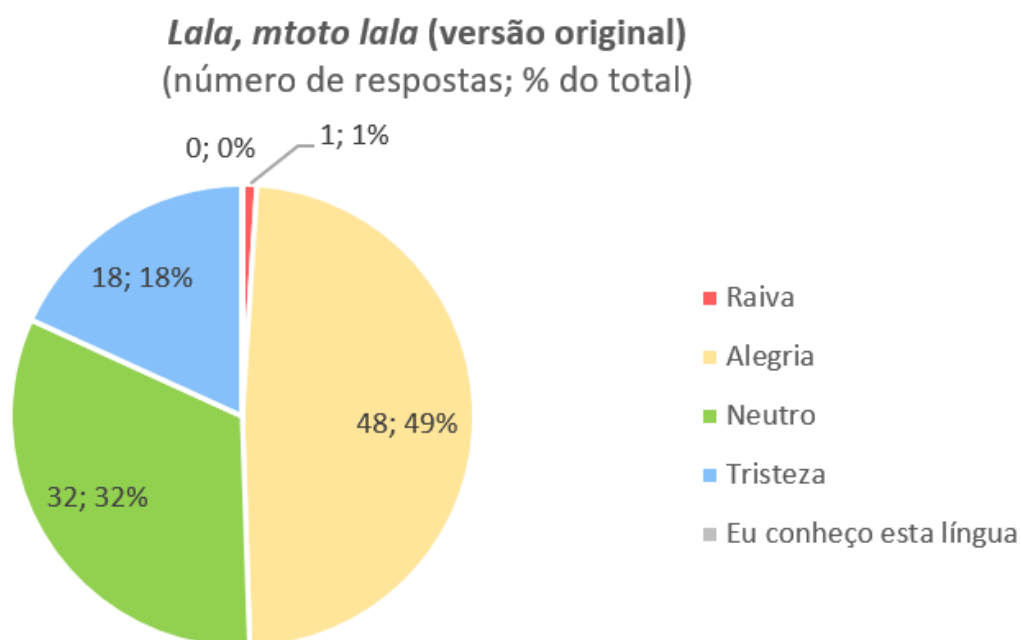
<sup>57</sup> Tradução livre, com base na tradução para o inglês disponível no site original: Durma, bebê, durma / Mamãe está vindo, durma / Ela vai te amamentar, durma.

**Figura 15** - Partitura de *Lala, mtoto lala*. Os intervalos de terça menor aparecem destacados em azul, e os intervalos de terça maior aparecem destacados em laranja.

La - la, mto-to la - la, mto-to la - la, mto-to la - la. Ma-ma a-na-ku - ja,  
 6 la - la a - ku - pe - ma - zi - wa, la - la.

A falta de acidentes na melodia poderia sugerir a tonalidade de Lá menor, mas o fato de a melodia começar e terminar com Dó, bem como os finais de frase também terminados em dó (compassos quatro e seis) fazem com que seja provável perceber essa melodia como sendo em Dó maior. A melodia apresenta oito intervalos de terça, metade de terça maior e metade de terça menor. A percepção dos informantes no que se refere à gravação original está ilustrada no gráfico 9.

**Gráfico 9** - Percepção de sentimentos expressos na versão original de *Lala, mtoto lala*, em que há uma proporção igual de intervalos de terça maior e de terça menor.



Como se pode observar, quase metade (48%) dos informantes interpretou a melodia como expressando Alegria, enquanto apenas 18% julgaram a emoção como Tristeza. Se apenas os intervalos de terça maior ou menor fossem determinantes na interpretação da carga emocional de uma melodia, seria de se esperar que a proporção de Alegria para Tristeza fosse mais próxima de 1:1, enquanto aqui observamos uma proporção de 2,6:1. Mais uma vez, a percepção da tonalidade como sendo maior (e, portanto, “Não-menor”) parece ter sido mais determinante na interpretação dos ouvintes do que o número de ocorrências de terças maiores ou menores.

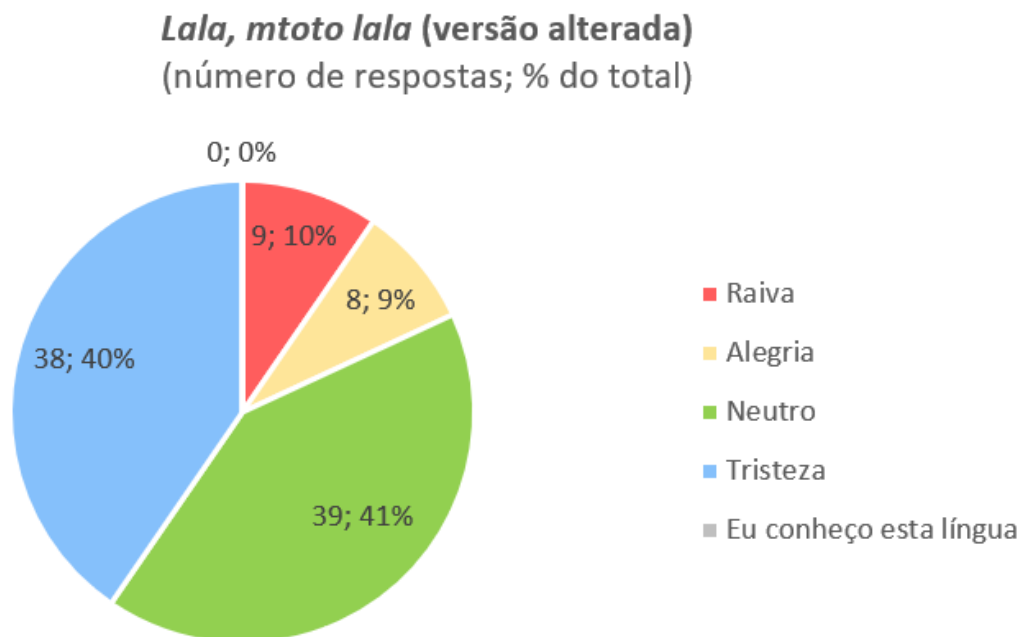
Para a versão alterada, optamos por transformar todos os intervalos em menores, mantendo dois dos intervalos de 3m já presentes na melodia (nos compassos dois e sete), transformando as notas Si e Lá, bem como quase todos os Mi em bemol e fazendo pequenos ajustes quando necessário (ver compassos 2 e 3). Obtivemos, assim, uma melodia que se encaixa, ainda que não perfeitamente, no tom de Dó menor.

**Figura 16** - Partitura da versão alterada de *Lala, mtoto lala*. Os intervalos destacados em azul são intervalos de terça menor.

La - la, mto-to la - la, mto-to la - la, mto-to la - la. Ma-ma a-na-ku - ja,

6  
la - la a - ku - pe - ma - zi - wa, la - la.

**Gráfico 10** - Percepção de sentimentos expressos na versão alterada de *Lala, mtoto lala*, em que todos os intervalos de terça menores.



As alterações digitais nas gravações fizeram com que a porcentagem de pessoas que interpretaram a gravação como expressando Alegria caísse drasticamente, enquanto a percepção de Tristeza teve um crescimento bastante acentuado. Nesta canção, também se verificou alta significância estatística na diferença de percepção de alegria e tristeza ( $p=0,0000$  e  $p= 0,0005$  respectivamente). Esse caso específico parece configurar uma correlação entre a presença de intervalos de terça menor e a percepção de tristeza. Porém, diante do que viemos observando até agora, parece mais razoável que essa diferença na percepção se deva à interpretação da tonalidade da melodia, e não (necessariamente apenas) de intervalos específicos.

### ***Tuu, tuu, tupakarulla***

Melodia oriunda da Finlândia, com letra em finlandês, *Tuu, Tuu, Tupakarulla*<sup>58</sup> apresenta dois intervalos de terça menor e três de terça maior. Apesar de não ouvirmos nenhum Si, a única nota que seria modificada pela armadura de chave correspondente a Fá maior ou a Ré menor, o fato de a melodia começar e terminar em uma nota Ré, bem como a presença de

<sup>58</sup> Tradução livre, com base na tradução para o inglês disponível no site original: Tuu, tuu, rolo de tabaco / como você chegou até aqui? Eu vim pela estrada de Turku / a trilha dos bois do povo Häme.

Ré em uma nota longa de final de frase (compasso 10) amparam a classificação dessa melodia como estando em Ré menor.

**Figura 17** - Partitura de *Tuu, tuu, tupakarulla*. Os intervalos de terça menor aparecem destacados em azul, e os intervalos de terça maior aparecem destacados em laranja.

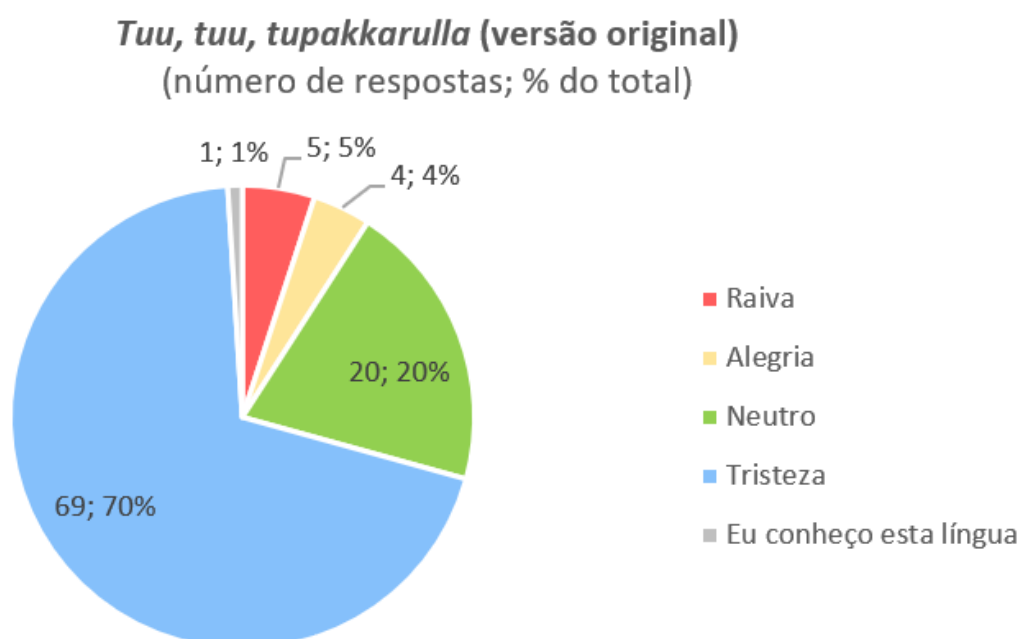
Tuu, tuu, tu-pak - ka - rul - la, mis-täs tie-sit tän ne tul - la?

11 Tu-lin pit-kin Tu-run tie - tä hä - mä - lis-ten här-kä - tie - tä.

Mais uma vez, a tonalidade, e não a proporção de terças maiores e menores, parece estar correlacionada com a percepção dos informantes acerca da emoção expressa. Ainda que metade dos intervalos de terça presentes na melodia sejam maiores, apenas 4% dos informantes interpretaram a canção como sendo alegre.

Considerando que a melodia de *Tuu, tuu, tupakarulla* começa e termina em Ré, as alterações na segunda versão se basearam na escala de Ré maior, a qual apresenta Dó e Fá sustenidos, com todos os intervalos Ré-Fá consistindo de uma terça maior.

**Gráfico 11** - Percepção de sentimentos expressos na versão original de *Tuu, tuu, tupakarulla*, em que os intervalos de terça são predominantemente maiores.



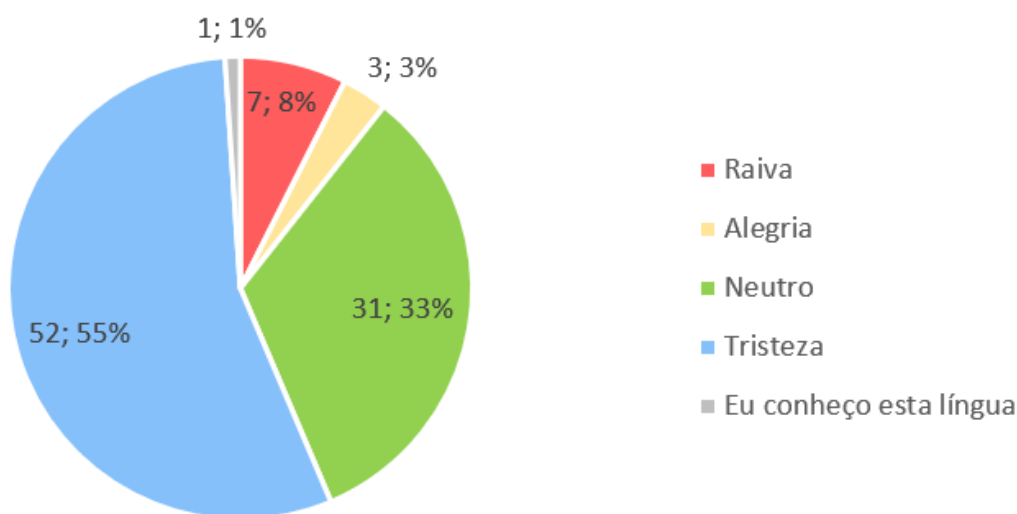
Quando expostos à versão alterada da canção, o número de informantes categorizando *Tuu, tuu, tupakarulla* como triste cai (apesar de continuar alto), a percepção de Raiva aumenta levemente, a de Alegria se mantém praticamente a mesma, e a de neutralidade aumenta. No entanto, apenas a variação de Neutralidade apresentou significância estatística ( $p=0,0289$ ), não sendo possível atribuir com um mínimo de segurança a variação na percepção das outras emoções à alteração dos intervalos ( $p = 0,5512$  para Alegria,  $p = 0,3556$  para Tristeza e  $p = 0,3409$  para Raiva).

**Figura 18** - Partitura da versão alterada de *Tuu, tuu, tupakarulla*. Destacado em azul, o intervalo de terça menor, e em laranja, os intervalos de terça maior.

The musical score is in G major (one sharp) and 8/8 time. The lyrics are: Tuu, tuu, tu-pak - ka - rul - la, mis-täs tie-sit tän ne tul - la? (line 1) and Tu-lin pit-kin Tu-run tie - tä hä - mä - lis-ten här-kä - tie - tä. (line 2). Intervals are highlighted: orange boxes around 'tu-pak - ka - rul - la', 'mis-täs', 'tän ne tul - la?', and 'Tu-lin'; a blue box around 'tie-sit'.

**Gráfico 12** - Percepção de sentimentos expressos na versão original de *Tuu, tuu, tupakarulla*, em que os intervalos de terça são predominantemente maiores.

***Tuu, tuu, tupakkarulla* (versão alterada)**  
(número de respostas; % do total)



### 3.5 Discussão

Como vimos, a presença de intervalos de terça maior e/ou menor pode alterar drasticamente a percepção de um enunciado cantado, mas não de forma tão simples como supomos a princípio. Em vez de uma correlação direta entre intervalos de terça menor e tristeza, o que se observou foi uma associação entre *tonalidades* (modos) menores e a percepção de tristeza. A presença recorrente de intervalos de terça maior ou menor, no *corpus* deste estudo e para os informantes que responderam ao questionário, não foi diretamente determinante na percepção dos informantes, mas sim para reforçar ou alterar a percepção de *modo* e *tom* da melodia – esta, sim, determinante. Melodias que evocavam uma tonalidade/modo maior foram consideradas, no geral, mais alegres do que melodias que evocavam uma tonalidade/modo menor.

Talvez essas conclusões possam parecer óbvias para alguns músicos, mas consideramos os resultados pertinentes aos demonstrarem certa fragilidade dos pressupostos que alicerçam alguns dos trabalhos empíricos que mencionamos anteriormente. Curtis e Bharucha (2010) partem da ideia de que o intervalo de terça menor é associado à expressão de tristeza na música, mas tal pressuposição é confrontada com a versão alterada de *Itsuki-no Komori-uta*, seus intervalos de terça menor e a associação da melodia com a emoção alegria por parte dos informantes. Schreuder, van Eerten e Gilbers (2004) se baseiam na associação do *modo* menor em música com tristeza, e a presença de *intervalos* de terça menor com o modo menor, para interpretar a presença de intervalos menores em contornos prosódicos estilizados como evidência de que *falamos em modo menor quando estamos tristes*. Ora, está empiricamente bastante claro, agora, que as associações entre intervalo de terça menor e modo menor são relativas.

Talvez a perspectiva de Bowling *et al.* (2012), associando a expressão de afetos negativos a intervalos menores em geral, e não à terça menor especificamente, esteja mais próxima da realidade. Cespedes-Guevara e Eerola (2018) defendem uma relativização dessas ideias, postulando que o que a música expressa não são exatamente emoções, mas duas dimensões de afetos (valência e ativação), as quais integram, mas não definem as emoções humanas. Defendem também que a capacidade de identificar emoções na música não advém diretamente da existência de limites categóricos no estímulo musical, mas de processos que facilitam a percepção categórica de emoções. Em ambos os casos, são relativizações razoáveis e que devem ser levadas em consideração na interpretação dos processos cognitivos, mas que

frequentemente não são observadas nos estudos interdisciplinares. Ainda que este não tenha sido um dos nossos objetivos iniciais, os resultados dão peso à necessidade de relativização.

Os resultados deste experimento também mostram a relevância de um aspecto nem sempre devidamente explicitado na literatura, mas que quando ignorado pode levar a resultados inesperados, como aconteceu em nosso experimento: em uma melodia, nem todos os intervalos de terça menor são iguais. Os intervalos de terça decisivos para estabelecer o modo (maior ou menor) de uma melodia não são os estabelecidos entre uma nota qualquer e a nota adjacente, mas sim os estabelecidos entre uma nota da melodia e a nota tônica da melodia em questão. Prova disto é encontrada na Tabela 1, adaptada de Bowling *et al.* (2010):

**Tabela 1** - Frequência de intervalos de terça maior e menor em melodias eruditas e folclóricas.

	<b>Melodias eruditas em modo maior</b>	<b>Melodias folclóricas em modo maior</b>	<b>Melodias eruditas em modo menor</b>	<b>Melodias folclóricas em modo menor</b>
<b>Intervalos tônicos</b> (definidos a partir do número de semitons entre uma nota da melodia e a tônica)				
<b>Terça maior</b>	18.2	16.8	0.7	0.0
<b>Terça menor</b>	0.8	0.0	15.8	15.6
<b>Intervalos melódicos</b> (definidos a partir da distância, em semitons, entre notas adjacentes na melodia)				
<b>Terça maior</b>	6.9	8.6	5.4	6.2
<b>Terça menor</b>	9.4	11.6	10.4	12.1

Fonte: Adaptado de Bowling *et al.* (2010, p. 496).

Pode-se observar que, seja em melodias eruditas ou em melodias folclóricas, os intervalos de terça maior formados entre notas adjacentes são menos frequentes do que os intervalos de terça menor, independentemente do modo dessas melodias. A diferença nas proporções se encontra nos intervalos de terça estabelecidos entre uma nota da melodia e a nota tônica. Isto explica por que, em nosso experimento, um aumento no número dos intervalos melódicos de terça menor não acarretou necessariamente maior percepção de tristeza.

É possível que a diferença de importância entre os tipos de intervalos de terça maior e de terça menor não surpreenda teóricos da música e músicos mais experientes. Porém, esta distinção não é observada de forma rigorosa em todos os estudos consultados durante a preparação do experimento desta tese, principalmente aqueles que buscaram verificar correspondências de intervalos na fala e na música.

Schreuder, van Eerten e Gilbers (2004), por exemplo, estilizaram o contorno melódico da fala de voluntários lendo a fala de personagens literários expressando tristeza e alegria, e o

transpuseram para notação musical, para então verificar quais intervalos foram mais recorrentes nas gravações. No artigo, os pesquisadores afirmam classificar o “modo da fala” de acordo com a distância entre as duas notas adjacentes mais comuns observadas nas amostras. Os autores afirmam que a presença de intervalos de terça maior na fala do personagem alegre configura a fala como “fala em modo maior”, enquanto a presença de intervalos de terça menor na fala do personagem triste consistiria em uma “fala em modo menor”. Porém, os intervalos observados pelos três pesquisadores corresponderiam aos intervalos melódicos na tabela adaptada de Bowling *et al.* (2010) e, como vimos, não é a frequência dos intervalos melódicos que apresenta discrepância entre melodias em modo maior e menor. Curtis e Bharucha (2010) sequer mencionam a questão da nota tônica em sua definição de intervalo menor, o que nos leva a deduzir que estão se referindo à noção de intervalo melódico. Em suma: os pesquisadores associam *intervalos* melódicos de terça menor ao *modo* menor e, conseqüentemente, à expressão de tristeza. É um salto aparentemente pequeno, mas que dá margem a erros.

É claro que o artigo de Bowling *et al.* (2010) foi publicado depois do estudo de Schreuder, van Eerten e Gilbers (2004), e, possivelmente, também depois de Curtis e Bharucha (2010). As datas de publicação, no entanto, não podem ser consideradas um alibi para a imprecisão: Bowling *et al.* (2010) afirmam claramente que seus resultados corroboram dados anteriores, obtidos por Vos e Troost (1989). Curiosamente, não é comum que se mencione o trabalho de Vos e Troost nas referências bibliográficas dos artigos mais recentes comparando música e fala, tampouco a diferenciação entre intervalos melódicos e tônicos explicitada por Bowling *et al.* (2010). Cespedes-Guevara e Eerola (2018), Mantell e Pfordresher (2013), Park *et al.* (2015) e Quinto, Thompson e Keating (2013) são alguns dos estudos a aproximarem fala e música (no que se refere à expressão e percepção de emoções) sem, contudo, abordar a importância da nota tônica nos intervalos de terça menor para o estabelecimento do modo e da carga emocional da melodia. Vos e Troost (1989) não são citados em nenhum destes artigos.

### 3.6 Considerações finais

Os resultados de nosso segundo experimento mostram claramente que uma associação simples e direta entre intervalos melódicos de terça menor e expressão ou percepção de tristeza é equivocada. Assim como na língua, na música também os elementos funcionam dentro das regras<sup>59</sup> de um sistema, e tudo deve ser observado dentro de um contexto.

---

<sup>59</sup> Não confundir, aqui, *regra* com *norma*.

Evidentemente, trabalhar na vanguarda e na interdisciplinaridade é um grande desafio, e não estamos menosprezando os estudos mencionados cujos resultados nossos experimentos não corroboraram. Ao mesmo tempo, é preciso reconhecer que acadêmicos da área de psicologia, neurociência e de linguística, sozinhos, nem sempre possuem todo o conhecimento necessário para desenvolver pesquisas interdisciplinares tão complexas abordando aspectos de composição, categorização e teoria musical. Tal afirmação não deve ser lida como um libelo: trata-se, antes, de um convite para que, no futuro, mais acadêmicos da área de música contribuam com pesquisadores de outras áreas na supervisão de seus experimentos e na interpretação dos dados obtidos.

Devemos, agora, seguir adiante. O terceiro e último experimento deste trabalho será relatado a seguir.

## 4. EXPERIMENTO 3: QUÃO MELODIOSA ESSA LÍNGUA SOA?

### 4.1 Introdução

O senso comum frequentemente caracteriza os idiomas do mundo como mais ou menos musicais. Textos escritos por leigos ou entusiastas na internet abordando os idiomas mais bonitos do mundo frequentemente trazem metáforas musicais em abundância: o idioma francês supostamente tem uma entoação “melodiosa” e é uma língua que soa extremamente musical para os ouvidos não-nativos (ANTOLIK, 2020), o italiano falado “soa como uma canção” para os não-nativos e o finlandês apresenta uma “cadência” adorável e um *staccato* sutil (SEELINGER, 2018).

Antes de nos aprofundarmos neste tema, é importante deixar claro que “melodiosa” e “musical”, a princípio, não são sinônimos, *stricto sensu*. Ainda que tradição ocidental normalmente divida a música entre melodia, harmonia e ritmo, mesmo uma observação pouco cuidadosa das manifestações musicais ao redor do mundo revela como essa divisão pode ser, ocasionalmente, bastante frágil. Uma apresentação de taiko certamente é musical, mas não apresenta melodia. Uma composição atonal, cantada “a capella”, apresenta melodia, mas certamente seria percebida como pouco musical (ou mesmo como “não sendo música”) para muitos ouvintes. A questão fica ainda mais complexa quando levamos em conta gêneros muito populares e difundidos atualmente, como o rap e o hip hop: os vocais desse gênero seriam uma forma de canto ou uma declamação ritmada de poesia? E o resultado final, seria uma fala com acompanhamento ou uma forma de música? Não nos aprofundaremos nessa discussão. Porém, para uma discussão sólida sobre os fenômenos prosódicos e musicais, é importante saber que melodia e música não são conceitos equivalentes.

De acordo com London (2001, n.p.) ritmo e altura melódica (em inglês, *pitch*) são os dois parâmetros primários da estrutura musical. “Melodia”, por definição, se refere a sons de altura definida ordenados no tempo (RINGER, 2001, n.p.), enquanto “ritmo” se refere a padrões de duração<sup>60</sup>. Se aceitarmos essas definições, podemos aceitar sem maiores problemas que nem todo padrão rítmico possui uma melodia. Considerando que nem todo som tem uma frequência fundamental distinguível, é perfeitamente possível conceber uma sequência rítmica expressa, por exemplo, por uma sucessão de chiados com durações diferentes, estabelecendo um padrão

---

<sup>60</sup> A distinção entre ritmo e melodia é comparável à distinção entre ritmo e entoação, como já vimos anteriormente, na seção 1.2.

de durações regular. Por outro lado, se observarmos os conceitos com atenção, veremos que conceber uma melodia sem ritmo não é uma tarefa tão simples. Observemos o trecho a seguir:

Podemos ter peças musicais “desprovidas” de ritmo, constituídas de notas longas, cujas alturas se fundem em fluxo contínuo. Mas exatamente o que marcaria a particularidade de uma tal peça é o fato de “não ter ritmo”; ou seja, a referência à ausência de algo dá a dimensão de quanto este algo é importante. (MEDEIROS, 2010, p. 51)

É difícil ter certeza se, neste trecho de Medeiros, a autora usa as aspas para mostrar que não há um termo consagrado para se referir a peças musicais desprovidas de ritmo, ou se as usa para relativizar a ideia de peças desprovidas de ritmo, implicando que o elemento rítmico ainda está presente nessas peças, pela sua ausência, mesmo que não seja percebido. Neste sentido, acreditamos ser bastante razoável a distinção posta por London (2001, n.p.), segundo o qual "ritmo" se refere a padrões de duração, enquanto o adjetivo "rítmico" implicaria a noção de um padrão de duração mais ou menos regular. Ou seja, havendo durações, há ritmo, o que não significa necessariamente que essas durações sejam percebidas como um padrão rítmico. Seguindo essa lógica, uma peça como a composição hipotética descrita por Medeiros seria, sim, dotada de ritmo, mesmo que os ouvintes não a percebessem como uma peça rítmica. Consequentemente, em última análise, não existe melodia sem ritmo, já que os sons sequenciais de uma melodia, necessariamente, apresentam uma duração no tempo. Por outro lado, uma peça executada unicamente por instrumentos percussivos, sem alturas definidas, pode ser considerada uma peça sem melodia.

Tendo em mente os fatos elencados até aqui, está claro que a música pode conter uma melodia, mas não é delimitada pela presença de uma melodia e nem se confunde com ela. Ao mesmo tempo, para a grande maioria da população leiga, problematizar a equivalência entre *melodiosa* e *musical* exigira uma dose de abstração da qual, muitas vezes, as pessoas não lançam mão. Além disso, não é raro empregar o adjetivo *melodioso* não para se referir a algo que contém ou é similar a uma melodia, mas para caracterizar um estímulo como sendo harmonioso, agradável e eufônico. Assim sendo, optamos por utilizar as palavras *musical* e *melodiosa* como sinônimos intercambiáveis durante a coleta de dados. Fazemos isso cientes da imprecisão teórica e com consciência de que a concepção de *musical* e *melodiosa* varia bastante, a depender do conhecimento técnico, teórico, linguístico e do próprio repertório de cada um. Porém, acreditamos que, se o intuito é verificar se há algo palpável e concreto por trás de uma ideia do senso comum, é necessário utilizar a linguagem do senso comum, ao menos na interação com informantes de formações heterogêneas.

Em uma tentativa de procurar parâmetros concretos para uma avaliação primordialmente subjetiva, esses artigos na internet elencam hipóteses pelas quais um idioma seria mais “musical” do que outro. Especula-se que essa percepção poderia decorrer do fato de uma língua apresentar, como ritmo tipológico básico, o ritmo silábico (MARASCO, 2016). Outra hipótese levantada seria a percepção de maior musicalidade estar condicionada a um sistema fonológico rico em sons vocálicos, ou condicionada à menor frequência de *clusters* consonantais<sup>61</sup> em posição de coda silábica (ERIKSSON, 2020). O uso de termos técnicos nos textos mencionados pode até disfarçar, mas não consegue esconder totalmente o fato de que essas supostas explicações costumam ter pouco ou nenhum rigor científico. Se a forma graciosa como *buongiorno* flui da boca dos falantes se deve à presença de vogais longas e poucos *clusters* consonantais no italiano (ERIKSSON, 2020), não seria natural que o japonês, cujo padrão silábico é ainda mais restritivo no que se refere à combinação ou simples presença de consoantes em ataque ou *coda* silábica, fosse considerado tão fluido, musical e belo quanto a língua oficial da Itália? Um texto no site Mental Floss, *What Makes French Sound Sexy?* (MARASCO, 2016), chega a apontar a presença de fricativas como desencadeadora de mudança na qualidade de voz dos francófonos, tornando as vozes dos falantes mais rouca e aspirada, e tal fator, combinado com as vogais arredondadas, conferiria apelo sexual inerente aos falantes de francês (MARASCO, 2016)<sup>62</sup>.

Em que pese a existência de um corpo de pesquisas sólidas acerca das atitudes linguísticas de vários grupos (ver, como exemplo, CARRIE, 2017; e SCHÜPPERT; HILTON; GOOSKENS, 2015), parece haver uma carência de conhecimento científico no que se refere especificamente a essas avaliações acerca de musicalidade. De fato, no levantamento bibliográfico que embasou o trabalho aqui apresentado, não conseguimos encontrar qualquer trabalho que explicasse, sistematizasse ou simplesmente aferisse a ocorrências destes embasamentos. Isto é bastante curioso, já que, como vimos, o uso de termos como “musical” e “melodioso” para se referir ao som das línguas naturais é bastante difundido, mesmo entre as pessoas que não tiveram qualquer tipo de educação musical.

Talvez o mais próximo que chegamos de uma explicação para esse fenômeno até o momento seja o trabalho de Reiterer et al. (2020). Nele, os pesquisadores notaram que parece haver uma correlação entre a percepção de musicalidade e a estrutura silábica, sendo que

---

<sup>61</sup> Como explica Crystal (2008, p. 81), o termo *cluster* se refere a qualquer sequência de consoantes adjacentes, mas em especial aquelas que ocorrem em posição inicial ou final dentro de uma sílaba, como as sequências /br/ em Brás ou /rfst/na palavra alemã Herfst .

<sup>62</sup> Convenientemente, o texto não menciona que consoantes fricativas e vogais arredondadas não são exclusividade da língua francesa, sendo características também, por exemplo, do alemão e do holandês.

línguas com maior ocorrência de sílabas abertas (ou seja, terminadas em vogal) e maior proporção de vogais na fala seriam consideradas mais musicais.

No entanto, ao contrário do que ocorre para as correlações entre capacidades musicais e habilidades fonológicas (ver, por exemplo, Degé e Schwarzer, 2011; Milovanov e Tervaniemi, 2011; Christiner e Reiterer, 2015), as evidências para uma possível causalidade entre estrutura silábica e musicalidade são, na melhor das hipóteses, inferenciais. Ao mencionar essa potencial correlação, Reiterer e os demais autores apresentam como evidência uma afirmação descontextualizada atribuída à professora Patti Adank, da University College London, e a presença frequente de comentários associando línguas como o italiano a uma suposta musicalidade em fóruns de discussão na internet - cabe deixar claro, aqui, que Reiterer *et al.* (2020) não têm pretensão de que tais evidências sejam fortes o suficiente para sustentar uma explicação científica, mas chama a atenção o fato de que mesmo quatro autores baseados em quatro países diferentes não conseguiram encontrar bibliografia científica que abordasse tal percepção difundida de forma empírica e adequada.

Ainda no artigo de 2020, Reiterer *et al.* analisaram as atitudes em relação a dezesseis idiomas (a maioria esmagadora de origem indo-europeia) por meio de 22 escalas semânticas bipolares, inclusive o quão melodiosa a língua soava para os participantes. Infelizmente, os resultados referentes à etiqueta semântica “melodia” não são destrinchados no artigo, e apenas somos informados que as línguas latinas foram percebidas como mais melodiosas do que as línguas eslavas e germânicas (REITERER *et al.*, 2020, p. 192). Mesmo apresentados de forma vaga, os resultados parecem ir ao encontro das evidências anedóticas e percepções subjetivas mencionadas anteriormente.

No entanto, a questão não é tão simples. O trabalho de Reiterer *et al.* (2020) reflete uma percepção majoritariamente ocidental: dos 45 participantes, 38 eram europeus, em contraste com três norte-americanos e apenas um brasileiro e um asiático. Ainda que assumíssemos que uma maior proporção de sons vocálicos nos enunciados falados e uma maior ocorrência de sílabas abertas fossem correlatos de uma maior percepção de musicalidade, ainda restaria explicar o porquê de tais características serem associadas, principalmente, a línguas latinas, notadamente o italiano, o francês e, até certo ponto, o espanhol. A fonologia do japonês, por exemplo, proíbe consoantes em posição de coda silábica (ou seja, não permite consoantes depois da vogal que ocupa o núcleo silábico), e mesmo assim não são encontradas muitas mensagens em fóruns de discussão sobre como o japonês soa “musical”. Seria possível que línguas indígenas, africanas ou asiáticas que tenham sistemas fonológicos com características semelhantes às línguas românicas também seriam percebidas como musicais?

Também é importante mencionar que, no que se refere às atitudes acerca de um idioma, a noção de *musicalidade* é frequentemente associada a, ou mesmo usada como um sinônimo de beleza e agradabilidade. O estudo de Reiterer *et al.* (2020) deixa essa associação clara no índice que chamaram de eroticismo, o quão “atraente” e “agradável” uma língua parece ao ouvinte. No estudo, este índice foi associado com ritmo silábico, alta proporção de vogais em relação ao número de fonemas e maior índice de sonoridade. É importante ter isso em mente: ainda que estejamos pensando na suposta musicalidade de um idioma, essa noção está associada, em algum nível, à “beleza” de uma língua, de modo que não seria necessariamente uma temeridade ou um salto demasiado grande usar “musical” e “melodiosa” como sinônimos de “bonita” e “agradável”. Um exemplo claro do campo semântico relativo à musicalidade sendo um sinal de beleza em oposição ao feio e desagradável pode ser encontrado em um texto publicado no blog da agência de tradução BeTranslated: *“Why is it that certain languages sound poetic and melodious while others grate on our nerves?”* (WHAT, 2017, n.p.), em tradução livre: “por que é que algumas línguas soam poéticas e melodiosas enquanto outras nos dão nos nervos?”<sup>63</sup>

Assim sendo, o que faz, afinal, uma língua soar como música para nossos ouvidos? Ainda não há uma resposta definitiva. No intuito de levantar evidências que possam ajudar a lançar luz sobre esse fenômeno, o terceiro e último experimento deste trabalho se debruça sobre essa questão, na tentativa de verificar se há algum fator linguístico ou acústico acarretando tal percepção, ou se a diferença na suposta musicalidade dos idiomas é explicada puramente por questões afetivas e particulares de quem ouve, averiguando a percepção do quão musical alguns idiomas soam para pessoas que não são falantes desses idiomas. A lógica por trás da hipótese será esmiuçada a seguir.

No senso comum, a noção de canto está intimamente ligada à noção de melodia. Quando alguém canta uma canção, muito provavelmente canta a linha melódica em vez de imitar os padrões rítmicos de um instrumento de percussão. De fato, essa ligação entre canto e melodia é forte a ponto de definir a modalidade “canto”, pelo menos na percepção da maioria das pessoas. Tal afirmação pode ser facilmente confirmada se pensarmos em um dos gêneros musicais mais populares na atualidade, o rap. Ao contrário do que ocorre em virtualmente todas as outras manifestações da música popular (aqui entendida como oposição a erudita), um dos aspectos que definem o gênero rap é justamente a tendência deste gênero musical de alturas

---

<sup>63</sup> Tal metáfora ignora, é claro, que nem toda melodia e nem toda poesia nos soa bela e agradável.

melódicas definidas<sup>64</sup>, o que faz com que os vocais no rap sejam percebidos não como canto, mas como prosa rimada e enunciada ritmicamente (KOMANIECKI, 2019, p. 108). O papel de destaque da melodia na percepção de musicalidade poderia certamente ser questionado e problematizado, engendrando uma discussão que, no entanto, não nos é relevante no momento. Para nossos propósitos, basta ter em mente que a noção de musicalidade está mais associada à noção de melodia do que de ritmo.

Mais a fundo na questão, Kolinsky et al. (2009, p. 15), em um artigo com o subtítulo sugestivo *Vowels sing but consonants speak*<sup>65</sup>, verificaram que as consoantes são processadas de maneira mais independentemente da informação melódica do que as vogais. Em outras palavras, há maior interferência da informação melódica sobre o processamento da informação vocálica do que sobre o processamento das informações consonantais. Também cabe mencionar a importância da frequência fundamental ( $f_0$ ) enquanto parâmetro acústico na diferenciação entre fala e canto (MEDEIROS; CABRAL, 2018, p. 542). Mais especificamente, trata-se da percepção mais ou menos consensual de que a saliência perceptiva das alturas melódicas bem-definidas caracterizaria um enunciado como cantado, em oposição aos contornos melódicos definidos de forma mais vagamente da fala (DEUTSCH; HENTHORN; LAPIDIS, 2011, p. 2245).

Tendo em mente todos os fatores elencados acima, parece razoável supor que um idioma será percebido como mais melodioso / musical quanto maior a proporção de sons vocálicos no enunciado e quanto maior seu índice de sonoridade. Esta é a nossa hipótese<sup>66</sup>.

## 4.2 Conceitos prévios

A articulação de sons consonantais é classificada comumente em função de três parâmetros: vozeamento, modo de articulação e local de articulação (CAGLIARI, 1981, p. 17). No que se refere ao local de articulação, as consoantes são classificadas com o nome do lugar onde ocorre o maior fechamento dos articuladores obstruindo a corrente de ar - por exemplo, se o estreitamento do canal do aparelho fonador ocorrer na região do alvéolo, o som será chamado de “alveolar” (CAGLIARI, 1981, p. 22).

---

<sup>64</sup> Evidentemente, tal afirmação é uma generalização, e não significa que não existam ocorrências de rappers modulando a altura melódica de suas vozes intencionalmente, usando uma variedade de técnicas (KOMANIECKI, 2019, p. 108).

<sup>65</sup> “As vogais cantam, mas as consoantes falam”, em tradução livre,

<sup>66</sup> Como veremos mais adiante, os dados colhidos tornaram clara a importância dos aspectos relacionados ao contorno da altura melódica no que se refere à percepção de musicalidade, pelo menos no *corpus* em questão.

Consoantes plosivas, também chamadas de oclusivas (MASSINI-CAGLIARI, CAGLIARI, 2001, p. 122), são aquelas em que os articuladores produzem uma obstrução completa da passagem da corrente de ar através da boca, com o véu palatino levantado e o ar oriundo dos pulmões encaminhando-se para a cavidade oral (SILVA, 2003, p. 34).

As consoantes nasais são aquelas em que os articuladores produzem uma obstrução completa da passagem da corrente de ar através da boca, com o véu palatino abaixado e o ar oriundo dos pulmões dirigindo-se simultaneamente às cavidades nasal e oral (SILVA, 2003, p. 34).

Durante a produção das consoantes fricativas, os articuladores se aproximam, produzindo fricção quando ocorre a passagem central da corrente de ar – sem, no entanto, causar obstrução completa do fluxo, mas sim parcial (SILVA, 2003, p. 34).

Depois dessas considerações sobre segmentos consonantais, passemos a um nível superior da organização fonológica. Os segmentos se juntam em sílabas, e é a partir do nível silábico que podemos perceber as relações de proeminências (que, por sua vez, serão a base para a emergência do ritmo, sobre o qual falaremos em breve). É a essas relações que nos referimos quando usamos termos como “tonicidade” e “acento”. Quando, no ensino fundamental, os alunos aprendem a identificar a sílaba tônica de uma palavra, sendo orientados a perceber qual das sílabas é a mais “forte”, estão basicamente tentando assimilar conscientemente uma relação de proeminência.

Os modelos fonológicos mais recentes (não-lineares) têm definido “acento” como uma relação de *proeminência* entre sílabas: as mais proeminentes são as tônicas ou acentuadas e as menos proeminentes, as átonas. Essas proeminências fonológicas são atualizadas no nível fonético (acústico, articulatório) de maneiras diferentes, em cada língua, constituindo uma de suas características próprias. (MASSINI-CAGLIARI, 1992, p. 9)

No caso do Português, as sílabas tônicas se diferem das átonas em termos de produção, mas também acusticamente, em termos de percepção, como explica Silva (2003):

Uma sílaba tônica ou acentuada é produzida com um pulso torácico reforçado. Portanto, na produção de uma sílaba acentuada temos um jato de ar mais forte (em relação às sílabas não acentuadas ou átonas). A vogal acentuada é auditivamente percebida como tendo duração mais longa e também como sendo pronunciada de maneira mais alta (no sentido de falar alto). Este aumento de volume permite-nos identificar as vogais acentuadas das vogais não acentuadas - que são pronunciadas com o volume mais baixo e portanto percebidas auditivamente de maneira distinta. (SILVA, 2003, p. 77)

Vale ressaltar que a tonicidade não é necessariamente um fenômeno binário, “sim ou não”, “tônica ou átona”. No português, por exemplo, uma sílaba pode carregar o acento

primário, o acento secundário ou ser totalmente isenta de acento (SILVA, 2003, p. 77). Isto pode ser facilmente notado quando pronunciamos palavras longas, como “paralelepípedo” em que a sílaba “pa” e a primeira ocorrência de “le” são claramente mais “fortes” do que “ra” e a segunda ocorrência de “le”, mas menos “fortes” do que “pi”.

Por fim, a noção de ritmo é, de certa forma, decorrente do acento, pois se refere à maneira como as línguas organizam no tempo os elementos salientes da fala (MASSINI-CAGLIARI, CAGLIARI, 2001, p. 116), ou seja, o ritmo está ligado às relações entre os acentos (primário e secundário) e a ausência de acentos (SILVA, 2003, p. 77). Não se deve confundir ritmo com taxa de elocução, a qual se refere à quantidade de palavras, sílabas, etc. enunciadas em um espaço de tempo:

[...] um mesmo padrão rítmico pode ser dito com maior ou menor velocidade de fala - assim como uma estrutura musical não perde o ritmo se executada mais rápida ou mais lentamente (variação de andamento: uma valsa, por exemplo, pode ser executada bem lentamente ou com um andamento mais rápido, mas continuará sendo uma valsa). (MASSINI-CAGLIARI, CAGLIARI, 2001, p. 117)

Por fim, lembramos que todas as sílabas da fala são pronunciadas com certa altura melódica (MASSINI-CAGLIARI, CAGLIARI, 2001, p. 117), e que a organização destas alturas melódicas ao longo do tempo faz emergir padrões entoacionais.<sup>67</sup>

### 4.3 Metodologia

As próximas seções abordam os procedimentos metodológicos deste experimento. Esclarecemos os parâmetros e procedimentos que guiaram a seleção e preparação dos estímulos e, posteriormente, a participação dos informantes na coleta de dados propriamente dita.

#### 4.3.1 Estímulos

Mais uma vez, os estímulos para o experimento foram retirados do site *Lullabies of the World* (THE VOICE PROJECT, 2013). Além das partituras e das interpretações musicais das cantigas, o site apresenta arquivos de áudio no formato mp3 em que as letras das cantigas são declamadas. Optamos por usar as cantigas declamadas em vez de outras possíveis amostras de fala nos idiomas por questões de conveniência (poupando tempo de procurar e selecionar estímulos) e normalização (os textos declamados são todos do mesmo gênero, letra de música).

---

<sup>67</sup> Ver seção 1.2.

O próximo passo foi escolher quais línguas utilizar. Para tal, foram considerados os seguintes critérios:

- **qualidade das gravações:** mesmo estando todas disponibilizadas no mesmo site, a qualidade das gravações não é padronizada. Optou-se por não utilizar gravações que apresentassem excesso de eco (como a canção libanesa *Yalla tnam Rima*, em árabe<sup>68</sup>), chiados ou som “abafado”;

- **estilo de declamação:** alguns falantes não apenas declamavam o texto, mas imprimiam às sílabas as mesmas durações das notas musicais (como na canção estoniana *Viire takka*<sup>69</sup>). Considerando que o objetivo é justamente verificar a percepção de musicalidade no idioma, optou-se por não utilizar essas declamações com padrão rítmico muito marcado, pois elas poderiam soar como “mais musicais” do que gravações em que o texto fosse apenas declamado;

- **família linguística:** no intuito de evitar qualquer influência do conteúdo semântico na avaliação dos informantes, foram descartadas as gravações em português e em espanhol (idioma que apresenta alto nível de inteligibilidade com o português). Também evitou-se utilizar línguas mais conhecidas (REITERER *et al.*, 2020, p. 185) e, portanto, mais facilmente identificáveis, como o francês, o italiano e o inglês, para minimizar o risco de que estereótipos e associações culturais influenciassem as avaliações. Uma exceção permitida foi o uso da gravação em alemão, idioma facilmente identificável e relativamente difundido, mas que se mostra relevante neste tipo de estudo devido à sua frequente classificação como feio e pouco melodioso – pelo menos no senso comum;

- **sexo do falante:** pesquisas sobre atitudes linguísticas (SCHÜPPERT; HILTON; GOOSKENS, 2015; CARRIE, 2017) mostraram uma influência clara do sexo do falante sobre as avaliações dos estímulos auditivos. Os motivos pelos quais isso acontece fogem do escopo deste trabalho e podem ser encontrados nos artigos aqui citados. Para nossos propósitos, basta dizer que um estímulo no mesmo idioma pode elicitar respostas levemente distintas a depender se a pessoa que fala é do sexo feminino ou masculino e que, para controlar as variáveis, os informantes deveriam ser expostos a declamações feitas por falantes do mesmo sexo.

---

<sup>68</sup> Disponível em <<https://www.lullabiesoftheworld.org/lie33194-Yalla-tnam-Rima.html>>, acesso em 02 de março de 2021.

<sup>69</sup> Disponível em <<https://www.lullabiesoftheworld.org/lie33228-Viire-takka.html>>, acesso em 02 de março de 2021.

Considerando todos esses fatores, elencamos as gravações oriundas da Armênia, da Índia e da Islândia (declamadas em vozes femininas), e da Alemanha, da Hungria e da Noruega (declamadas em vozes masculinas).

As declamações foram analisadas *foneticamente*. Acredita-se que uma descrição puramente fonética seja mais adequada para os propósitos deste trabalho, uma vez que o ouvinte que não seja falante do idioma não está familiarizado com o sistema fonológico da língua e, assim, tem acesso apenas ao aspecto fonético do idioma. Evidentemente, uma vez que o ouvinte tente interpretar e/ou reproduzir o estímulo no idioma desconhecido, provavelmente o fará com o filtro do sistema fonológico de sua língua materna e/ou de outras línguas das quais seja falante. No entanto, esta questão trata de um outro aspecto, e não a percepção de uma das características do estímulo sonoro, que é o que pretendemos aferir.

Evidentemente, é pouquíssimo provável que a declamação da letra de uma canção apresente exatamente os mesmos padrões entoacionais verificáveis em uma amostra de fala espontânea. No entanto, acreditamos que tal fato não invalida nossas observações e resultados, uma vez que todos os arquivos de áudio apresentavam declamações. As diferenças observadas na percepção das declamações não devem ser atribuídas ao simples fato de elas serem declamações e não amostras de fala espontânea. Por outro lado, seria possível especular que, estando os declamadores familiarizados com as canções, os contornos melódicos das canções influenciariam, em algum grau, a fonação (por exemplo, músicas com uma tessitura mais ampla levando a maiores variações de  $f_0$  enquanto o falante declama). O quadro 2 aponta a tessitura das canções de acordo com as partituras, informação que será retomada mais adiante na discussão dos resultados.

**Quadro 2** - Tessitura das cantigas de ninar.

<b>Cantiga de ninar</b>	<b>Nota mais grave</b>	<b>Nota mais aguda</b>	<b>Tessitura</b>
Akna Oror	A3	E $\flat$ 4	3 tons
Aludjal el, csucsuljal el	D3	D4	1 oitava
Der Mond ist aufgegangen	F3	D4	5 tons
Gjendines badnlat	B2	B3	1 oitava
Jasoda Hari palne	C3	C4	1 oitava
Sofðu unga ástin mín	D3	B $\flat$ 3	4 tons e 1 semitom

A seguir, faremos uma breve descrição dos estímulos utilizados no questionário *online*. Considerando a tendência, mencionada anteriormente, de vozes masculinas e femininas serem

percebidas de maneira diferente, no questionário da pesquisa os estímulos foram apresentados de modo a alternar vozes masculinas e femininas, para minimizar o risco de *priming*. De acordo com Bargh (2006), *priming* é uma forma de ativação inconsciente de estruturas de conhecimento social, que pode afetar quase todas as formas de representação social e, por consequência, gerar distorções em estudos envolvendo percepção e cognição. Um exemplo, reportado por Srull e Wyer (1979), mostrou que expor participantes de um experimento a palavras relacionadas a "bondade" levou esses participantes a encararem uma pessoa-alvo como mais bondosa. Em outras palavras, o efeito *priming* faz com que a exposição a um estímulo afete a percepção do estímulo seguinte. Nossa opção por alternar entre estímulos com vozes femininas e masculinas consiste em uma tentativa de evitar que as vozes fossem avaliadas "em bloco". Aqui, no entanto, apresentaremos primeiro os estímulos que traziam vozes masculinas, para então passar para os estímulos com vozes femininas. O texto falado no áudio aparece transcrito a seguir, seguido da descrição das estruturas silábicas em termos simples de Consoante e Vogal, bem como eventuais explicações que julgamos serem pertinentes.

### **Gravação nº 01: Der Mond ist aufgegangen**

(alemão, voz masculina)

*Der Mond ist aufgegangen,  
die goldnen Sternlein prangen  
am Himmel hell und klar.  
Der Wald steht schwarz und schweiget,  
und aus den Wiesen steigt  
der weiße Nebel wunderbar.*<sup>70</sup>

O quadro 3 mostra a classificação dos segmentos da declamação em alemão, sendo que abaixo de cada sílaba está discriminada a estrutura silábica percebida foneticamente. Na última coluna, estão discriminados o número de segmentos consonantais (C), vocálicos (V) e a soma dessas duas categorias, resultando no número total de segmentos do verso respectivo.

---

<sup>70</sup> Apresentaremos uma tradução livre em português da letra de cada cantiga, baseadas nas traduções para o inglês, disponibilizadas pelo próprio site. Começamos pela primeira cantiga, *Der Mond ist aufgegangen*: A lua nasceu / as pequenas estrelas douradas brilham / no céu claro e claro / A floresta inerte, escura e silenciosa / e dos prados sobe / a névoa branca, milagrosamente.

**Quadro 3** - Análise silábica de *Der Mond is aufgegangen*.

Células/Segmentos (Vogais = V, Consoantes = C)								Quantidades de segmentos (S): Consoantes (C) e vogais (V)
Der	Mond	ist	auf	ge	ga	ngen		
CVV	CVCC	VCC	VVC	CV	CV	CVC		11C+9V=20s
die	Gold	nen	Stern	lein	pra	ngen		
CV	CVCC	CVC	CCVVC	CVC	CCV	CVC		15C+8V=23s
am	Hi	mmel	hell	und	klar			
VC	CV	CVC	CVC	VCC	CCVC			11C+6V=17s
Der	Wald	steht	schwarz	und	schwei	get		
CVV	CVCC	CCVC	CCVC	VCC	CCVV	CVC		16C+9V=25s
und	Aus	den	Wie	sen	stei	get		
VCC	VVC	CVC	CV	CVC	CCVV	CVC		12C+9V=21s
der	Weiß	ße	Ne	bel	wun	der	bar	
CVV	CVV	CV	CV	CVC	CVC	CV	CVV	10C+11V=21s
Total de segmentos: 127, sendo 75 consonantais (59%) e 53 vocálicos (41%).								

Os sons [ʊ] (como em *aus*) e [j] (como em *weiß*) foram considerados como vogais, ainda que fonologicamente pudessem ser interpretados como consoantes. O dígrafo <eh> em *steht* foi considerado como um único som vocálico. As combinações V+<r> foram interpretadas caso a caso, o que leva a resultados diferentes: na palavra *der*, logo no primeiro verso, a combinação <er> foi interpretada como duas vogais, uma vez que à vogal [ɛ] se segue uma breve vogal [a], como alofone de /r/. Porém, a mesma combinação de letras foi interpretada como uma vogal simples [e] na última palavra, *wunderbar*, pois o falante realiza a combinação <er> apenas como um [e], não alterando a qualidade do som vocálico antes do ataque da consoante seguinte, [b]. Consoantes duplas, como o <ß> em *weiß* e o <mm> em *Himmel* foram consideradas como parte do ataque da sílaba seguinte, haja vista que todas as línguas do mundo apresentam em seus sistemas fonológicos sílabas no padrão CV, mas nem todas permitem a ocorrência de consoante no travamento silábico. Eventuais diferenças entre o esperado pelo sistema fonológico da língua alemã e a análise aqui apresentada se devem à pronúncia na gravação em questão, que pode não ser totalmente representativa dos fenômenos fonológicos do idioma.

A seguir, apresentamos a letra da canção *Aludjal el, csucsuljal el*, oriunda da Hungria.

### Gravação n° 02: Aludjal el, csucsuljal el

(húngaro, voz masculina)

*Aludjál el, csucsuljál el, drága kicsim, aludjál,  
aludj szépen, aludjál, csillagokról álmodjál.*<sup>71</sup>

O quadro 4 mostra a classificação dos segmentos da declamação em húngaro, sendo que abaixo de cada sílaba está discriminada a estrutura silábica percebida foneticamente. Na última coluna, estão discriminados o número de segmentos consonantais (C), vocálicos (V) e a soma dessas duas categorias, resultando no número total de segmentos do verso respectivo.

**Quadro 4 - Análise silábica de *Aludjal el, csucsuljal el*.**

Células/Segmentos (Vogais = V, Consoantes = C)								Quantidades de segmentos: Consoantes (C) e vogais (V)
A	lu	djál	el	csu	csul	jál	el	
v	cv	cvc	vc	cv	cv	vvc	vc	8C+9V=17s
drá	ga	ki	csim	a	lu	djál		
CCV	CV	CV	CVC	V	CV	CVVC		9C+8V=17s
a	ludj	szé	pen	a	lu	djál		
V	CVC	CV	CVC	V	CV	CVVC		8C+8V=16s
csi	lla	go	król	ál	mo	djál		
CV	CV	CV	CCVC	VC	CV	CVVC		10C+8V=18s
Total de segmentos: 68, sendo 35 consonantais (51,47%) e 33 vocálicos (48,52%).								

Passemos, agora, para a letra da próxima canção: a cantiga de ninar *Gjendines badnlat*, da Noruega.

### Gravação n° 03: Gjendines badnlat

(norueguês, voz masculina)

*Barnet legges i vuggen ned,  
stundom græder og stundom ler.  
Sove nu, sove nu, i Jesu navn,  
Jesu bevare barnet.*<sup>72</sup>

<sup>71</sup> Em tradução livre: Vá dormir, meu doce pequenino / vá dormir e sonhe com as estrelas / Cubra-se e durma / Durma, durma, e sonhe com as estrelas.

<sup>72</sup> Em tradução livre: A criança deitada no berço / Ora chora, ora ri / Durma agora, durma agora em nome de Jesus / Jesus protege a criança.

Sendo uma língua de origem germânica, não é surpreendente que o norueguês tenha apresentado uma proporção de consoantes e vogais mais próxima à do alemão e mais distinta do húngaro – fato que pode ser verificado no quadro 5.

**Quadro 5** - Análise silábica de *Gjendines badnlat*.

Células/Segmentos (Vogais = V, Consoantes = C)								Quantidades de segmentos: Consoantes (C) e vogais (V)
Bar	net	leg	ges	i	vug	gen	ned	
cvc	cv	cv	cvc	v	cvc	cvc	cv	11C+8V=19s
stun	dom	græ	der	og	stun	dom	ler	
CVC	CVC	CV	CVC	V	CVC	CVC	CV	12C+8V=20s
So	ve	nu	so	ve	nu			
CV	CV	CV	CV	CV	CV			6C+6V=12s
i	Je	su	navn					
V	CV	CV	CVCC					5C+4V=9s
Je	su	be	va	re	bar	net		
CV	CV	CV	CV	CV	CVC	CV		8C+7V=15s
Total de segmentos: 75, sendo 42 consonantais (56%) e 33 vocálicos (44%).								

Seguimos agora com a letra de *Akna Oror*, cantiga de ninar da Armênia.

#### **Gravação nº 04: Akna Oror**

(armênio, voz feminina)

*Aghvor es, chunis khalat,*

*ertam ov berim pekhalat, oror.*

*Ertam lusynkan berim,*

*lusun astghery pekhalat, oror.*<sup>73</sup>

No quadro 6, podemos observar que, apesar de o armênio ser um idioma bastante distante do alemão, a proporção de consoantes e vogais nas amostras dessas duas línguas foi muito parecida.

<sup>73</sup> Em tradução livre: Você é lindo, sem defeitos / Onde poderia eu encontrar alguém como você? / Deixe-me trazer àquele que é perfeito / estrelas da lua. O segundo verso da cantiga, que não foi usado como estímulo neste estudo e, portanto, não aparece transcrito aqui, termina com o reconhecimento de que a criança, na verdade, tem um único defeito: não dormir.

**Quadro 6** - Análise silábica de *Akna Oror*.

Células/Segmentos (Vogais = V, Consoantes = C)										Quantidades de segmentos: Consoantes (C) e vogais (V)
Agh	vor	es	chu	nis	kha	lat				
VC	CVC	VC	CV	CVC	CV	CVC				10C+7V=17s
er	tam	ov	be	rim	pe	kha	lat	o	ror	
CVC	CVC	VC	CV	CVC	CV	CV	CVC	V	CVC	14C+10V=24s
Er	tam	lu	syn	kan	be	rim				
CVC	CVC	CV	CVC	CVC	CV	CVC				12C+7V=19s
lu	sun	ast	ghe	ry	pe	kha	lat	o	ror	
CV	CVC	VCC	CV	VC	CV	CV	CVC	V	CVC	13C+10V=23s
Total de segmentos: 83, sendo 49 consonantais (59,03%) e 34 vocálicos (40,96%).										

O texto da penúltima gravação, apresentado aqui, é da cantiga de ninar *Jasoda Hari palne*, originária da Índia. De acordo com as notas disponíveis no próprio *site*, a letra faz referência a Jashoda, mãe de criação do deus hindu Hari, também conhecido como Krishna.

#### Gravação nº 5: *Jasoda Hari palne*

(dialecto Hindi, voz feminina)

*Jasodā Hari Pālāne Jhulāvai*

*Halarāvai Dularaī Malhāve*

*Joi Soi Kachu Gāvai*

*Mere Lālako Āvu Nīndariyā*

*Kāheṅa Āni Sulāvai*<sup>74</sup>

Note-se a presença de uma consoante aproximante [v] no final da palavra *dularaī* e de um [a] bastante breve no meio da palavra *Malhāve*, nenhuma das quais é marcada pela ortografia, e a ausência de um segundo [a] na palavra *Lālako* (quadro 7).

<sup>74</sup> Em tradução livre: Yashoda embala Hari em seu berço / Balançando para lá e para cá / Yasoda canta pequenas canções / O sono vem dos céus / Para dentro dos olhos do meu querido.

Quadro 7 - Análise silábica de *Jasoda Hari palne*.

Células/Segmentos (Vogais = V, Consoantes = C)											Quantidades de segmentos: Consoantes (C) e vogais (V)	
Ja	so	dā	Ha	ri	Pā	la	ne	Jhu	lā	vai		
CV	CV	CV	CV	CV	CV	CV	CV	CV	CV	CV		11C+11V=22s
Ha	la	rā	vai	Du	la	ra	ī	Ma	l	hā	ve	
CV	CV	CV	CV	CV	CV	CV	CV	CV	CV	CV	CV	12C+12V=24s
Joi	Soi	Ka	chu	Gā	vai							
CVV	CVV	CV	CV	CV	CV							6C+8V=14s
Me	re	Lā	la	ko	Āvu	Nīn	da	ri	yā			
CV	CV	CV	C	CV	VV	CVC	CV	CV	CV			10C+10V=20s
Kā	he	na	Ā	ni	Su	lā	vai					
CV	CV	CV	C	CV	CV	CV	CV					8C+7V=15s
Total de segmentos: 95, sendo 47 consonantais (49,47%) e 48 vocálicos (50,52%).												

Finalmente, temos a letra da cantiga de ninar *Sofðu unga ástin mín*, proveniente da Islândia.

### Gravação n° 6: *Sofðu unga ástin mín*

(islandês, voz feminina)

*Sofðu unga ástin mín,*

*úti regnið grætur.*

*Mamma geymir gullin þin,*

*gamla leggi' og völuskrín.<sup>75</sup>*

Digno de nota no caso do islandês é o fato de o grafema <á> em *ástin* ser pronunciado como [aʊ], o que pode ser bastante contraintuitivo para quem está acostumado com a ortografia do português. Ao contrário do alemão, aqui os dois <mm> em *Mamma* são interpretados como duas consoantes, uma em posição de coda e outra em posição de ataque silábico. Isso se deve pelo notável alongamento da articulação [m] pela falante, ao contrário do que ocorreu na gravação com o falante de alemão na palavra *Himmel*. Também pode surpreender que o grafema <æ> represente dois sons vocálicos, mas o sistema fonológico islandês não apresenta uma vogal anterior quase aberta não arredondada como no inglês cat. Na ortografia islandesa, a letra <æ> representa um ditongo, [aɪ]. Por fim, no último verso, na sequência *leggi' og*, o <i> é

<sup>75</sup> Em tradução livre: Durma agora, suavemente, meu amorzinho / Lá fora a chuva cai / A mamãe cuida do seu tesouro / sua coleção de ossos e o baú para pedras / Não devemos ficar acordados noites de escuridão adentro.

pronunciado como uma consoante aproximante, [j], não constituindo núcleo silábico (quadro 8).

**Quadro 8** - Análise silábica de *Sofðu unga ástin mín*.

Células/Segmentos (Vogais = V, Consoantes = C)							Quantidades de segmentos: Consoantes (C) e vogais (V)	
Sof	ðu	un	ga	ás	tin	mín		
CVC	CV	VC	CV	VVC	CVC	CVC		10C+8V=18s
ú	ti	reg	nið	græ	tur			
V	CV	CVC	CVC	CCVV	CVC			9C+7V=16s
Mam	ma	gey	mir	gul	lin	þin		
CVC	CV	CVV	CVC	CVC	CV	CVC		11C+8V=19s
gam	la	le	ggi'	og	vö	lus	krín	
CVC	CV	CV	CC	VC	CV	CVC	CCVC	13C+7V=20s
Total de segmentos: 73, sendo 43 consonantais (58,9%) e 30 vocálicos (41,09%).								

#### 4.3.2 Informantes

Os informantes neste experimento com as declamações foram os mesmos participantes do experimento anterior, com as melodias sendo cantadas. Porém, enquanto no experimento anterior os quase 200 informantes foram separados em dois grupos distintos (sendo que cada grupo ouviu uma das duas versões das melodias, mas não ambas), no experimento abordado no presente capítulo não houve divisão.

Em vez disso, todos os informantes foram expostos à mesma seleção de seis trechos das letras das canções sendo declamadas. Tal configuração foi possível porque a plataforma gratuita Google Docs permite que, em um mesmo formulário, sejam criadas seções diferentes. No caso da primeira seção, correspondente ao experimento das melodias, cada grupo respondia a perguntas diferentes. Porém, na segunda seção, ambos os grupos responderam às mesmas perguntas.

No total, foram 192 informantes, cujas idades variavam de 14 a 65 anos (média de 34,1 anos, mediana 32 anos). Esses informantes falavam, em média, 3 línguas (o número de línguas variou entre 1 e 6). Nenhum dos informantes declarou não ser falante de português, seja como língua materna ou língua estrangeira. No tocante a envolvimento ativo com música (cantando, tocando um instrumento, compondo ou qualquer outra atividade do tipo), 55 (28%) afirmaram nunca ter tido envolvimento ativo e não manifestaram desejo de ter; 56 (29%) nunca tiveram envolvimento ativo, mas gostariam de ter; 44 (22,9%) afirmaram já ter tido envolvimento ativo

no passado; e 37 (19%) afirmaram ter envolvimento ativo com música regularmente. Informações sobre sexo e gênero dos informantes não foram coletadas.

#### 4.3.1 Estatística

Esta seção aborda os conceitos estatísticos a serem utilizados na discussão dos dados do terceiro experimento desta tese. Para uma introdução aos critérios que guiaram nossas análises, remetemos à seção 3.3.2.

Um dos testes sobre os quais nós nos voltaremos mais adiante tem o objetivo de verificar se falantes de um idioma (alemão) atribuem a essa língua graus de musicalidade significativamente diferentes dos atribuídos por não-falantes. Os graus de musicalidade atribuídos por nossos informantes a cada língua são os nossos escores, os quais foram mensurados ordinalmente (“1 – Nada melodiosa / musical” é inferior a 2, e assim sucessivamente até “5 – Muito melodiosa / musical”).

Os falantes e os não-falantes do idioma constituem duas amostras independentes (ver seção 3.3.2). No caso, o intuito é descobrir se os escores da amostra “falantes de alemão” são significativamente diferentes dos escores da amostra “não-falantes de alemão”. Para isso, lançamos mão do teste de Mann-Whitney<sup>76</sup>, também chamado de Teste U (AYRES *et al.*, 2007, p. 135 e 348), cujo objetivo é justamente comparar duas amostras independentes, sejam seus tamanhos iguais ou desiguais.

Ainda nesse sentido, podemos observar o quanto os indivíduos que compõem cada uma das amostras concordam entre si, isto é, quão alto é o grau de discrepância dos escores dentro de uma mesma amostra. Para tanto, lançamos mão das medidas desvio padrão, variância e coeficiente de variação.

O desvio padrão é uma medida do grau de dispersão de um conjunto de dados (no caso, os escores de musicalidade). Em outras palavras, trata-se de uma medida que indica o quão uniforme esses escores são. Quanto mais próximo de zero for o valor do desvio padrão, mais homogêneos são os dados e, portanto, menos discrepância entre os escores dentro de uma amostra (KHAN ACADEMY BRASIL, 2014a, GOUVEIA, s.d., VARIÂNCIA..., s.d.).

---

<sup>76</sup> O manual do software BioEstat apresenta como exemplo de aplicação do teste Mann-Whitney verificar se há diferença nos níveis de fosfatase sanguínea em dois grupos de pacientes de malária, cada um acometido por um tipo de parasita. No nosso caso, temos dois grupos independentes (uma pessoa não pode ser falante e não-falante ao mesmo tempo), e buscamos avaliar se há diferença nos níveis de musicalidade percebida por esses dois grupos (falantes e não-falantes).

A variância, assim como o desvio padrão, também é uma medida de dispersão. A variância determina o quanto os dados de um conjunto se afastam da média. A diferença entre os dois é que o desvio padrão é definido como a raiz quadrada da variância, e é expresso na mesma unidade de medida dos dados.

Já o coeficiente de variação indica a dispersão que exclui a influência da ordem de grandeza variável (sendo útil, por exemplo, quando se quer comparar a dispersão em torno da média de idade e de altura de dois grupos de pessoas). Um coeficiente de variação menor ou igual a 15% indica baixa dispersão e, portanto, dados homogêneos, enquanto um coeficiente maior que 30% indica alta dispersão e dados heterogêneos (AYRES *et al.*, 2007, p. 144 e 341, GOUVEIA, s.d., KHAN ACADEMY BRASIL, 2014a).

Por fim, um último aspecto estatístico será relevante nesta seção: verificar o grau de associação de duas variáveis. À associação entre duas variáveis se dá o nome de correlação, ou seja, a presença ou a intensidade observada de uma variável está diretamente associada à presença ou à intensidade observada de outra variável (AYRES *et al.*, 2007, p. 341, BELISARIO; SANTOS, s.d., KHAN ACADEMY BRASIL, 2014b). Por exemplo, pode-se verificar se a variável “quantidade de consoantes na estrutura silábica” está associada aos escores de musicalidade atribuídos pelos informantes, bem como se essa associação é positiva ou negativa (maior quantidade de consoantes associada a maior musicalidade, ou quantidade maior de consoantes associada a menos musicalidade).

Note-se, porém, que correlação não implica causalidade. No exemplo em questão: ainda que verificássemos uma forte correlação negativa entre quantidade de consoantes e musicalidade, um teste de correlação não é capaz de nos dizer se a maior quantidade de consoantes diminui a musicalidade da língua ou vice-versa, muito menos nos mostra se um terceiro fator não considerado é o causador dessa correlação. Nesta subseção, utilizamos o teste de Correlação Linear de Pearson (AYRES *et al.*, 2007, p. 85 e 341) para medir a associação entre duas variáveis que serão abordadas mais adiante.

#### **4.4 Resultados**

Como já mencionado, o site do projeto *Lullabies of the World* apresenta, além de interpretações artísticas das canções de ninar coletadas, gravações de áudio das letras dessas canções sendo declamadas por falantes de seus respectivos idiomas, a fim de mostrar a pronúncia do texto de forma mais clara do que em uma interpretação musical. Os informantes do presente estudo foram apresentados a seis trechos curtos, em seis idiomas diferentes, dessas

gravações. Foi pedido a esses informantes que avaliassem o quão melodiosas/musicais essas línguas soavam, em uma escala bipolar de um a cinco, sendo um “Nada melodiosa/musical” e cinco “Muito melodiosa/musical”. Também foi pedido que informassem se conheciam a língua em questão (com as opções “Sim”, “Não” ou “Não tenho certeza”) e, se assim desejassem, escrevessem qual língua acreditavam estar ouvindo (com a pergunta “Que língua você acha que é essa?” seguida de um campo de texto no qual poderiam digitar seus palpites).

A tabela 2 apresenta as médias aritméticas das respostas obtidas para cada idioma, bem como seus respectivos valores de Variância e de Desvio Padrão. Sombreados em azul acinzentado estão os menores valores de cada linha, e sombreados em laranja, os maiores valores.

**Tabela 2** - Dados numéricos acerca da percepção da musicalidade de seis idiomas por parte dos informantes.

	<b>Armênio</b>	<b>Alemão</b>	<b>Húngaro</b>	<b>Hindi</b>	<b>Islandês</b>	<b>Norueguês</b>
Tamanho da amostra	193	193	193	193	193	193
Média aritmética	3,0311	2,6736	2,8964	3,2176	2,8898	3,3679
Variância	1,4574	1,3981	1,3121	1,3170	1,2409	1,3067
Desvio padrão	1,2072	1,1824	1,1455	1,1476	1,1140	1,1431

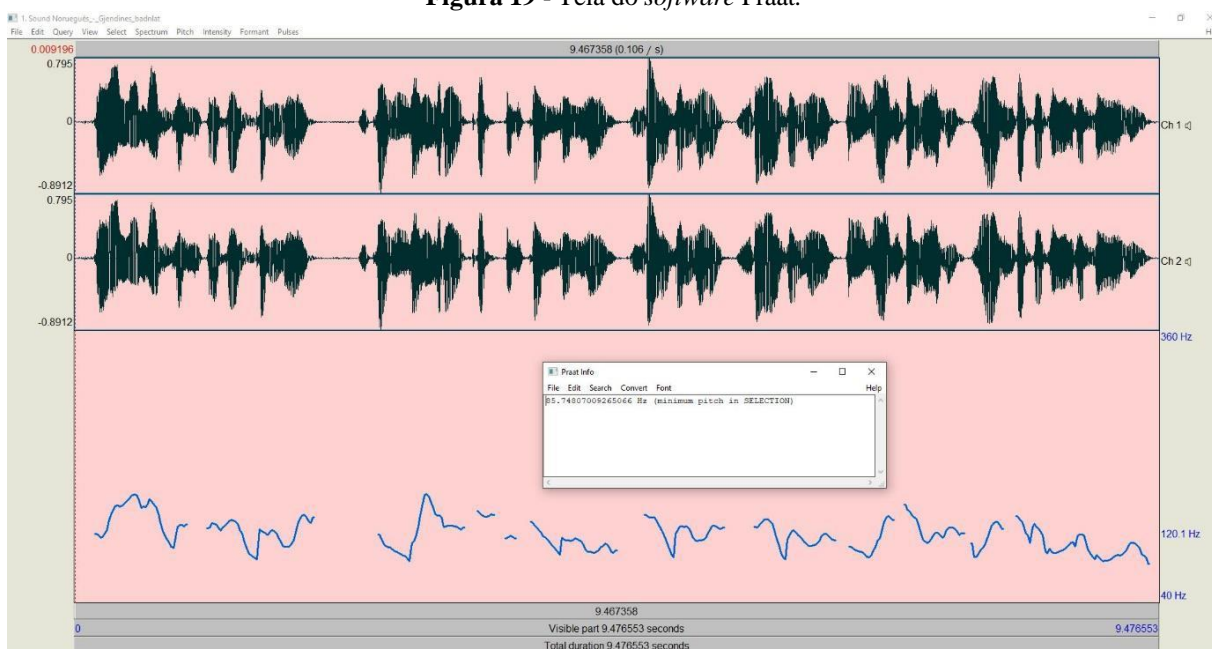
Os valores da média aritmética mostram que a língua alemã foi considerada a menos musical/melodiosa dentre as seis línguas avaliadas neste experimento, enquanto a língua norueguesa foi considerada a mais musical/melodiosa. O idioma armênio, por sua vez, foi o idioma que mais dividiu a opinião dos informantes, como atestado pelos valores de variância e desvio padrão. O idioma alemão foi o segundo mais divisivo, enquanto o islandês foi avaliado de forma homogênea.

É pouquíssimo provável que um único fator seja suficiente para explicar a disparidade na percepção desses idiomas. Isso não significa, no entanto, que não possamos buscar correlações ou associações entre parâmetros acústicos das gravações e as pontuações obtidas por cada língua. Considerando que o aspecto mais mencionado como resposta à pergunta “Na sua opinião, o que faz uma língua soar musical e/ou melodiosa?” foi a altura melódica, podemos verificar se está correlacionada com a percepção de musicalidade.

O contorno de *pitch* das alturas melódicas foi analisado através do *software* Praat. Além do espectrograma do som (como se pode ver na metade superior da figura 19), o programa permite a visualização gráfica de componentes do sinal sonoro, como a altura melódica (as linhas azuis, na figura 19), a intensidade e os formantes. O *software* permite, com um clique, obter o valor mínimo ou o valor máximo da altura melódica do áudio selecionado (a janela com

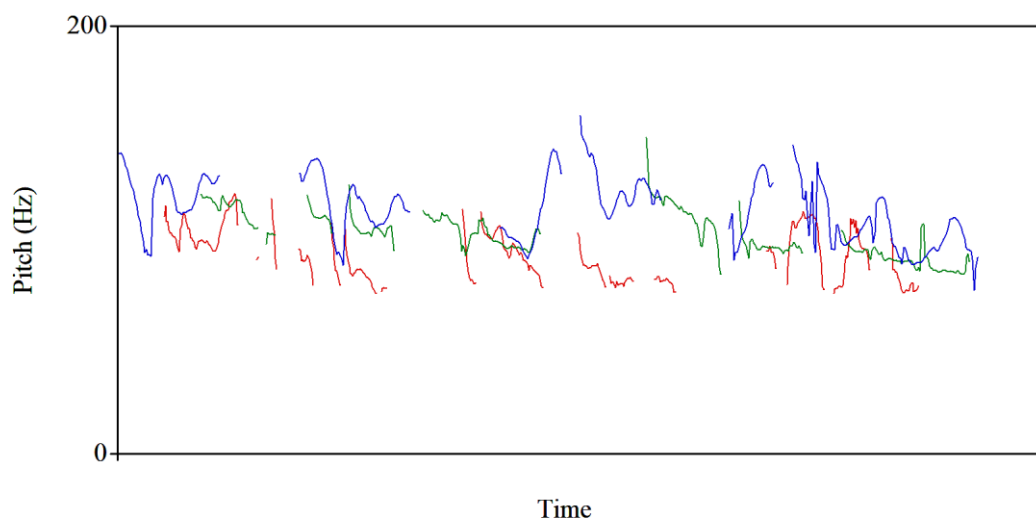
o resultado do comando *Get Minimum Pitch* aparece sobreposta à tela na figura 19). No nosso caso, selecionamos o áudio inteiro e obtivemos os valores da altura melódica mínima e da altura melódica máxima de toda a fonação. Subtraindo o valor mínimo do valor máximo, calculamos a variação da frequência fundamental ( $f_0$ ) em cada amostra de fala.

**Figura 19** - Tela do software Praat.

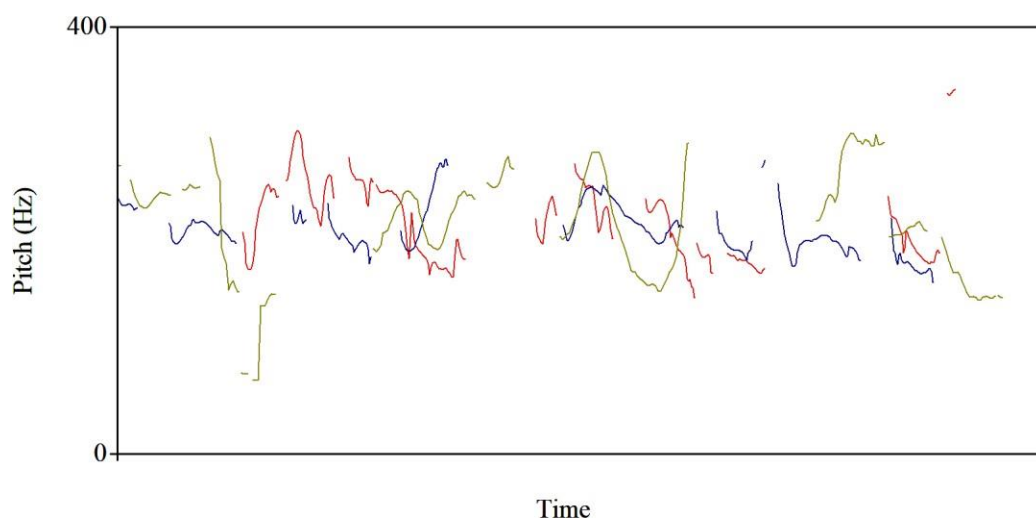


Metade das amostras apresentava falantes do sexo masculino, enquanto a outra metade trazia vozes de falantes do sexo feminino, e verificou-se que os valores de variação da frequência fundamental eram muito maiores para as vozes femininas do que para as vozes masculinas (figuras 20 e 21). Isto, junto ao fato já mencionado em outros trabalhos que vozes de homens parecem acarretar percepções distintas das vozes de mulheres, nos levou a analisar uma possível associação separando o grupo das vozes femininas das vozes masculinas.

**Figura 20** - Variação da altura melódica (em Hz, no eixo vertical) nas fonações de vozes masculinas ao longo do tempo (no eixo horizontal). Alemão em vermelho, húngaro em verde e norueguês em azul.



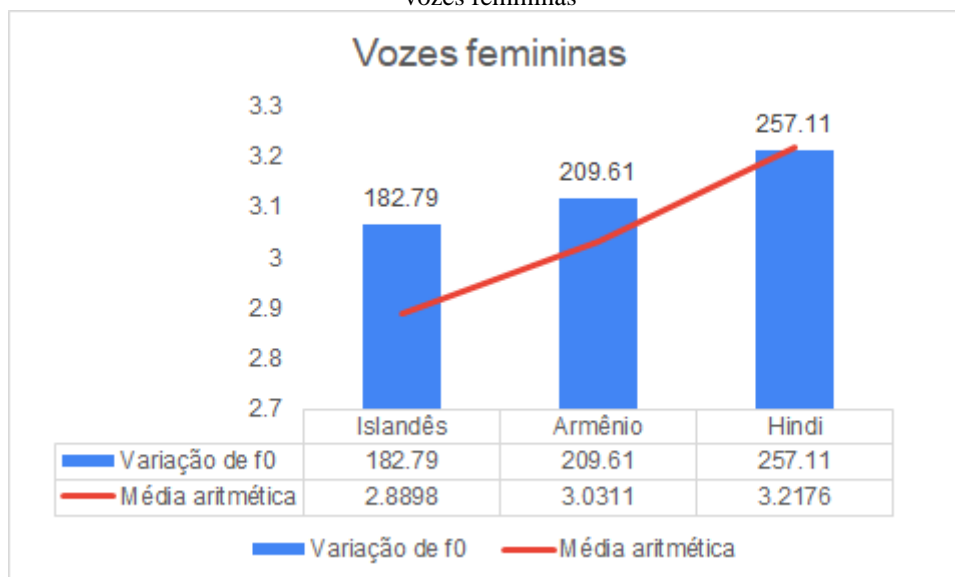
**Figura 21** - Variação da altura melódica (em Hz, no eixo vertical) nas fonações de vozes femininas ao longo do tempo (no eixo horizontal). Armênio em vermelho, hindi em ocre e islandês em azul.



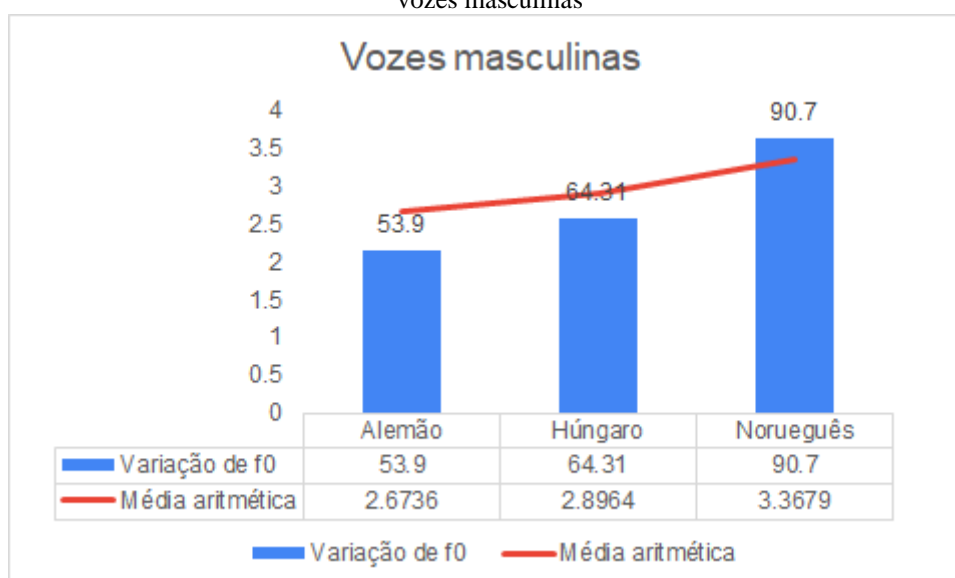
O número reduzido de amostras (três em cada grupo) diminui a precisão e o poder de testes estatísticos para confirmar a significância dos valores apontados acima. No entanto, em uma perspectiva mais qualitativa do que quantitativa, os dados parecem configurar um indício de uma possível correlação entre variação de altura melódica e musicalidade, especialmente se considerarmos que em ambos os grupos a tendência observada foi a mesma: uma maior variação de *pitch* estaria associada a uma maior percepção de musicalidade do idioma. Os gráficos 13 e

14 mostram a relação entre a variação de altura melódica e a média aritmética da percepção de musicalidade para cada idioma.

**Gráfico 13** - Relação entre variação de f0 e média do nível de musicalidade percebido em amostras de fala com vozes femininas



**Gráfico 14** - Relação entre variação de f0 e média do nível de musicalidade percebido em amostras de fala com vozes masculinas



Uma hipótese apreciada é a de que a maior variação de altura melódica fosse resultado da influência da melodia na declamação, ou um tipo de “coação”, por assim dizer, imposta aos padrões entoacionais pelo gênero “texto declamado”. Obviamente, não é possível discriminar

*se e nem até que ponto* houve alteração dos padrões entoacionais no momento de declamar os textos. Afinal, não dispomos de amostras de fala espontânea desses leitores para fazer uma comparação mais minuciosa. Ainda assim, acreditamos que a variação observada entre as amostras é, provavelmente, melhor explicada pelo sistema linguístico de cada língua do que por influências musicais ou restrições de gênero.

Mostramos, anteriormente, que a tessitura das seis melodias é bastante próxima. Em outras palavras, há pouca variação entre a distância entre a nota mais grave e a nota mais aguda de cada melodia. Parece-nos razoável supor que, caso a variação de altura melódica fosse um reflexo da melodia, deveríamos perceber algum tipo de correlação entre melodia e texto declamado – isto é, melodias com tessitura maior levando a maior variação de  $f_0$  quando o texto fosse declamado, e tessituras de tamanho mais ou menos iguais levando a declamações com variações de  $f_0$  mais ou menos correspondentes. Porém, não é possível estabelecer essa relação. Ao mesmo tempo, não acreditamos que a maior variação de  $f_0$  em uma ou outra língua deva ser atribuída ao ato de declamar um texto originalmente musical. Afinal, todos os seis estímulos tinham a mesma natureza no que se refere a gênero. Mesmo assumindo que o fato de ser um texto declamado leve a uma maior variação de  $f_0$ , lembramos que tal coação estaria operando sobre todos os seis falantes. Portanto, julgamos que as diferenças na variação de  $f_0$  se devam a fatores do sistema linguístico intrínseco a cada idioma, e não a influências musicais ou de gênero do discurso.

Evidentemente, pode-se pensar que cada falante seja mais ou menos susceptível a tais coações. É provável que um ator ou um músico experiente declamasse um mesmo texto de forma muito distinta da que um leigo declamaria. Sempre há brechas para especulação e questionamento. Porém, cedo ou tarde é necessário assumir o risco e generalizar. Talvez as amostras utilizadas neste trabalho não sejam representativas de *todos* os falantes de armênio, *todos* os falantes de alemão e assim por diante, em todas as situações. Acreditamos que a argumentação apresentada até aqui seja o suficiente para dar respaldo considerável às nossas deduções de que a variação de  $f_0$  é reflexo de diferenças linguísticas entre os idiomas.

Assim como no estudo de Reiterer *et al.* (2020), a língua alemã foi considerada pelos informantes pouco musical/melodiosa. Considerando que o mesmo estudo verificou alto impacto da familiaridade dos informantes na forma como estes avaliavam as línguas em questão (REITERER *et al.*, 2020p. 188), levantamos a hipótese de que falantes de alemão pudessem atribuir maior musicalidade à língua do que os não falantes, uma vez que, sendo falantes do idioma, apresentariam maior nível de familiaridade do que os não falantes. Para verificar se havia diferença significativa na percepção da musicalidade do idioma alemão entre falantes e

não-falantes, recorremos ao teste Mann-Whitney, um teste não-paramétrico destinado a comparar duas amostras independentes. O teste apontou diferença significativa entre os dois grupos, com falantes de alemão considerando a língua mais musical/melodiosa do que os não-falantes ( $p=0,0010$ ). Ademais, os falantes de alemão apresentavam maior coesão em suas respostas, fato revelado pelos menores valores de Variância, de Desvio Padrão e Coeficiente de Variação. Tais valores podem ser observados na tabela 3.

**Tabela 3** - Diferença nos valores relativos à percepção de musicalidade da língua alemã por parte de falantes e de não-falantes do idioma.

	Falantes de alemão	Não-falantes de alemão
Tamanho da amostra	19	178
Mínimo	2	1
Máximo	5	5
Mediana	3	2
Média aritmética	3,5789	2,5899
Variância	1,1462	1,3280
Desvio padrão	1,0706	1,1524
Coeficiente de variação	29,91%	44,50%

Em que pese a distribuição desigual no tamanho das amostras, esse resultado parece ser mais uma evidência no sentido de que tendemos a considerar mais melodiosas (e, por extensão, mais belas) as línguas com as quais temos maior familiaridade. Uma segunda hipótese, derivada desta primeira, seria a de que os falantes de alemão atribuiriam notas mais altas no quesito musicalidade a todas as línguas da mesma família no estudo (além do alemão, o islandês e o norueguês). Tal hipótese parecia ser confirmada pelos valores de média aritmética, variância e desvio padrão, mas a significância estatística não foi confirmada pelo teste de Mann-Whitney, o que nos levou a descartá-la, se não como totalmente falsa, ao menos como não podendo ser sustentada com convicção pelos resultados deste estudo.

Curiosamente, também observamos que as pessoas que falavam mais idiomas apresentaram tendência a atribuir nota mais alta no quesito musicalidade aos idiomas aos quais eram apresentadas. Analisando a pontuação combinada dos seis idiomas aos quais os informantes foram expostos, a média aritmética foi de 2,7895 para os falantes de apenas uma língua, 3 para os falantes de duas ou três línguas, e 3,1327 para os falantes de quatro línguas ou

mais. Agrupando as respostas em dois grandes grupos, os falantes monolíngues e bilíngues pontuaram os idiomas com uma média de 2,8095, enquanto os políglotas atribuíram uma média de 3,1352 aos estímulos. A significância dessa diferença foi confirmada pelo teste Mann-Whitney ( $p$ -valor  $<0,0001$ ).

Por fim, voltamos a uma das hipóteses iniciais, a de que línguas com maior proporção de vogais seriam avaliadas como mais musicais. O teste de Correlação Linear de Pearson, teste utilizado para medir o grau de associação entre duas variáveis, mostrou uma leve correlação negativa  $r(\text{Pearson})=-0,4233$  entre as variáveis porcentagem de consoantes e musicalidade. Em outras palavras: quando a proporção de consoantes cresce, a percepção de musicalidade cai, e vice-versa. Porém, é importante ressaltar que o número extremamente pequeno de amostras (seis) exige que tal resultado seja relativizado.

#### **4.5 O que faz uma língua soar mais musical e melodiosa?**

Como vimos, não é incomum que, ao expressar o quão bonita e agradável uma língua soa aos ouvidos, as pessoas usem adjetivos como “musical” e “melodiosa”, que são frequentemente atribuídos a línguas que nos apetezem. Por mais intuitivas que essas metáforas possam parecer, há de se perguntar o que exatamente se está a expressar por elas. Afinal, todas as línguas orais do mundo compartilham com a música a natureza concreta primeiramente sonora e organizam seus sons ao longo do tempo obedecendo regras sistemáticas. Por que, então, nem todas as línguas são percebidas como (igualmente) musicais? A seção anterior procurou responder de forma mais objetiva se, de fato, alguns idiomas são considerados mais musicais do que outros, bem como verificar se a proporção de vogais e consoantes, a familiaridade dos informantes com o idioma em questão e a variação da altura melódica do estímulo poderiam ser fatores associados a essa percepção. A seguir, ainda no intuito de levantar alguns critérios objetivos que levem as pessoas a compararem uma língua com música, voltamo-nos para as crenças que os falantes têm sobre a musicalidade dos idiomas, bem como suas racionalizações acerca dessas percepções. No mesmo questionário online em que os participantes avaliaram os estímulos em seis línguas, perguntamos:

*Na sua opinião, o que faz uma língua soar musical e/ou melodiosa? Não se preocupe com "certo" ou "errado", apenas compartilhe a sua percepção.*

Não era obrigatório que os informantes respondessem a essa pergunta no questionário. Por isso, do total de 194 respostas ao questionário, tivemos apenas 186 respostas a essa questão específica.

Considerando que nem todas as pessoas que responderam tinham formação linguística e/ou dominavam a metalinguística fonético-fonológica e musical, foi necessário tentar fazer aproximações, na tentativa de interpretar o que os informantes responderam. 30 das 186 respostas se mostraram demasiadamente confusas ou vagas, entre elas os exemplos abaixo:

(1) Talvez seja a variação do som das vogais, por exemplo, "os" e "és" abertos e fechados ou pode ser também o "chiado" que fazemos. É difícil de explicar.

(2) línguas aonde o tom da palavra é tão importante quanto o fonema. (sic)

(3) Fonemas predominantes

(4) Sonoridade sucinta

No exemplo 1, o informante não deixa claro se o “chiado” é um elemento que aumenta ou diminui a musicalidade da língua. Também não é possível saber o que o informante quis dizer com “a variação do som das vogais”, se se referia à presença de fonemas vocálicos distintos ou se quis dizer que alguns fonemas vocálicos eram mais musicais do que outros. Já em 3, não é possível ter certeza se, na opinião expressa nessa resposta, as consoantes tornam a língua mais ou menos melodiosa, nem se o participante se refere a um grupo específico de consoantes ou a consoantes no geral (em oposição a sons vocálicos).

Quando a resposta não apresentava elementos que pudessem conduzir a uma categorização minimamente sólida, ela foi desconsiderada. Como exemplo, podemos citar participantes que parecem não ter entendido a pergunta e responderam “A musica conseguir transmitir emoções pela melodia” (sic) e “A melodia que acompanha a letra da música, os instrumentos musicais”.

Elencamos, abaixo, os parâmetros que, a nosso ver, podem ser extraídos das respostas dos informantes, bem como exemplos de respostas que mencionavam esses parâmetros. Quando mais de um parâmetro foi mencionado na mesma resposta (por exemplo, no caso de uma resposta mencionando tanto a velocidade de fala quanto a predominância de vogais), a resposta foi “pontuada” uma vez para cada parâmetro mencionado – o que explica a soma de

menções aos parâmetros (222) ser superior ao total de respostas a essa questão (186). Um caso desse tipo é a resposta “A variação da entonação para mais baixa, alta nas sentenças (sic) e perguntas, sons longos ou curtos”, em que a menção à distribuição dos sons ao longo do tempo (sons longos ou curtos) foi interpretada como sendo uma alusão ao ritmo, enquanto a menção à entonação foi interpretada como altura melódica. Os autores das respostas transcritas não são identificados, pois a participação no experimento era anônima. Também é importante ressaltar que as respostas abaixo não foram alteradas ou corrigidas, de modo que foram mantidos desvios ortográficos ou gramaticais que elas apresentassem.

### **Parâmetro 01: Altura melódica (entonação)**

Exemplo de resposta: *Acho que as variações de intensidade e altura (frequência em Hz) são os fatores mais relevantes. A ritmicidade também conta, mas isso varia muito de indivíduo e seu estado de espírito (...). Aqui no RS, estado onde moro, vejo que as pessoas falam de forma mais arrastada e com muita variação de alturas, o que passa a ideia de que gaúchos falam de forma mais "cantada".* (Grifos nossos)

A altura melódica (ou *pitch*) se refere à frequência fundamental ( $f_0$ ) ao longo da fonação, seja essa  $f_0$  estável ou variável. Menções à altura e à frequência do som foram interpretadas como se referindo a esse parâmetro. O termo “entonação” também foi interpretado como se referindo à altura melódica, uma vez que esse termo era frequentemente colocado em oposição ao ritmo.

### **Parâmetro 02: Consoantes**

Exemplo de resposta: *Penso que a presença de consoantes que truncam o fluxo das vogais.* (Grifo nosso)

A resposta acima parece contradizer não só o senso comum, como também a tendência geral observada entre os demais informantes. No entanto, trata-se de uma resposta válida, ainda que contraintuitiva, para a pergunta “o que faz uma língua soar musical e/ou melodiosa”. Considerando que não podemos confirmar diretamente com o informante a intenção por trás dessa resposta, optamos por fazer a menção às consoantes como um parâmetro distinto, a fim de manter a integridade científica do estudo.

### **Parâmetro 03: Continuidade do fluxo de som**

Exemplo de resposta: *Quando ela parece fluir mais naturalmente sem "uma trava" de palavras, por exemplo, semelhante ao francês, onde parece que as palavras são uma só, e diferente do alemão, onde é marcado o fim da palavra.* (Grifos nossos)

Vários informantes mencionaram a sensação de que línguas mais musicais apresentam menos interrupções da fonação, e a impressão de que as palavras se juntam em um longo segmento sonoro. Palavras como “fluidez”, “conexão” e “linearidade” foram algumas que apontaram para esse parâmetro, o qual optamos por chamar de continuidade.

### **Parâmetro 04: Estrutura silábica**

Exemplo de resposta: *A "ligadura" sonora entre as palavras, sílabas "rendondas" (consoante + vogal)* (Grifo nosso)

Respostas que fizeram menção explícita à estrutura silábica ou à forma como as palavras se organizam em sílabas foram interpretadas como se referindo a um parâmetro específico, aqui chamado de estrutura silábica.

### **Parâmetro 05: Facilidade**

Exemplo de resposta: *Geralmente a entonação com a qual as palavras são ditas, se ocorrem muitas aliterações, se as palavras são curtas ou de fácil pronúncia* (Grifo nosso)

Nenhuma das respostas que apontou esse parâmetro definiu exatamente o que torna uma língua mais “fácil” de ser pronunciada do que outra. Ainda assim, optamos por estabelecer esse critério, já que palavras-chave relacionadas a facilidade apareceram mais de uma vez.

### **Parâmetro 06: Particularidades do falante**

Exemplo de resposta: (1) *O sentimento, o desejo das pessoas naquele momento. Se é uma poesia, uma pergunta<sup>77</sup>, uma sentença triste ou feliz, um relato de algo acontecido. Cada uma dessas opções intimamos a língua, automaticamente, de forma diferente.* (Grifo nosso)

(2) *A forma com a qual língua é falada, os maneirismos, independente da complexidade fonética da língua.* (Grifo nosso)

---

<sup>77</sup> Foi-nos apontado que a resposta reproduzida neste trecho aponta consciência, por parte do informante, que a declamação de uma poesia pode ter uma “melodia” diferente, e concordamos com essa afirmação. Apesar disso, o senso comum ainda vai na direção de que há línguas mais adequadas à poesia e à música (ou pelo menos, a alguns estilos de música) do que outras.

Alguns falantes pontuaram que, mais do que o sistema fonológico em si, são os aspectos relacionados ao falante e à situação da enunciação – como o tipo de texto, o timbre de voz e as atitudes do falante – que definem o quão musical ou melodiosa uma língua soa. Essas respostas foram agrupadas como se referindo a um parâmetro extralinguístico, aqui referenciado através do termo Falante, por se referir a aspectos particulares do usuário que atualiza o sistema linguístico.

### **Parâmetro 07: Familiaridade**

Exemplo de resposta: *Acho que são mais melodiosas aos meus ouvidos as línguas que eu já ouvi associada a músicas com mais frequência no meu dia a dia. E também acho que línguas latinas que a pronuncia eu já me acostumei e acho que o som das vogais e das palavras parece mais "espontâneo" de ser falado e mais prazeroso de ser ouvido.* (Grifos nossos)

Houve respostas em que os falantes mencionaram que percebiam (maior) musicalidade em línguas que já conheciam, ou que julgavam parecidas com os idiomas com os quais estavam acostumados. Respostas que associavam a musicalidade com o conhecimento ou a sensação de familiaridade com um idioma foram agrupadas sob esse rótulo.

### **Parâmetro 08: Nasalização**

Exemplo de resposta: *A mudança na entonação das palavras e os locais da árvore respiratória utilizados para produção do som (movimentos da língua, sons anasalados...)* (Grifo nosso)

Segmentos nasais são produzidos com o abaixamento do véu palatino, acarretando ressonância na cavidade nasal, em oposição aos segmentos orais, em que a passagem do ar para a cavidade nasal é bloqueada pelo véu palatino levantado (SILVA, 2003, p. 26). No caso das vogais, especificamente, a alteração da configuração da cavidade bucal na produção desses segmentos influencia a qualidade vocálica (SILVA, 2003, p. 91).

### **Parâmetro 09: Relatividade: Todas as línguas são/podem ser musicais**

Exemplo de resposta: *Todas as línguas me parecem melodiosas. Cada língua tem sua maneira de mostrar sentimentos e isso é música pura* (Grifo nosso)

Um grupo de informantes rejeitou explicitamente que se pudesse chegar a parâmetros objetivos para pontuar a musicalidade dos idiomas. Essas respostas validavam todos os idiomas como dotados de uma musicalidade única, ou especificavam que a musicalidade não estava no idioma em si, mas na percepção do ouvinte. Essas respostas receberam a etiqueta “Relativo”, em oposição às classificações mais categóricas dos outros parâmetros.

### Parâmetro 10: Repetição

Exemplos de respostas: (1) *Pergunta difícil. Acho que línguas que repitam alguma sonoridade com frequência. Algo que pareça uma rima ou uma marcação de ritmo (Grifo nosso)*

(2) *Pelo que percebi da minha percepção das línguas do teste, era a ausência de "r" e muitas repetições, tem línguas que "fluem" melhor, pra mim tipo o espanhol e o francês. (Grifo nosso)*

Ocasionalmente, alguns informantes evocaram ideias relacionadas à repetição, através de construções como “as palavras terem finais parecidos”. Note-se que a noção de repetição inclui, mas não se limita aos segmentos em posição de coda silábica na última sílaba das palavras.

### Parâmetro 11: Ritmo

Exemplos de respostas: (1) *O ritmo com que pronunciamos as palavras e o tempo utilizado na pronúnciação de cada palavra/frase (Grifo nosso).*

(2) *Ter un ritmo bem definido, uma cadência perceptível e marcante (Grifos nossos)*

O ritmo se refere à distribuição de acentos ao longo do tempo<sup>78</sup>. Sílabas acentuadas tendem a ser mais longas e pronunciadas com maior intensidade do que sílabas átonas. O acento, no entanto, não está intimamente ligado à variação da altura melódica, justificando a sua separação em um parâmetro à parte.

### Parâmetro 12: Menos sons guturais

Exemplos de respostas: *Me parece que as línguas germânicas, ou do leste europeu, mais guturais são menos melodiosas, talvez pela ausência de vogais ou pelas posições do aparelho vocal drante (sic) a fala. (Grifo nosso)*

Alguns informantes apontaram alguns grupos fonêmicos específicos como fatores que levavam a uma diminuição da percepção de musicalidade de um idioma. Mais especificamente, associaram *menor* musicalidade ao modo de articulação de consoantes cujo lugar de articulação se encontra na parte posterior do aparelho fonatório. Como registrado no dicionário Priberam (GUTURAL, 2021), o termo gutural se refere, na linguagem comum, a algo que sai ou provém da garganta, ou aos sons articulados na região do palato mole, da faringe ou da laringe, na linguagem especializada. Considerando que a úvula pende do véu palatino e que a glote é uma abertura da laringe, consideramos que a palavra *gutural* pode ser usada, para os objetivos deste

---

<sup>78</sup> Ver seção 4.2.

trabalho, como um termo guarda-chuva que nomeie este parâmetro, englobando consoantes tradicionalmente classificadas como velares, uvulares e glotais<sup>79</sup>.

### **Parâmetro 13: Menos sons plosivos**

Exemplo de resposta: *Som das vogais e poucas consoantes plosivas* (Grifo nosso)

Assim como as consoantes com lugar de articulação na garganta (no caso, no véu palatino, na úvula e na glote), nestas respostas os informantes associaram *menor* musicalidade ao modo de articulação de consoantes oclusivas, também chamadas de plosivas (ver seção 4.2).

### **Parâmetro 14: Variedade**

Exemplo de resposta: *Não ter muita repetição dos mesmos vocábulos* (Grifo nosso)

A fim de rigor científico, incluímos o parâmetro “Variedade”, em oposição a “Repetição”. As respostas enquadradas neste parâmetro se referiram tanto à variedade de sons quanto à variedade de palavras na fala (lexical).

### **Parâmetro 15: Velocidade / Taxa de elocução**

Exemplos de respostas: *Quando a fala é cantada e mais lenta* (Grifo nosso)

A taxa de elocução se refere à quantidade de palavras ou sílabas pronunciadas em um determinado espaço de tempo. Não deve ser confundida com o ritmo<sup>80</sup>.

### **Parâmetro 16: Sons vozeados**

Exemplos de respostas: *Fonemas suaves. Geralmente línguas latinas são mais palatáveis e agradam mais. Línguas germânicas ou com pronúncia semelhante soam mais ásperas* (Grifo nosso)

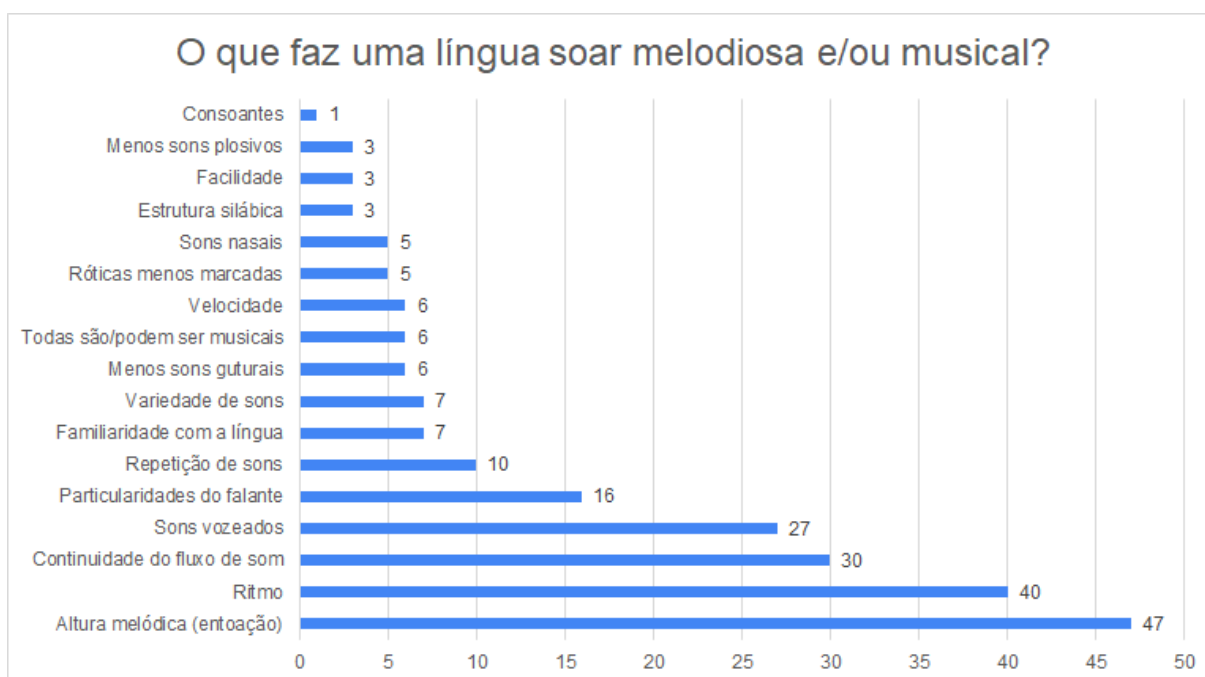
Essa categoria engloba tantos sons de vogais quanto sons consonantais vozeados. Ainda que nenhum participante tenha utilizado exatamente o termo “vozeamento”, foram frequentes as menções a “sons suaves”, “macios” e consoantes “menos pesadas”. A metáfora do vozeamento como elemento que confere leveza e suavidade aos fonemas está presente inclusive em materiais didáticos, e por isso assumimos com certo grau de confiança que os informantes se referem ao vozeamento ao empregar termos como “suave”. Assim, acreditamos que fazia sentido agrupar as consoantes vozeadas e as vogais em uma única categoria, em vez de criar dois parâmetros, “Vogais” e “Consoantes vozeadas”.

<sup>79</sup> Sobre a classificação de consoantes de acordo com local de articulação, ver seção 4.2.

<sup>80</sup> Ver seção 4.2.

Elencados os parâmetros mapeados nas respostas, passemos à análise quantitativa. O gráfico 15 mostra a frequência de menção a cada um dos parâmetros elencados acima, organizados em ordem do menos mencionado para o mais mencionado.

**Gráfico 15** - O que faz uma língua soar melodiosa e/ou musical?



Como se pode ver pelo gráfico, os parâmetros *Altura melódica (entoação)* e *Ritmo* foram os mais mencionados pelos informantes. Este resultado dialoga, ainda que não seja exatamente o mesmo, com o trabalho de Reiterer *et al.* (2020). Apesar das diferentes metodologias e naturezas dos fenômenos avaliados, nossos resultados concordam com Reiterer *et al.* (2020) no que se refere à importância do ritmo. Por outro lado, enquanto no estudo mencionado encontrou-se uma associação forte entre a familiaridade com o idioma e alto eroticismo<sup>81</sup> (REITERER *et al.*, 2020, p. 193), no nosso estudo esse aspecto se mostra presente, mas está longe de ser um dos mais influentes na percepção de musicalidade (ou, no mínimo, não está entre os aspectos que os informantes *conscientemente* parecem *considerar* mais relevantes). O alto número de respostas apontando a entoação como um fator importante, porém, dá força à interpretação expressa na seção 4.4, de que a maior variação de altura melódica estaria associada à maior percepção de musicalidade em uma amostra de fala.

<sup>81</sup> Termo usado pelos pesquisadores para se referir ao quão atraente e bonita a língua em questão soa.

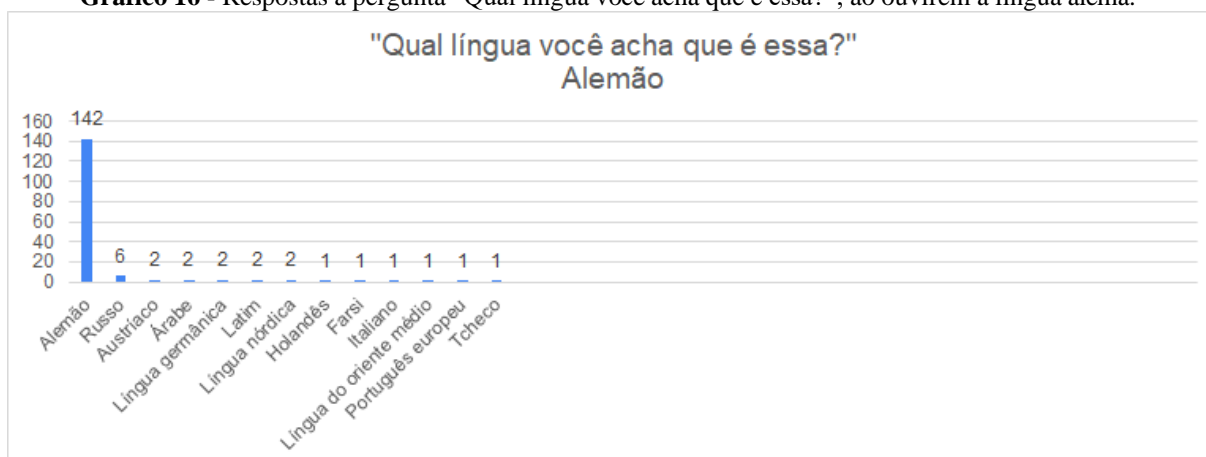
#### 4.6 Que língua você acha que é essa?

Além de tentar verificar quais aspectos os informantes julgavam mais influentes na percepção de musicalidade de um idioma, também pedimos que os informantes deixassem seu palpite acerca de qual língua estavam ouvindo, se assim desejassem. A princípio, o intuito dessa questão era averiguar possíveis falsas identificações (por exemplo, um informante afirmando conhecer a língua da gravação ao ouvir o estímulo em húngaro, mas dizendo que a língua em questão era o francês).

Nesse sentido, observamos um número baixo de falsas identificações. A língua mais falsamente identificada foi o hindi, mal identificado por seis informantes que afirmavam saber de qual língua se tratava (quatro pensaram ser espanhol, uma pessoa pensou ser árabe, e uma pessoa pensou ser português, talvez pensando se tratar de outra variedade), seguida do norueguês (três informantes afirmaram ser o francês, um afirmou ser sueco), do islandês (dois informantes pensaram se tratar de alemão) e do húngaro (um informante pensando estar ouvindo a língua árabe). Todos os respondentes que afirmaram conhecer a língua alemã acertaram o palpite.

Buscamos, então, observar possíveis padrões emergentes da distribuição dos palpites que buscavam identificar os idiomas – distribuição essa que comentaremos a seguir. Começaremos pelos palpites dos informantes ao ouvirem a língua alemã (gráfico 16), que no estudo de Reiterer *et al.* (2020) foi reconhecida por quase todos os informantes.

**Gráfico 16** - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua alemã.



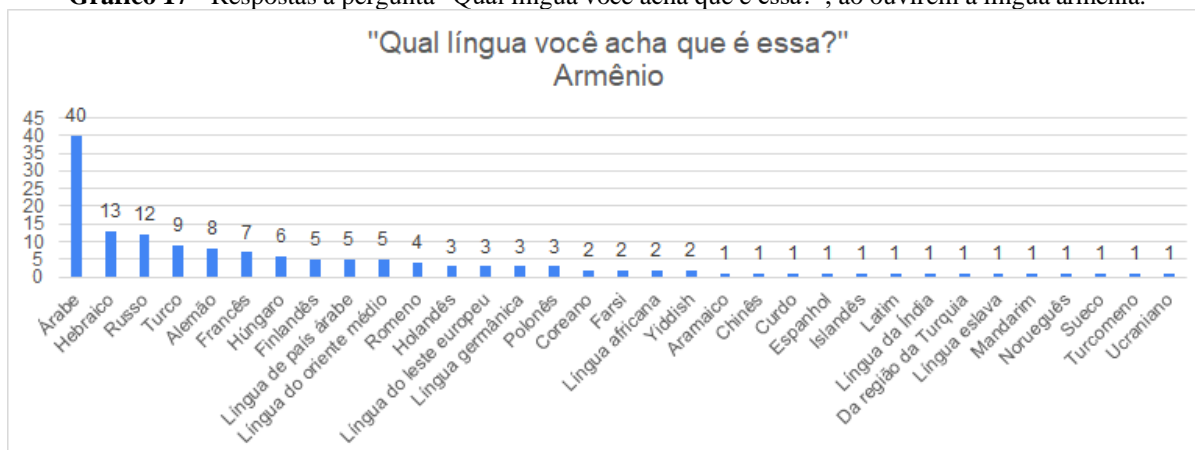
De forma semelhante aos resultados encontrados no estudo de Reiterer *et al.* (2020), a língua alemã apresentou alto índice de identificação. É possível que isso se deva tanto à notoriedade deste idioma no cotidiano das pessoas (haja vista a relativamente grande população

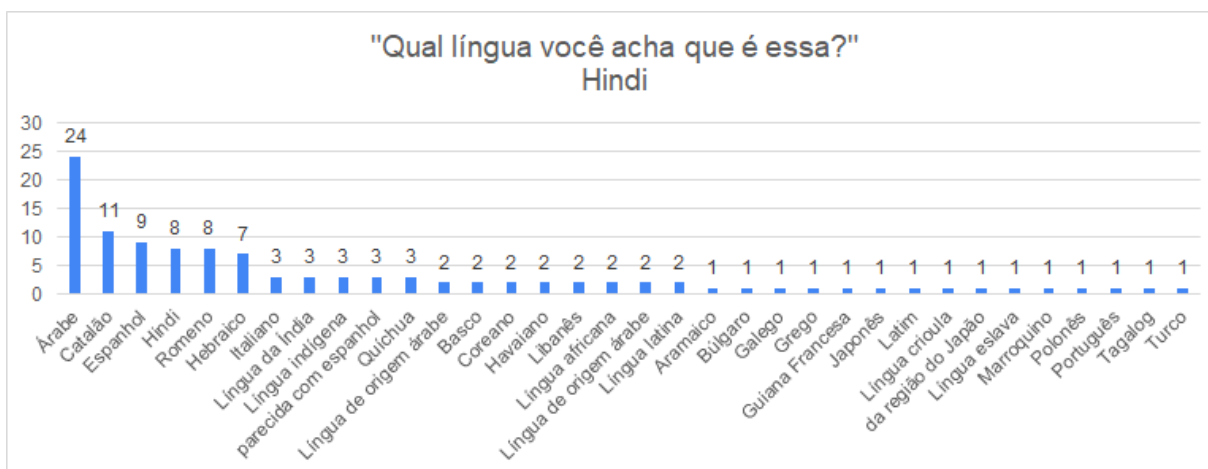
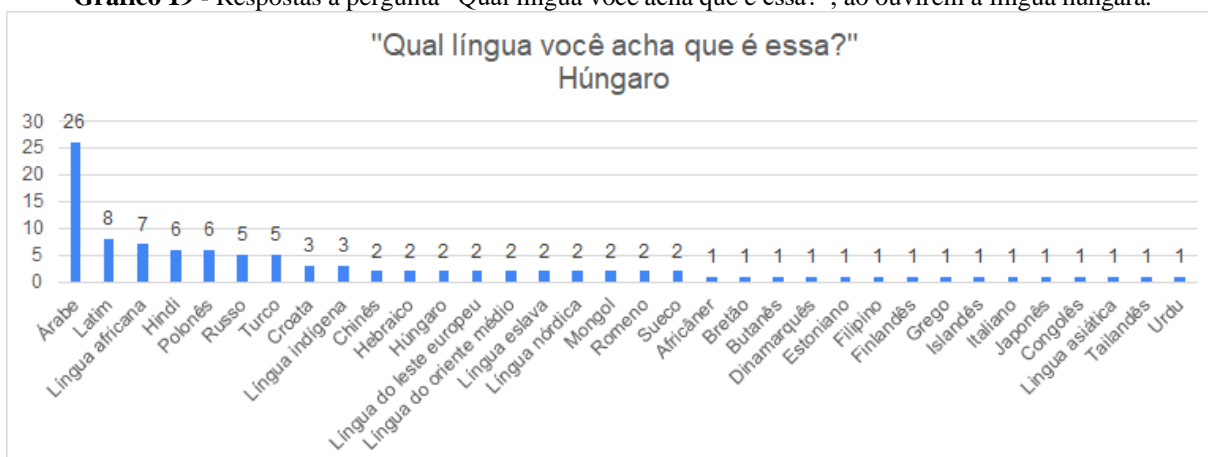
descendente de alemães no Brasil, bem como o grande número de celebridades, cientistas e demais personalidades de origem alemã influentes, entre os quais se poderia citar a modelo Gisele Bündchen, o tenista Gustavo Kuerten e, internacionalmente, o físico Albert Einstein). As respostas à pergunta “Na sua opinião, o que faz uma língua soar musical ou melodiosa” que mencionavam, especificamente, o alemão ou línguas germânicas, associaram o idioma à alta ocorrência de consoantes, a um som mais ríspido e truncado. Parece claro que a maioria dos informantes tem uma ideia razoavelmente consistente acerca da identidade fonológica da língua alemã ou, no mínimo, das línguas de origem germânica.

Identificar a maioria das línguas utilizadas neste estudo pode ser considerada uma tarefa traiçoeira, já que buscamos ativamente selecionar línguas menos difundidas e conhecidas (à exceção do alemão). Assim, não surpreende que o índice de identificação para as línguas a seguir seja extremamente baixo. É mais interessante tentar analisar e explicar o que leva os informantes a dar os palpites expressos nas respostas.

O armênio (gráfico 17) é uma língua indo-europeia que constitui sua própria família, ou seja, não possui “parentes próximos”. Também apresenta relativamente poucos falantes no mundo e não é um idioma de alta penetração científica e cultural. Ainda que exista uma comunidade de imigrantes e descendentes de imigrantes armênios no Brasil, seria pouquíssimo provável que qualquer dos informantes tivesse conhecimento do idioma, ou tivesse tido contato o suficiente para formar uma imagem mental da identidade fonológica da língua. O mesmo pode ser dito do húngaro (gráfico 19) (com menos de 20 milhões de falantes no mundo e pouca difusão fora do seu país de origem). O dialeto Hindi Maithili (gráfico 18), por outro lado, é falado por uma população maior, de aproximadamente 34 milhões de pessoas, mas também apresenta baixa difusão internacional. Em comum, os três idiomas foram identificados pelos informantes dessa pesquisa, predominantemente, como árabe.

**Gráfico 17** - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua armênia.



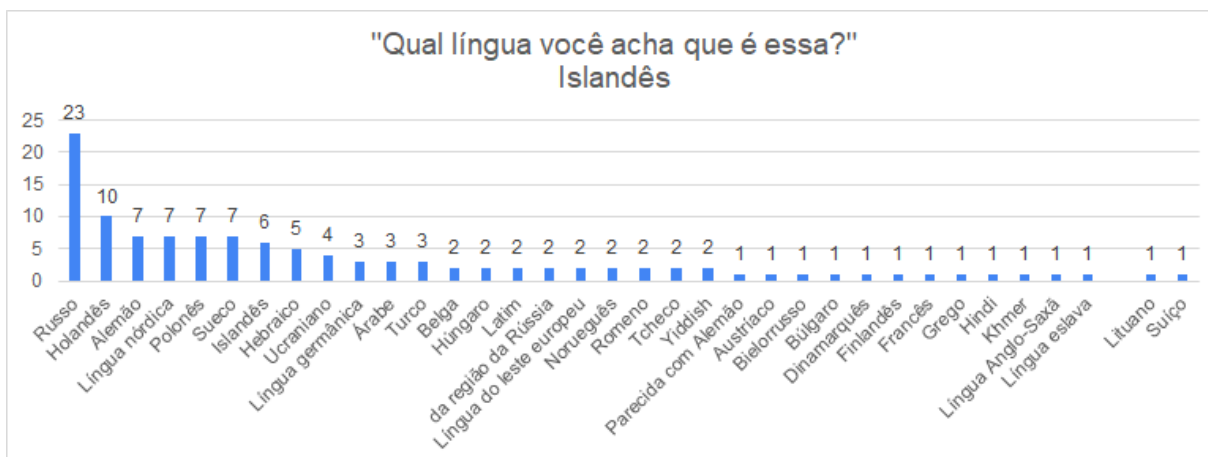
**Gráfico 18** - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua hindi.**Gráfico 19** - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua húngara.

Os dados colhidos não permitem dar uma resposta definitiva para o motivo, mas podemos esboçar interpretações. A amostra de fala do armênio apresenta várias ocorrências do que, para ouvidos brasileiros, seria interpretado como sons de <R>, consoantes fricativas velares [χ] e [ʁ], uvulares [x] e glotais [h]. As amostras de hindi e húngaro, por sua vez, têm presença saliente de consoantes líquidas [l] e [r] e africadas [tʃ] e [dʒ] (especialmente no húngaro). Tem-se, assim, evidência de que essas consoantes são associadas, pelos informantes da pesquisa, primordialmente à língua árabe, e que a saliência desses sons levou os informantes a classificarem o armênio, o hindi e o húngaro como sendo “árabe”.

O islandês (gráfico 20) foi percebido, basicamente, como uma língua eslava e/ou germânica. O alto índice de sons consonantais na amostra de fala, 58,9%, pode ser o fator principal para essa percepção. É interessante notar, porém, que a língua com presença frequente

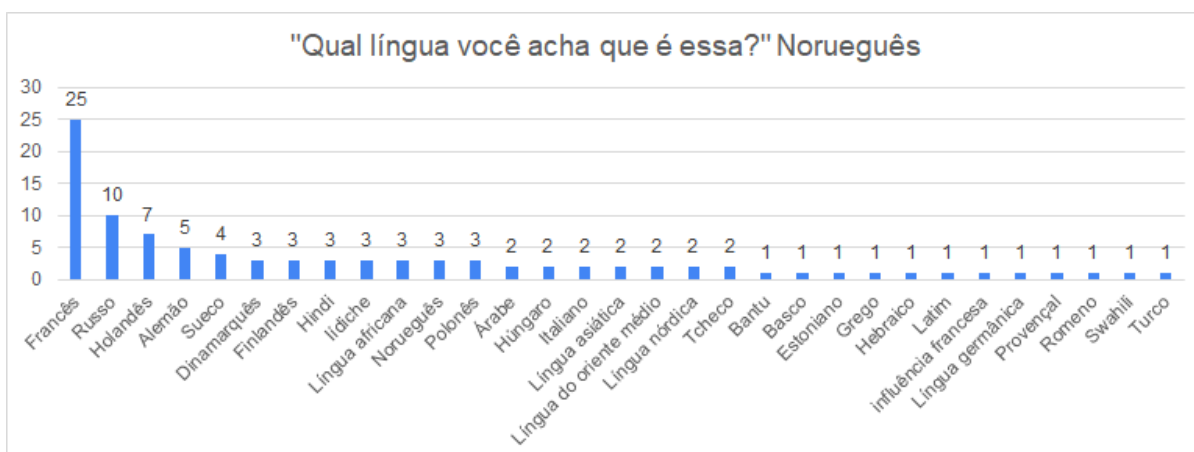
de consoantes evocou o idioma russo, e não o alemão, para a maioria dos informantes. Isto indica, como também foi sugerido por algumas respostas à questão “Na sua opinião, o que faz uma língua soar musical e/ou melodiosa?”, que a alta concentração de consoantes está ligada tanto às línguas germânicas quanto às línguas eslavas como o russo e o polonês.

**Gráfico 20** - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua islandesa.



O número alto de informantes que pensou que estava ouvindo francês quando foram expostos à língua norueguesa (gráfico 21) foi bastante surpreendente. Afinal, o norueguês é uma língua de outra família, a germânica, com alta proporção de consoantes (56%). As consoantes abundantes, de fato, parecem ter sido um fator relevante, haja vista as respostas dispersas, porém numerosas, associando o estímulo a línguas eslavas e germânicas. Há de se questionar, ainda assim, o que levou uma quantia considerável de informantes a associar a fonologia do norueguês ao francês.

**Gráfico 21** - Respostas à pergunta “Qual língua você acha que é essa?”, ao ouvirem a língua norueguesa.



Tendo em mente o alto nível de musicalidade atribuído ao norueguês, bem como a alta variação na altura melódica nesta amostra de fala e estereótipos linguísticos bastante difundidos, nossa suposição é que tantos informantes tenham pensado que o norueguês era francês *devido à alta musicalidade* atribuída ao idioma francês. Como mostram algumas das respostas à questão “Na sua opinião, o que faz uma língua soar musical e/ou melodiosa?”, o senso comum vai contra a noção de que línguas germânicas sejam melodiosas. Assim, ao ouvir uma língua com várias consoantes, mas que considera musical e melodiosa, o informante pode ser levado a tentar encontrar uma língua latina para “etiquetar” o estímulo sonoro. O francês, ao contrário do espanhol, distinto do português para não ser minimamente compreendido, mas ainda possui o estereótipo de língua “bonita e musical”. Esse pode ser, então, o raciocínio que levou informantes a classificarem uma língua germânica como sendo o prestigioso idioma francês.

Voltemos, por fim, à alta identificação da língua alemã. As línguas com maior proporção consonantal em relação ao total de sons, o norueguês (56%), o islandês (58,9%) e o armênio (59,03%) foram incorretamente identificadas como alemão por alguns participantes, húngaro (51,47%) e hindi (49,47%) não. Para além da proporção de consoantes, porém, pode haver outro fator relevante para a identificação: a presença, em posição de ataque silábico, das sequências [ft] (como em *steht* e *Sternlein*) e [fv] (como em *Schwarz* e *schwei*), é restrita à língua alemã. Tanto quanto, ou até mais, do que o “excesso” de consoantes, algumas sequências consonantais, especialmente em ataque silábico, podem ter guiado os informantes na identificação da língua alemã.

#### 4.7 Discussão

A discussão a seguir, obviamente tem como base os dados colhidos em nosso experimento. Contudo, é inevitável analisá-los em comparação com os trabalhos de Reiterer *et al.* (2020) e Kogan e Reiterer (2021), uma vez que, até onde pudemos averiguar, estes dois são os trabalhos que mais se aproximam tematicamente do último experimento abordado nesta tese. Além disso, até onde sabemos, estes são os únicos trabalhos completos e já publicados com os quais compartilhamos objetivos e metodologia de forma mais do que tangencial.

Aqui, pedimos aos informantes que avaliassem fatores que impactam a *musicalidade* do idioma. Reiterer *et al.* (2020) avaliaram os correlatos com o quão bonita e agradável uma língua soava para os participantes (em uma medida que os autores chamaram *eroticism*). A categoria

referida por *eroticism* foi retomada no estudo de 2021, no qual o quesito musicalidade foi agrupado por Kogan e Reiterer (2021) sob a categoria “Beleza”. Kogan e Reiterer (2021) afirmam que os escores de musicalidade foram semelhantes do de Eros/*eroticism*, de modo que podemos comparar os resultados de musicalidade, beleza e Eros.

Sobre Eros / *eroticism*, que apontam o quão atraente, sedutora e romântica uma língua soa, os pesquisadores afirmaram:

[...] *the eroticism of a language's sound seemed affected to a higher degree by inherent phonetic variables (isochrony: the more syllable-timed, the more erotic; sonority: the more “sonorous”, the more erotic; and vocalic share: the more vocalic share in the language sample, the more erotic)*<sup>82</sup>. (REITERER et al., 2020, p. 189).

Apesar de não contradizer aberta e totalmente a afirmação de Reiterer et al. (2020) de que línguas de ritmo silábico tendem a ser percebidas como mais agradáveis do que línguas de ritmo acentual, os resultados aqui apresentados não parecem seguir na mesma direção. Por exemplo, o norueguês - idioma de ritmo acentual com proporção vocálica relativamente baixa - foi avaliado como a língua mais musical entre as amostras com um falante do sexo masculino. Da mesma forma, a proporção de vogais em relação ao número total de fonemas não foi associada à maior musicalidade. Aqui, a variação na altura melódica pareceu ser um preditor mais confiável do nível de musicalidade. Neste sentido, nossos resultados vão na contramão a um estudo posterior de Reiterer, publicado em 2021 em coautoria com outra pesquisadora (KOGAN; REITERER, 2021). Segundo este outro estudo:

*The most important discovery was the trade-off between speech tempo and so-called linguistic melody (pitch variance): the faster the language, the flatter/more atonal it is in terms of the pitch (speech melody), making it highly appealing acoustically (sounding beautiful and sexy), but not so melodious in a “musical” sense.*<sup>83</sup> (KOGAN; REITERER, 2021, p. 1).

O que observamos em nosso último experimento foi justamente o contrário, com maior variação de altura melódica associada a maiores escores de musicalidade. Há de se considerar, aqui, a demografia bastante distinta dos informantes que avaliaram as línguas no estudo de Kogan e Reiterer (2021) (todos oriundos da Europa central) e o fato de que todas as línguas

<sup>82</sup> Em tradução livre: (...) o *eroticismo* do som de uma língua pareceu ser afetado em um grau mais alto por variáveis fonéticas inerentes (*isocronia*: quanto mais silábico o ritmo, mais erótico; *sonoridade*: quanto mais “sonora”, mais erótica; e *proporção vocálica*: quanto maior a *proporção de vogais* na amostra da língua, mais erótica).

<sup>83</sup> Em tradução livre: A descoberta mais importante foi a relação inversa entre a taxa de elocução e a chamada melodia linguística (variação de altura melódica): quanto mais rápida a língua, mais plana/atonal ela é em termos de altura melódica (melodia da fala), tornando-a altamente atraente acusticamente (soando bonito e sexy), mas não tão melodiosa no sentido “musical”.

faladas eram línguas oriundas da Europa (embora não necessariamente indo-europeias) e associadas oficialmente a pelo menos um país europeu.

É curioso notar que, em última instância, os dois estudos (REITERER *et al.*, 2020, KOGAN; REITERER, 2021) reforçam o senso comum: línguas românicas esteticamente aprazíveis, valorização do inglês em vários aspectos, associação da língua alemã a valores de ordem e disciplina. Coincidentemente ou não, estas são associações comuns também aos respectivos países, às culturas e, em algum grau, aos próprios falantes da língua. Dizer que os alemães são organizados e metódicos é lugar comum, muito mais comum de dizer que são *sexy* e charmosos. No senso comum, a sofisticação estética, charme e prazeres sensoriais são características atribuídas às artes e culinária francesa e italiana, mas não a organização e a eficiência. Assim sendo, uma questão tão pertinente quanto incômoda persiste: até que ponto os informantes dos estudos de 2020 e 2021 estavam julgando efetivamente aspectos acústicos, e até que ponto estavam (consciente ou inconscientemente) reproduzindo estereótipos sobre as famílias linguísticas e seus falantes? Afinal, mesmo que não se consiga identificar exatamente qual idioma se está ouvindo, os estereótipos são familiares (línguas eslavas e germânicas ásperas e com mais consoantes, línguas latinas com menos consoantes e mais românticas, etc) e foram observados, em algum grau, também no nosso experimento.

Não tiramos, de forma alguma, o mérito de Reiterer e seus coautores. Pelo contrário, a disposição de abordar um tópico potencialmente tão controverso e pouco estudado merece, no mínimo, reconhecimento pela ousadia. Mas as avaliações linguísticas nos parecem alinhadas um pouco *demais* com os estereótipos nacionais e culturais presentes no senso comum ocidental, o que nos leva a encarar os resultados apresentados com cautela.

Por outro lado, nossos resultados fazem coro, ainda que discretamente, com as conclusões de Hilton *et al.* (2021) sobre a percepção do sueco e do dinamarquês. Ao pedir que informantes que não tinham qualquer conhecimento dos dois idiomas (não sendo, portanto, capazes de reconhecê-los, diferenciá-los e menos ainda entendê-los) avaliassem amostras de fala de ambos os idiomas, verificou-se que os informantes avaliavam a língua sueca de forma mais positiva do que a língua dinamarquesa. Além disso, a diferença na avaliação desaparecia quando as amostras eram “monotonizadas”, isto é, quando os arquivos de som foram digitalmente alterados para remover a variação de altura melódica. Em outras palavras: não apenas o sueco foi percebido de forma mais positiva do que o dinamarquês, como essa percepção não foi influenciada por fatores extralinguísticos e pode ser atribuída, pelo menos em parte considerável, às diferenças na entoação dos dois idiomas. Sendo o sueco a língua que apresentava maior variância na altura melódica, os resultados de Hilton *et al.* (2021) são

consoantes com nossa interpretação de que maior variedade de *pitch* leva a uma maior percepção de musicalidade e agradabilidade da língua.

Evidentemente, não temos a pretensão de oferecer uma resposta definitiva à questão “o que faz uma língua soar melodiosa e/ou musical”. Porém, acreditamos que nossos resultados são evidência de que, em algum grau, a proporção de consoantes e a amplitude da variação da altura melódica, bem como a familiaridade do avaliador com a língua em questão.

#### **4.8 Considerações finais**

No terceiro e último experimento desta tese, enveredamos por um território que, até onde pudemos verificar, ainda era pouquíssimo explorado pelos estudos linguísticos publicados internacionalmente na língua inglesa, e ainda menos abordado na produção científica em língua portuguesa.

Basicamente, vimos que a maior variação da entoação está associada à maior percepção de musicalidade. Por outro lado, a maior proporção de consoantes no estímulo está associada à menor percepção de musicalidade. Tais fatos são consoantes com a opinião dos informantes: uma análise das respostas colhidas mostrou que estes consideram, além da entoação (variação da altura melódica), ritmo e menor presença de sons desvozeados na fala como índices de maior musicalidade. O índice de reconhecimento das seis línguas que formaram nosso grupo de amostras, em geral, foi baixo, com exceção da língua alemã. Por fim, os estereótipos acústicos parecem marcar presença, com a abundância de consoantes remetendo ao árabe e a línguas germânicas. Nesse mesmo sentido, levantamos a hipótese de que o norueguês tenha sido erroneamente identificado como francês por parte considerável dos informantes devido à consagrada associação entre a língua francesa e valores estéticos positivos. O norueguês apresentando maior variação da altura melódica, é percebido como altamente musical; não sendo reconhecido como as línguas tradicionalmente consideradas “atraentes” e “musicais”, como o espanhol e o italiano, acabaria sendo incorretamente rotulado como francês.

## CONCLUSÕES

Buscamos, ao longo deste trabalho, investigar alguns aspectos da relação entre (as noções de) linguagem verbal e música, tanto do ponto de vista do que essas linguagens expressam, quanto do ponto de vista da delimitação entre essas duas linguagens.

É importante, também, ressaltar que a seção 2, sobre o experimento com as amostras bissilábicas, tratou da codificação, enquanto as seções 3 e 4, acerca da percepção das músicas e da musicalidade das línguas, trataram da decodificação dos estímulos. Essa diferenciação é relevante, porque diferenças na codificação nem sempre resultam em diferenças relevantes na decodificação (SCHERER, BANSE, WALLBOT, 2001).

Buscamos, sempre que possível e até onde foi viável, fugir do lugar comum. Ao percebermos que já havia uma quantidade considerável de trabalhos se voltando para aspectos como a taxa de elocução e a intensidade do som na expressão de emoções na fala e na música, optamos por focar na possível correspondência do uso de intervalos melódicos nesses dois domínios. Ao encontrarmos poucos trabalhos abordando a ideia tão presente no senso comum de que algumas línguas soam mais bonitas e melodiosas do que outras, não hesitamos em abordar esse tema.

Trabalhar na interdisciplinaridade é sempre um desafio, mas um desafio do qual não nos furtamos cujo aspecto mais gratificante talvez seja justamente ver uma miríade de opções de encaminhamentos futuros se abrindo. É sobre esses possíveis encaminhamentos que nos debruçaremos no final deste trabalho.

Como mostraram os resultados do experimento número um, as evidências de uma correspondência do uso de intervalos musicais na fala e na música ainda não são sólidas. A escassez de pesquisas nesse sentido deve ser revertida, a fim de poder corroborar, compreender ou descartar essa hipótese com maior segurança. A correspondência do uso de intervalos na fala e na música existe, de fato? Essa correspondência é específica de acordo com a comunidade linguística e a tradição musical, ou pode ser observada entre culturas diferentes? Ainda é cedo para dizer, daí a importância de reproduzir e aperfeiçoar as metodologias, e aplicá-las a comunidades linguísticas e culturais distintas.

Os resultados do experimento número dois deixaram claro que é inadequado dizer que o intervalo de terça menor está associado à expressão de tristeza na música. Tal intervalo pode ser um dos fatores, e tomado isoladamente pode, realmente, remeter a afetos negativos. Porém, quando em um contexto melódico, não parece conduzir necessariamente à percepção de tristeza.

Este resultado pode relativizar um pouco as conclusões de trabalhos como Curtis e Bharucha (2010) e Schreuder, van Eerten e Gilbers (2004).

Por fim, os resultados do último experimento evidenciam a necessidade de expandir os trabalhos da área chamada por Kogan e Reiterer (2021) de fonostética. Se quisermos entender como e pelo que são guiadas as percepções estéticas dos indivíduos ao avaliarem uma língua como mais ou menos agradável e musical aos ouvidos, faz-se mister utilizar como estímulos línguas desconhecidas do grande público e de outras famílias que não a indo-europeia, bem como verificar a avaliação de pessoas de grupos étnico-culturais distantes da tradição ocidental/europeia, bem como levar em consideração a possível influência de estereótipos socioculturais latentes sobre as avaliações.

Gostaríamos de, metaforicamente, pensar neste trabalho como um modesto tijolo de evidências a ser alocado em uma das paredes de uma grande construção: a compreensão da natureza e da dinâmica da relação entre a língua e a música. Esta construção já demonstra alicerces sólidos, mas ainda está longe de ser completada. Esperamos que esta nossa contribuição possa servir como apoio para os muitos outros tijolos que, com esperança e otimismo, acreditamos estarem prestes a se juntar a nós nessa empreitada. Que venham as próximas peças deste quebra-cabeça. Este final não é senão o início de uma jornada maior.

## REFERÊNCIAS

- ANTOLIK, L. And the most beautiful languages in the world are... **Education First**. 19 de fevereiro de 2020. Disponível em: <<https://www.ef.com/wwen/blog/language/most-beautiful-languages-in-the-world/>> Acesso em 12 de maio de 2021.
- AYRES, M.; AYRES JR, M.; AYRES, D.L.; SANTOS, A. S. **BioEstat**: Aplicações estatísticas nas áreas das ciências biológicas e médicas: Belém: Sociedade Civil de Mamirauá, 2007.
- BAILEY, T. M.; PLUNKETT, K.; SCARPA, E. A cross-linguistic study in learning prosodic rhythms: Rules, constraints, and similarity. **Language and Speech**, v. 42, n. 1, p. 1-38, 1999.
- BANSE, R., SCHERER, K. Acoustic Profiles in Vocal Emotion Expression. **Journal of Personality and Social Psychology**, v. 70, n. 3, p. 614-636, 1996.
- BÄNZIGER, T., SCHERER, K. R. The role of intonation in emotional expressions. **Speech Communication**, nº 46, p. 352-267, 2005
- BARBOSA, P. A. Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. **Revista de Estudos da Linguagem**, v. 20, n. 1, p. 11-27, 2012. <<https://doi.org/10.17851/2237-2083.20.1.11-27>>
- BARGH, J. What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. **European Journal of Social Psychology**, 36(2), p. 147-168, 2006.
- BARRETO, D. A. R. J.; COSTA, T.; MIELO, S.; MASSINI-CAGLIARI, G. O prefixo 'des-' e palavras de valor negativo no período arcaico da língua portuguesa, vistos a partir de uma breve análise das Cantigas de Santa Maria. **Filologia e Linguística Portuguesa**, v. 21, n. 2, p. 211-224 2019.

BELISARIO, A.; SANTOS, A. Correlação não é causalidade, mas o que é então? **Escola de Dados**, s. d. Disponível em: <https://escoladedados.org/tutoriais/correlacao-nao-e-causalidade-mas-o-que-e-entao/>. Acesso em 28 março 2022.

BOERSMA, P.; WEENINK, D. **Praat**: doing phonetics by computer. *Software*. Versão 6.1.16, 2020. Disponível em <<http://www.praat.org/>>. Acesso em 05 julho 2020.

BOWLING, D. L.; GILL, K.; CHOI, J. D.; PRINZ, J.; PURVES, D. Major and minor music compared to excited and subdued speech. **The Journal of the Acoustical Society of America**, v. 127, p. 491-503, 2010.

BOWLING, D. L.; SUNDARARAJAN, J.; HAN, S.; PURVES, D. Expression of Emotion in Eastern and Western Music Mirrors Vocalization. **PLoS ONE**, v. 7, n. 3, 2012.

BRIGHT, William. Language and Music: Areas for Cooperation. **Ethnomusicology**, [S.l.], v. 7, n. 1, p. 26-32, jan. 1963.

BRITANNICA, The Editors of Encyclopaedia. "doctrine of the affections". **Encyclopedia Britannica**, 24 mar. 2014. Disponível em <<https://www.britannica.com/art/doctrine-of-the-affections>>. Acesso em 09 março 2022.

BRITANNICA, The Editors of Encyclopaedia. "scat". **Encyclopedia Britannica**, 6 ago. 2012. Disponível em: <<https://www.britannica.com/art/scat-music>>. Acesso em 30 março 2022.

BUELOW, G. Affects, theory of the. **Grove Music Online**, 2001. Disponível em <<https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000000253>> Acesso em 09 de março de 2022.

CAGLIARI, L. C. **Elementos de Fonética do Português Brasileiro**. Tese de livre-docência. UNICAMP, 1981.

CAGLIARI, L. C. Marcadores prosódicos na escrita. In: Estudos Lingüísticos – Seminários do GEL, 18., 1989, Lorena, SP. **Anais...** Lorena, SP: GEL, p. 195-203, 1989.

CAGLIARI, L. C. **Alfabetização & linguística**. São Paulo: Scipione, 1997.

CAGLIARI, L. C. Prosódia: Ontem e Hoje. In: FONSECA-SILVA, M. C., PACHECO, V., LESSA-DE-OLIVEIRA, A. S. C. L. (Org.) **Em Torno da Língua(gem):** Questões e Análises. Vitória da Conquista: Edições Uesb, 2007.

CAGLIARI, L. C. **Análise fonológica:** introdução à teoria e à prática: com especial destaque para o modelo fonêmico. São Paulo: Mercado das Letras, 2002.

CAGLIARI, L. C. Marcadores prosódicos na escrita de obras literárias. In: Encontro Nacional da Associação Portuguesa de Linguística, 18., 2003, Lisboa. **Actas...** Lisboa: Colibri, p. 223-232, 2003.

CAGLIARI, L. C. **Alfabetizando sem o bá-bé-bi-bó-bu**. São Paulo: Scipione, 2008.

CARRIE, E. ‘British is professional, American is urban’: attitudes towards English reference accents in Spain. **International Journal of Applied Linguistics (United Kingdom)**, v. 27, n.2, p. 427-447, 2017.

CESPEDES-GUEVARA, J., EEROLA, T. Music communicates affects, not basic emotions - A constructionist account of attribution of emotional meanings to music. **Frontiers in Psychology**, v. 9, 2018. n.p.

CHRISTINER, M., REITERER, S. M. A Mozart is not a Pavarotti: singers outperform instrumentalists on foreign accent imitation. **Frontiers in Human Neuroscience**, v. 9, artigo 482, 2015.

COOK, N. D.; FUJISAWA, T.; TAKAMI, K. Application of a Psycho-acoustical Model of Harmony to Speech Prosody. In: BEL, B.; MARLIEN, I. (Eds.). **Proceedings of Speech Prosody**, p. 147- 150, 2004.

COOKE, D. **The Language of Music**. Londres: Oxford University Press, 1959.

CÓRDULA, M. S. M. **Entoação e sentidos: análise fonético-fonológica dos padrões entoacionais do português brasileiro e do inglês norte-americano no filme Shrek (2001)**. São Paulo: Cultura Acadêmica, 2013.

COSTA, D. S. **A interface música e linguística como instrumental metodológico para o estudo da prosódia do português arcaico**. Tese (Doutorado em em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras de Araraquara, Universidade Estadual Paulista, 2010.

COSTA, D. S. **Estudo do Acento Lexical no Português Arcaico por meio das Cantigas de Santa Maria**. Dissertação (Mestrado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras de Araraquara, Universidade Estadual Paulista, 2006.

COUPER-KUHLEN, E. **An Introduction to English Prosody**. Londres: Edward Arnold, 1986.

COUTINHO, E., DIBBEN, N. Psychoacoustic cues to emotion in speech prosody and music. In: **Cognition and Emotion**, v. 27, n. 4, p. 658-684, 2013.

CRYSTAL, D. **A Dictionary of Linguistics and Phonetics**. 6ª. edição. Oxford: Blackwell Publishing, 2008.

CUMMINS, F. Joint speech: The missing link between speech and music? **Percepta - Revista de Cognição Musical**, v. 1, n. 1, p. 17–32, 2013. < [https://doi.org/10.34018/2318-891x.1\(1\)17-32](https://doi.org/10.34018/2318-891x.1(1)17-32)>

CURTIS, M. E.; BHARUCHA, J. J. The Minor Third Communicates Sadness in Speech, Mirroring Its Use in Music. **Emotion**, v. 10, n. 3, 2010. Disponível em <http://ase.tufts.edu/psychology/music-cognition/pdfs/Curtis&Bharucha2010Emotion.pdf>. Acessado em 28/07/2017.

D’ALESSANDRO, C.; ROSSET, S.; ROSSI, J.-P. The pitch of short-duration fundamental frequency glissandos. **Journal of the Acoustical Society of America**, v. 104, n. 4, out. 1998, p. 2339-2348.

DAY-O'CONNELL, J. Speech, Song, and the Minor Third. **Music Perception: An Interdisciplinary Journal**, v. 30(5), p. 441–462, 2013.

DEGÉ, F., SCHWARZER, G. The effect of a music program on phonological awareness in preschoolers. **Frontiers in Psychology**, v. 2, artigo 124, 2011.

DEUTSCH, D.; HENTHORN, T.; LAPIDIS, R. Illusory transformation from speech to song. **The Journal of the Acoustical Society of America**, v. 129, n. 4, p. 2245–2252, 2011.

DRABKIN, W. Major (i). **Grove Music Online**. Oxford University Press, 2001a. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.17505>>. Acesso em 03 de março de 2021.

DRABKIN, W. Minor (i). **Grove Music Online**. Oxford University Press, 2001b. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.18743>>. Acesso em 03 de março de 2021.

DRABKIN, W. Scale (Fr. gamme; Ger. Tonleiter; It. gamma). **Grove Music Online**. Oxford University Press, 2001c. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.24691>>. Acesso em 03 de março de 2021.

DRABKIN, W. Third (Fr. tierce; Ger. Terz; It. terza). **Grove Music Online**. Oxford University Press, 2001d. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.27849>>. Acesso em 03 de março de 2021.

EBERHARD, D. M.; SIMONS, G. F.; FENNIG, C. D. (Eds.). **Ethnologue: Languages of the World**. 24ª edição. Dallas: SIL International, 2021. Disponível em: <<http://www.ethnologue.com>>. Acesso em 10 de junho de 2021.

ERIKSSON, M. Which Languages Are The Most Beautiful? **Babbel Magazine**. 03 de janeiro de 2020. Disponível em <<https://www.babbel.com/en/magazine/most-beautiful-languages>> Acesso em 12 de maio de 2021.

ESCOFFIER, N.; ZHONG, J.; SCHIRMER, A.; QIU, A. Emotional expressions in voice and music: Same code, same effect?. **Human Brain Mapping**, v. 34 (8), p. 1796-1810, 2013.

EWELLUL, M. Is Major Really Happy and Minor Sad? **School of Composition**. [2020?] Disponível em: <<https://www.schoolofcomposition.com/is-major-happy-and-minor-sad/>> Acesso em 16 de maio de 2021.

FALK, D. **Finding Our Tongues: Mothers, Infants & the Origins of Language**. Nova York: Basic Books, 2009.

FAVARO, G. S.; CANGEMI, A. C. F. G. A.; MASSINI-CAGLIARI, G. Uso de corpus poético para estudos de natureza fonológica no português arcaico: o caso da elisão. **Letras**, n. 60, p. 157-178, 2020.

FERNANDES, F. R. Tonal association in neutral and subject-narrow-focus sentences in Brazilian Portuguese: a comparison with European Portuguese. **Journal of Portuguese Linguistics**, v. 5, n. 1, p. 91-115, 2007.

FERNANDES-SVARTMAN, F. R. Acento secundário, atribuição tonal e ênfase em português brasileiro (PB). **Estudos Linguísticos**, v. 38, n. 1, p. 47-58, 2009.

FERREIRA NETTO, W. O acento na língua portuguesa. In: ARAÚJO, G.A. (org.) **O acento em português**. Abordagens fonológicas, v. 1, n. 1, p. 21-36, 2007.

FERREIRA NETTO, W. ExProsodia. **Revista da Propriedade Industrial–RPI**, v. 2038, 2010

FERREIRA NETTO, W.; CONSONI, F. Estratégias prosódicas da leitura em voz alta e da fala espontânea. **ALFA: Revista de Linguística**, v. 52, n. 2, 2009.

FERREIRA NETTO, W.; PERES, D. O. A variação tonal em português brasileiro a partir de frases espontâneas e textos lidos. **Estudos Linguísticos**, São Paulo, 38 (1): 9-19, jan.-abr. 2009.

FERREIRA NETTO, W.; PERES, D. O.; MARTINS, M. M.; VIEIRA, M. F. Automatic analysis of emotional intonation in Brazilian Portuguese. **The Journal of the Acoustical Society of America**, v. 135, n. 4, p. 2197-2198.

FERREIRA NETTO, W.; SOUZA, A. R.; VIEIRA, M.; PERES, D.; MARTINS, M. Desenvolvimento de algoritmo de análise automática da curva de frequência por meio de convoluções gaussianas do histograma de alturas. In: IV COLÓQUIO BRASILEIRO DE PROSÓDIA DA FALA. **Caderno de resumos**. Maceió: Universidade Federal de Alagoas, Faculdade de Letras, p. 235-240, 2013.

FONTE, J. S.; MASSINI-CAGLIARI, G. As vogais médias tônicas do português do século XVII a partir das rimas do cancioneiro barroco A Fénix Renascida. **Alfa: Revista de Linguística**, v. 65, 2021.

FROTA, S.; CRUZ, M.; FERNANDES-SVARTMAN, F.; COLLISCHONN, G.; FONSECA, A.; SERRA, C.; OLIVEIRA, P.; VIGÁRIO, M. Intonational variation in Portuguese: European and Brazilian varieties. In: Frota, S.; Prieto, P. (Eds). **Intonation in Romance**. Oxford: Oxford University Press, p. 235-283, 2015.

FRÜHHOLZ, S.; TROST, W.; GRANDJEAN, D. The role of the medial temporal limbic system in processing emotions in voice and music. In: **Progress in Neurobiology**, 123, p. 1-17, 2014.

FUJISAWA, T.; COOK, N. D. **Identifying Emotion in Speech Prosody Using Acoustical Cues of Harmony**. INTERSPEECH 2004 - ICSLP, 8th International Conference on Spoken Language Processing, Jeju Island, Korea, 2004.

GOSSELIN, N.; PAQUETTE, S.; PERETZ, I. Sensitivity to musical emotions in congenital amusia. **Cortex**, v. 71, p. 171-192, 2015.

GOUVEIA, R. Medidas de Dispersão. **Toda Matéria**, s. d. Disponível em: <https://www.todamateria.com.br/desvio-padrao/>. Acesso em 28 de março 2022.

GUTURAL. In: **Dicionário Priberam da Língua Portuguesa**. Lisboa: Priberam Informática S.A, 2001. Disponível em: <<https://dicionario.priberam.org/gutural>>. Acesso em 27 de maio de 2021.

HART, J. 'T., 1976, Psychoacoustic backgrounds of pitch contour stylisation. **IPO Annual Progress Report 11**. Eindhoven: Institute for Perception Research, 1976, p. 11-19.

HAYNES, B.; COOKE, P. Pitch. **Grove Music Online**. Oxford University Press, 2001. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.40883>>. Acesso em 16 de maio de 2021.

HILEY, D. Clef(i) (from Lat. clavis: 'key'). **Grove Music Online**. Oxford University Press, 2001. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.05927>>. Acesso em 26 de maio de 2021.

HILTON, N.; GOOSKENS, C.; SCHÜPPERT, A.; TANG, C. Is Swedish more beautiful than Danish? Matched guise investigations with unknown languages. **Nordic Journal of Linguistics**, p. 1-19, 2021.

HYER, B. Key (Fr. ton; Ger. Tonart; It. tono). **Grove Music Online**. Oxford University Press, 2001. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.14942>>. Acesso em 26 de maio de 2021.

INTERVAL Calculator. **Musictheory.net**. Disponível em: <<https://www.musictheory.net/calculators/interval>>. Acesso em 23 mar. 2022.

JACKENDOFF, R.; LERDAHL, F **A Deep Parallel Between Music and Language**. Bloomington: Indiana University Linguistics Club, 1980.

JACKENDOFF, R.; LERDAHL, F. Generative Music Theory and Its Relation to Psychology. In: **Journal of Music Theory**, Vol. 25, No. 1, 25th Anniversary Issue, p. 45-90, 1981.

JEKIEL, M. Comparing rhythm in speech and music: The case of English and Polish. **Yearbook of the Poznań Linguistic Meeting 1**, p. 55–71, 2014.

JUSLIN, P. N.; LAUKKA, P. Communication of emotions in vocal expression and music performance: Different channels, same code?. **Psychological Bulletin**. American Psychological Association (APA), v. 129, n. 5, p.770-814, 2003.

KHAN ACADEMY BRASIL. Amplitude, variância e desvio padrão como medidas de dispersão. **YouTube**, 26 setembro 2014a. Disponível em: <https://youtu.be/UdrtnBGSeSw>. Acesso em 28 março 2022.

KHAN ACADEMY BRASIL. Correlação e causalidade. **YouTube**, 22 outubro 2014b. Disponível em: <https://youtu.be/K4MXPV7jX8Q>. Acesso em 28 março 2022.

KASTNER, M. P.; CROWDER, R. G. Perception of the Major/Minor Distinction: IV . Emotional Connotations in Young Children. **Music Perception: An Interdisciplinary Journal**, v. 8, n. 2, p. 189–201, 1990.

KOGAN, V. V.; REITERER, S. M. Eros, Beauty, and Phon-Aesthetic Judgements of Language Sound. We Like It Flat and Fast, but Not Melodious. Comparing Phonetic and Acoustic Features of 16 European Languages. **Frontiers in Human Neuroscience**, v. 15, 2021. n.p.

KOLINSKY, R.; LIDJI, P.; PERETZ, I.; BESSON, M.; MORAIS, J. Processing interactions between phonology and melody: Vowels sing but consonants speak. **Cognition**, v. 112, n. 1, p. 1–20, 2009. Disponível em <<https://doi.org/10.1016/j.cognition.2009.02.014>>. Acesso em 25 de fevereiro de 2021.

KOMANIECKI, R. **Analyzing the Parameters of Flow in Rap Music**. Dissertação (Doctor of Philosophy) - Jacobs School of Music, Indiana University, 2019.

LAZZARI, M.; PETRY, I. Disfluência e narratividade: a importância do olhar prosódico sobre a aquisição da linguagem. **Revista da ABRALIN**, v. 19, n. 2, p. 1-5, 2020.

LEHISTE, I. **Suprasegmentals**. Cambridge : The M.I.T. Press, 1970.

LEVMAN, B. G. - The Genesis of Music and Language. **Ethnomusicology**, Vol. 36, No. 2, p. 147-170. University of Illinois Press on behalf of Society for Ethnomusicology, 1992.

LIBERMANN, M. **The intonational system of English**. 1975. Tese (Doutorado em Linguística) - Massachusetts Institute of Technology, Cambridge, 1975.

LIMA, C. F.; CASTRO, S. L. Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. **Emotion**, v. 11(5), p. 1021-1031, 2011.

LINDLEY, M. Interval. **Grove Music Online**. Oxford University Press, 2001. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.13865>>. Acesso em 16 de maio de 2021.

LONDON, J. Rhythm (from Gk. *rhythmos*; Lat. *rhythmus*; Fr. *rythme*; Ger. *Rhythmus*; 16th-, 17th-century Eng. *rithme*). **Grove Music Online**. Oxford University Press, 2001. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.45963>>. Acesso em 16 de maio de 2021.

LUNDQVIST, D.; LITTON, J. E. The Averaged Karolinska Directed Emotional Faces - AKDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998. ISBN 91-630-7164-9.

MA, W.; THOMPSON, W. F. Human emotions track changes in the acoustic environment. **Psychological And Cognitive Sciences**, v. 112, n. 47, 2015.

MACHADO, A. L. **Diferenças na percepção de emoções em enunciados verbais declamados e cantados**. Dissertação (mestrado em Linguística e Língua Portuguesa) - Faculdade de Ciências e Letras de Araraquara, Universidade Estadual Paulista, 2017.

MAEDA, S. **A Characterization of American English Intonation**. Tese (Doutorado em Engenharia Elétrica) – Massachusetts Institute of Technology – M.I.T., Cambridge, 1976.

MAGDICS, K. From the Melody of Speech to the Melody of Music. **Studia Musicologica Academiae Scientiarum Hungaricae**, v. 4, n. 3, p. 325–346, 1963.

MANTELL, J. T.; PFORDRESHER, P. Q. Vocal imitation of song and speech. **Cognition**, v. 127, n. 2, p. 177-202, 2013.

MATHUR, A.; VIJAYAKUMAR, S.; CHAKRABARTI, B.; SINGH, N. Emotional responses to Hindustani *raga* music: the role of musical structure. **Frontiers in Psychology**, v. 6, artigo 513, 2015. n.p.

MARASCO, L. What Makes French Sound Sexy? **Mental Floss**. 04 de fevereiro de 2016. Disponível em <<https://www.mentalfloss.com/article/74461/what-makes-french-sound-sexy>>. Acesso em 12 de março de 2021.

MARTINS, M. V. M.; FERREIRA NETTO, W. Proposal of description for an intonation pattern: The simulacrum of neutral intonation. **The Journal of the Acoustical Society of America**, v. 141, n. 5, p. 3701, 2017.

MASSINI-CAGLIARI, G. **Acento e Ritmo**. São Paulo: Contexto, 1992.

MASSINI-CAGLIARI, G. **Cantigas de amigo**: do ritmo poético ao lingüístico. Um estudo do percurso histórico da acentuação em Português. 1995. Tese (Doutorado em Lingüística)- IEL/UNICAMP, Campinas, 1995.

MASSINI-CAGLIARI, G. **Do poético ao lingüístico no ritmo dos trovadores**: três momentos da história do acento. Araraquara: FCL, Laboratório Editorial, UNESP; São Paulo: Cultura Acadêmica, 1999.

MASSINI-CAGLIARI, G. **O texto na alfabetização**: coesão e coerência. Campinas: Mercado das Letras, 2001.

MASSINI-CAGLIARI, G. **A música da fala dos trovadores**: Desvendando a prosódia medieval. São Paulo: Editora Unesp Digital, 2015.

MASSINI-CAGLIARI, G. Chapter Four Sandhi: A Comparative Study Between Archaic And Brazilian Portuguese. **Contemporary Phonology in Brazil**, 2021.

MASSINI-CAGLIARI, G.; CAGLIARI, L. C. Fonética. IN MUSSALIM, F., BENTES, A. C. **Introdução à Lingüística: domínios e fronteiras**, v. 1, p. 105-146. São Paulo: Cortez, 2001.

MATTHESON, J.; LENNEBERG, H. Johann Mattheson on Affect and Rhetoric in Music (I). **Journal of Music Theory**, v. 2, n. 1, p. 47-84, 1958.

MEDEIROS, B. R. **Descrição comparativa de aspectos fonético-acústicos selecionados da fala e do canto em português brasileiro**. Tese (Doutorado em Linguística). Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2002.

MEDEIROS, B. R. Ritmo na língua e na música: o elo possível. **Música em Perspectiva**, v. 2, n. 2, 2010.

MEDEIROS, B. R.; CABRAL, J. P. Acoustic distinctions between speech and singing: Is singing acoustically more stable than speech? **Proceedings of the International Conference on Speech Prosody**, p. 542–546, 2018.

MEDEIROS, B. R., CABRAL, J. P., MEIRELES, A. R., BACETI, A. A. A comparative study of fundamental frequency stability between speech and singing. **Speech Communication**, v. 128, p. 15–23, 2021. <https://doi.org/10.1016/j.specom.2021.02.003>

MEDEIROS, B. R.; CUMMINS, F. Speech and song synchronization: A comparative study. **Proceedings of the International Conference on Speech Prosody**, p. 748–751, 2014.

MEIRELES, A. R.; MEDEIROS, B. R.; CABRAL, J. P. Voice quality comparison between MPB singing and speech. **Sefo**, p. 3–5, 2020.

MEIRELES, A. R.; SIMÕES, A. R.; RIBEIRO, A. C.; MEDEIROS, B. R. Musical Speech: a New Methodology for Transcribing Speech Prosody. **Proc. Interspeech**, p. 334-338, 2017.

MERRILL, J.; LARROUY-MAESTRI, P. Vocal features of song and speech: Insights from Schoenberg's Pierrot lunaire. **Frontiers in Psychology**, v. 8, 2017. n.p.

MERTENS, P. **Prosogram user's guide (Oct 18, 2020)**. 2020. Disponível em: <<https://sites.google.com/site/prosogram/home>> Acesso em 10 de março de 2022.

MERTENS, P. The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. In BEL, B.; MARLIEN, I. (eds.) **Proceedings of Speech Prosody 2004**.

MILOVANOV, R., TERVANIEMI, M. The Interplay between Musical and Linguistic Aptitudes: A Review. **Frontiers in Psychology**, v. 2, artigo 321, 2011. n.p.

MORTON, J. B, TREHUB, S. E. Children's Understanding of Emotion in Speech. **Child Development**, v. 72, n. 3, p. 834-843. Wiley on behalf of the Society for Research in Child Development, 2001.

O'NEILL, P.; MASSINI-CAGLIARI, G. A discriminação e preconceito linguísticos no português brasileiro e outras línguas: sugestões e recomendações. **Revista Diálogos**, v. 9, n. 3, 2021.

PACHECO, V. Leitura e Prosódia: o Caso dos Sinais de Pontuação. In: FONSECA-SILVA, M. C., PACHECO, V., LESSA-DE-OLIVEIRA, A. S. C. L. (Org.) **Em Torno da Língua(gem): Questões e Análises**. Vitória da Conquista: Edições Uesb, 2007.

PAQUETTE, S.; TAKERKART, S.; SAGET, S.; PERETZ, I.; BELIN, P. Cross-classification of musical and vocal emotions in the auditory cortex. **Annals of the New York Academy of Sciences**, v. 1423, n. 1, p. 329-337, 2018.

PARK, M.; GUTYRCHIK, E.; WELKER, L.; CARL, P.; PÖPPEL, E., ZAYTSEVA, Y.; MEINDL, T.; BLAUTZIK, J. REISER, M.; BAO, Y. Sadness is unique: neural processing of emotions in speech prosody in musicians and non-musicians. **Frontiers in Human Neuroscience**, v. 8, 2015. n.p.

PATEL, A. D. The relationship of music to the melody of speech and to syntactic processing disorders in aphasia. **Annals of the New York Academy of Sciences**, v. 1060, n. 1, p. 59–70, 2005.

PATEL, A.D., IVERSEN, J.R., ROSENBERG, J.C. Comparing the rhythm and melody of speech and music: The case of British English and French. **Journal of the Acoustical Society of America**, n. 119, p. 3034–3047, 2006.

PERES, D. O. **A percepção da emoção na fala por nativos e não nativos**. Tese (Doutorado em Letras) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2016.

PIAZZA, E.A.; SWEENEY, T.D.; WESSEL, D.; SILVER, M.A.; WHITNEY, D. Humans Use Summary Statistics to Perceive Auditory Sequences. **Psychological Science**, v. 24(8), p.1389-1397, 2013.

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

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

PINKER, S. **How the mind works**. Nova York: W. W. Norton & Co., 1997.

POWERS, H. S.; WIERING, F.; PORTER, J.; COWDERY, J.; WIDDESS, R.; DAVIS, R. PERLMAN, M.; JONES, S.; MARETT, A. Mode (from Lat. modus: ‘measure’, ‘standard’ ; ‘manner’, ‘way’). **Grove Music Online**. Oxford University Press, 2001. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.43718>>. Acesso em 26 de maio de 2021.

QUINTO, L.; THOMPSON, W. F.; KEATING, F. L. Emotional communication in speech and music: the role of melodic and rhythmic contrasts. **Frontiers in Psychology**, v. 4, 2013. n.p.

RAIN, R. Music Interval Calculator. **Omni Calculator**, 2021. Disponível em: <<https://www.omnicalculator.com/other/music-interval>>. Acesso em 23 mar. 2022.

REITERER, S. M.; KOGAN, V.; SEITHER-PREISLER, A.; PESEK, G. Foreign language learning motivation: Phonetic chill or Latin lover effect? Does sound structure or social stereotyping drive FLL? **Psychology of Learning and Motivation** - Advances in Research and Theory, v. 72, p. 165–205, 2020.

RINGER, A. Melody. **Grove Music Online**. Oxford University Press, 2001. Disponível em: <<https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000018357>>. Acesso em 27 de março de 2022.

SALOMÃO, G. L. Expressão Vocal de Emoções: Metáforas Sonoras, Fala e Canto. In: MADUREIRA, S. (Org.) **Sonoridades: a Expressividade na Fala, no Canto e na Declamação**. São Paulo: Pontifícia Universidade Católica de São Paulo, p. 31-43 2016

SANTOS, C. A. A. A. **Vogais cantadas e tonicidade: estudo experimental comparativo entre fala e canto com foco na duração**. Tese (Doutorado em Linguística) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2017.

SANTOS, V. G.; FERNANDES-SVARTMAN, F. R. O padrão entoacional neutro do português de Guiné-Bissau: uma comparação preliminar com o português brasileiro. **Estudos Linguísticos**, v. 43, n. 1, p. 48-63, 2014.

SCARPA, E. M. Sobre o sujeito fluente. **Cadernos de estudos linguísticos**, v. 29, p. 163-184, 1995.

SCARPA, E. M. Sons preenchedores e guardadores de lugar: relações entre fatos sintáticos e prosódicos na aquisição da linguagem. In: E. M. Scarpa (Ed.) **Estudos de Prosódia**, p. 253-284. Campinas: Ed. Unicamp, 1999.

SCARPA, E. M. A criança e a prosódia: uma retrospectiva e novos desenvolvimentos. **Cadernos de estudos linguísticos**, v. 47, n. 2, p. 19-28, 2005.

SHELLENBERG, M. Does Language Determine Music in Tone Languages? **Ethnomusicology**, vol. 56, nº 2, p. 266-278, 2012.

SCHERER, K. R. Emotion as a multicomponent process: A model and some cross-cultural data. **Review of Personality and Social Psychology**, v. 5, p. 37-63, 1984.

SCHERER, K. R. A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology. **International Conference on Spoken Language Processing Proceedings of ICSLP 2000**, Beijing, China, 2000.

SCHERER, K. R. The affective and pragmatic coding of prosody. In: HUO, Q.; MA, B.; CHNG, E. S.; LI, H. (eds). **ISCSLP 06 Proceedings of the 5th International Conference on Chinese Spoken Language Processing**, p. 13-14, 2006.

SCHERER, K. R., BANSE, R., & WALLBOTT, H. G. Emotion inferences from vocal expression correlate across languages and cultures. **Journal of Cross-Cultural Psychology**, v. 32, n. 1, p. 76–92, 2001.

SCHERER, K. R., BANSE, R., WALLBOTT, H. G., GOLDBECK, T. Vocal cues in emotion encoding and decoding. **Motivation and Emotion**, v. 15, n. 2, p. 123–148, 1991.

SCHERER, K. R.; LADD, D. R., SILVERMAN, K. E. A. Vocal cues to speaker affect: Testing two models. In: **The Journal of the Acoustical Society of America** 76, 1346. Acoustical Society of America, 1984.

SCHREUDER, M. **Prosodic Processes in Language and Music**. Tese (Doutorado em Letras), Rijksuniversiteit Groningen, 2006.

SCHREUDER, M.; VAN EERTEN, L.; GILBERS, D. Speaking in Minor and Major Keys. **Proceedings of Speech Prosody**. Nara, Japão, p. 341-344, 2004.

SCHÜPPERT, A., HILTON, N. H., & GOOSKENS, C. Swedish is beautiful, Danish is ugly? Investigating the link between language attitudes and spoken word recognition. **Linguistics**, v. 53, n. 2, p. 375–403, 2015.

SEELINGER, L. The 10 Most Beautiful Languages in the World. **Culture Trip**. 28 de março de 2018. Disponível em: <<https://theculturetrip.com/europe/italy/articles/the-10-most-beautiful-languages-in-the-world/>> Acesso em 12 de maio de 2021.

SILVA, T. C. **Fonética e Fonologia do Português**: Roteiro de estudos e guia de exercícios. 7. ed. São Paulo: Contexto, 2003.

SILVA, W. **Correlatos prosódicos da expressão da ironia sarcástica no português brasileiro**. Tese (Doutorado em Linguística) - Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, 2019.

SILVA, W., BARBOSA, P. A., ABELIN, Å. Cross-cultural and cross-linguistic perception of authentic emotions through speech: An acoustic-phonetic study with Brazilian and Swedish listeners. **DELTA: Documentação de Estudos Em Linguística Teórica e Aplicada**, v. 32, n. 2, p. 449–480, 2016.

SIMÕES, A. R. M.; MEIRELES, A. R. Speech Prosody in Musical Notation: Spanish, Portuguese and English. **Proc. Speech Prosody 2016**, p. 212-216, 2016.

SLEVC, L. R. Language and music: sound, structure, and meaning. **WIREs Cogn Sci**, v. 3, p. 483–492, 2012.

SOUZA, A. R. **A intenção na melodia: estudo interdisciplinar sobre as relações entre entoação e gênero de discurso nas manifestações vocais (da fala ao canto)**. Tese (doutorado em Letras) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2014.

SRULL T.K., WYER R.S., JR. The role of category accessibility in the interpretation of information about persons: Some determinants and implications. **Journal of Personality and Social Psychology**, n. 37, p. 1660–1667, 1979.

STEINBEINS, N.; KOELSCH, S. Shared Neural Resources between Music and Language Indicate Semantic Processing of Musical Tension-Resolution Patterns. In: **Cerebral Cortex**, vol. 18(5), p. 1169-1178, 2008.

STEVENS, C. J. Music Perception and Cognition: A Review of Recent Cross-Cultural Research. **Topics in Cognitive Science**, v. 4, n. 4, p. 653-667, 2012.

SUBRAMANIAN, S. On a Sad Note. **Mint**, 12 de novembro de 2010. Disponível em: <<https://www.livemint.com/Leisure/Km0Py6jSEG8x7KXmzCKX1N/On-a-sad-note.html>> Acesso em 10 de março de 2022.

TAN, S.L.; PFORDRESHER, P.; HARRÉ, R. **Psychology of Music: From Sound to Significance**. E-book. Psychology Press, 2010.

THE VOICE PROJECT. **Lullabies of the World**, 2013. Disponível em: <<https://lullabiesoftheworld.org/>> Acesso em 10 de junho de 2021.

THOMPSON, W. F.; MARIN, M. M.; STEWART, L. Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 46, p. 19027–19032, 2012.

THOMPSON, W. F.; BALKWILL, L. Decoding speech prosody in five languages. **Semiotica**, v. 2006, n. 158, p.407-424, 2006.

TOLBERT, E. The Enigma of Music, the Voice of Reason: "Music," "Language," and Becoming Human. **New Literary History**, vol. 32, nº 3, Voice and Human Experience, 2001.

VARIÂNCIA e desvio padrão. **Toda Matéria**, s.d. Disponível em: <https://www.todamateria.com.br/variancia-e-desvio-padrao/>. Acesso em 28 março de 2022.

VASCONCELOS, A. N.; VIEIRA, N.; SCARPA, E. M. A constituição prosódica da enunciação na relação mãe-bebê. **Bakhtiniana**, São Paulo, vol. 16, nº 1, p. 39-60, 2021

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

VOS, P. G., TROOST, J. M. Ascending and Descending Melodic Intervals: Statistical Findings and Their Perceptual Relevance. **Music Perception: An Interdisciplinary Journal**, v. 6, n. 4, p. 383-396, 1989.

WHAT Makes a Language Sound Beautiful or Ugly? **BeTranslated**, Valencia, 19 ago 2017. Disponível em: <<https://www.betranslated.com/blog/makes-language-sound-beautiful-ugly/>>. Acesso em 12 maio 2021.

WATTENBERG, D. What Makes a Song Sad. **The Atlantic**. 10 de dezembro de 2010. Disponível em: <<https://www.theatlantic.com/entertainment/archive/2010/12/what-makes-a-song-sad/67709/>> Acesso em 18 de maio de 2021.

WILLIAMS, C. E.; STEVENS, K. N. Emotions and Speech: Some Acoustical Correlates. **Journal of The Acoustical Society Of America**, v. 52, n. 4, p.1238-1250, 1972.

WILSON, S. 7 times Björk used cutting-edge technology to shape her music. **Fact Magazine**, 24 de novembro de 2017. Disponível em <<https://www.factmag.com/2017/11/24/bjork-technology-instruments-software/>> Acesso em 10 de março de 2022.

WILSON, B.; BUELOW, G. J.; HOYT, P. A. Rhetoric and music. **Grove Music Online**. Oxford University Press, 2001. Disponível em: <<https://doi.org/10.1093/gmo/9781561592630.article.43166>>. Acesso em 25 de março de 2022.

WÖLLNER, C. Is empathy related to the perception of emotional expression in music? A multimodal time-series analysis. **Psychology of Aesthetics, Creativity, and the Arts**, v. 6, n.3, p. 214-223, 2012.

ZBIKOWSKI, L. M.. Music, Language, and What Falls in Between. **Ethnomusicology**, v. 56, n. 1, p.125-131, 2012

## 7. APÊNDICES

### APÊNDICE A – QUESTIONÁRIO: PÁGINA INICIAL

#### Pesquisa: fala, emoções e música

Olá, obrigado por estar aqui! Por favor, leia atentamente as instruções a seguir antes de continuar.

Esta é uma pesquisa sobre as emoções que as pessoas percebem em uma melodia, e também sobre quão melodiosas diferentes línguas soam (para pessoas que não são falantes dessas línguas). Você provavelmente vai demorar entre dez e quinze minutos para colaborar, não mais do que isso.

Você vai ouvir quatro gravações de uma pessoa cantando uma melodia de até 30 segundos. Depois de ouvir, selecione a imagem que, na sua opinião, expressa a mesma emoção que o áudio expressa. Não existe resposta certa ou errada, queremos saber apenas qual é a sua percepção.

A seguir, você vai ouvir seis trechos curtos, de até 30 segundos, de pessoas falando em uma língua estrangeira. Selecione, em uma escala de 1 a 5, o quão melodiosa a língua que a pessoa está falando soa para você.

Se, a qualquer momento, você tiver certeza que conhece a língua que está ouvindo e/ou conseguir entender o que está sendo cantado ou dito, selecione a opção "Eu conheço esta língua".

Quando acabar, responda algumas perguntas para que possamos traçar o perfil demográfico das pessoas que participaram. Sua participação nesta pesquisa é voluntária e sua identidade não será divulgada.

**\*POR FAVOR, RESPONDA A APENAS UMA VERSÃO DO FORMULÁRIO\***

Desde já, agradeço sua colaboração!

**\*Obrigatório**

Se você concorda com os termos acima, escolha uma das cores abaixo para começar o teste (não há resposta certa ou errada, a cor apenas decide quais gravações você vai ouvir) \*



Concordo, vamos nessa! Eu escolho a versão vermelha!



Concordo, vamos nessa! Eu escolho a versão azul!

Próxima

## APÊNDICE B – QUESTIONÁRIO: MÚSICA

Versão vermelha - Música 1/4

Qual dos rostos abaixo você acha que expressa a mesma emoção expressa pela música? Se você conhece essa língua e/ou entende as palavras que estão sendo cantadas, selecione "Eu conheço esta língua".

Qual dos rostos expressa a **mesma** emoção que a música expressa?

Lembre-se:  
Não há resposta certa ou errada!



Opção 1



Opção 2



Opção 3



Opção 4



Eu conheço esta língua

**APÊNDICE C – QUESTIONÁRIO: MUSICALIDADE DA LÍNGUA****Pesquisa: fala, emoções e música****\*Obrigatório****Quão melodiosa é essa língua? 1/6**

Em uma escala de 1 a 5, sendo 1 nada musical/melodiosa e 5 muito musical/melodiosa, quão melodiosa e/ou musical a língua falada nessa gravação soa para você?

Quão melodiosa essa língua  
soa para você?

Lembre-se:  
Não há resposta certa ou errada!



1 2 3 4 5

Nada melodiosa / musical

Muito melodiosa / musical

**Você conhece essa língua? \***

- Sim
- Não
- Não tenho certeza

**Que língua você acha que é essa?**

Sua resposta

Voltar

Próxima

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários

## APÊNDICE D – INSTRUÇÕES E CENÁRIOS FORNECIDOS AOS ATORES PARA GRAVAÇÃO

**Instruções:** Cada célula representa um contexto em que o seu personagem sente uma dentre 4 emoções básicas (raiva, alegria, tristeza, simpatia). Para cada gravação, diga o número da cena e a frase em vermelho (por exemplo, “Cena 01, Eu sei.”). Você pode gravar as cenas na ordem que preferir. Pode gravar todas as cenas com a frase “Sou eu”, depois todas com a frase “Tá bom”, e assim por diante, ou pode gravar todas as cenas que envolvem a emoção Raiva, depois todas as que envolvem Alegria. Apenas certifique-se de identificar o número da cena *antes* de falar a frase em vermelho. Se preferir, você pode gravar a descrição toda da cena antes da fala em vermelho, mas apenas o texto em vermelho será considerado para a análise. Se possível, faça esse procedimento duas vezes, em dois dias diferentes (totalizando 32 gravações). Muito obrigado pela sua participação!

### Frase 01 – Aqui.

	Cena 01 - Raiva	Cena 02 - Alegria	Cena 03 - Tristeza	Cena 04 - Simpatia
Sou eu.	Você está em um consultório médico. Sua consulta, agendada há meses, estava marcada para as 14h30, já são 16h00, você tem outro compromisso importante às 16h15 e o médico ainda não chegou. Finalmente, uma secretária mal-educada pergunta onde está o paciente que estava marcado para as 14h30, ao que você responde: <b>- Aqui.</b>	Você acabou de sair da sala do seu chefe. Devido ao seu ótimo desempenho no trabalho, ele te ofereceu uma promoção e um bom aumento salarial. Você dá de cara com um de seus colegas e já quer dar as boas notícias: “Sabe onde está a pessoa que ganhou um ótimo aumento porque vai passar a ocupar o cargo de gerência? <b>Aqui.</b> ”	Sua mãe acaba de falecer e você está no hospital assinando alguns papéis para dar seguimento aos procedimentos. A responsável chega e pergunta se os documentos já estão prontos. Ao entregar os papeis a ela, você diz: <b>- Aqui.</b>	Você é professor em uma universidade e está a caminho para a sua sala quando ouve alguém, provavelmente um aluno, perguntando onde pode encontrar você. Você se vira e percebe que, por algum motivo, ele parece nervoso e intimidado. Tentando acalmá-lo um pouco, você sorri, acena, e responde: <b>-Aqui.</b>

**Frase 02 – Tá bom.**

	Cena 05 - Raiva	Cena 06 - Alegria	Cena 07 - Tristeza	Cena 08 - Simpatia
Tá bom.	Você estava pronto para sair e encontrar seus amigos quando sua mãe te aborda e te dá uma longa bronca por você ainda não ter arrumado o seu quarto. Por mais que você tente argumentar, ela te proíbe de sair e manda você dar um jeito no seu quarto imediatamente. A contragosto, você responde: <b>Tá bom.</b>	Você teve uma noite muito agradável e divertida, jantando num ótimo restaurante com um amigo que você não via há tempos. É hora de ir embora, vocês pedem a conta, e ele insiste em pagar a conta pra você, como um presente. Você aceita e diz: <b>Tá bom.</b>	Você é pai de família, com três filhos pequenos. Seu emprego está por um fio devido à recessão e o iminente corte de pessoal. Seu chefe te fala que você, terá que trabalhar mais horas e que você tem que jogar em time pelo bem da empresa. Desolado e mesmo sabendo que não poderá ver sua filha que tem poucos meses de vida tanto quanto gostaria, você compreende que tem que se manter nesse emprego e responde: <b>Tá bom.</b>	É o começo do expediente, e a nova auxiliar de limpeza do lugar onde você trabalha deixou um frasco de produto de limpeza na sua sala. Logo que você entra, ela chega, afobada, pedindo desculpas por ter deixado o produto na sua sala. Você não se importa, sabe que ela é nova e não causou nenhum problema. Você quer que ela fique tranquila e quando ela se explica e diz que já vai guardar o produto, você responde: <b>Tá bom.</b>

**Frase 03 – Vem cá.**

	Cena 09 - Raiva	Cena 10 - Alegria	Cena 11 - Tristeza	Cena 12 - Simpatia
Vem cá.	Depois de um dia estressante no trabalho, você chega em casa e flagra seu filho mexendo na sua pasta de documentos, sendo que você já se cansou de falar para ele não fazer isso. Seu filho percebe que está encrencado e tenta sair correndo para fugir da bronca, ao que você responde: <b>Vem cá.</b>	Você acaba de receber uma ligação do seu advogado dizendo que ganhou uma causa na justiça. Você quer contar a ótima novidade para alguém e chama a sua colega de trabalho: <b>- Vem cá.</b>	Sua irmã entra na sala e encontra você sentado no sofá. Sua namorada acabou de terminar com você. Sua irmã pergunta qual é o problema, mas você está tão mal que só quer um abraço. Você pede pra ela vir até você: <b>- Vem cá.</b>	Você descobriu que seu irmão foi abandonado pela esposa. Você vai até a casa dele para tentar animá-lo ou, pelo menos, consolá-lo. A porta está aberta, você entra e encontra ele na sala, sentado no sofá. Você se senta do lado dele e percebe que ele precisa de afeto. Sem perguntar nada, simplesmente oferece seu abraço e diz:

			<b>-Vem cá.</b>
--	--	--	-----------------

### Frases 04 – Eu sei.

	Cena 13 - Raiva	Cena 14 - Alegria	Cena 15 - Tristeza	Cena 16 - Simpatia
Eu sei.	Sua sogra está te visitando nesse domingo e passou a manhã te criticando. Ela não se cansa de dar palpites e falar com você como se você não soubesse o que está fazendo. Você está temperando a salada e ela começa a soltar mais uma série de instruções e palpites e, por fim, pergunta se você sabe como temperar direito a salada. Você responde: <b>Eu sei.</b>	A empresa em que você trabalha oferece participação dos lucros aos funcionários. Desta vez, você e seus colegas receberam bem mais dinheiro do que esperavam. A animação é geral e, em uma conversa, um colega menciona que esse dinheiro veio em ótima hora, pois você vai sair de férias no mês que vem e, com isso, vai poder fazer aquela viagem em família que tanto queria. Você responde: <b>Eu sei.</b>	Você está no velório da sua mãe, que morreu após sofrer um ataque cardíaco. Seu melhor amigo tenta te consolar, dizendo que pelo menos sua mãe não passou o fim da vida debilitada, presa a uma cama, e que por mais que seja uma situação triste, ela fica feliz por sua mãe não ter sofrido muito. Você responde: <b>Eu sei.</b>	Você é um terapeuta e seu paciente acaba de te contar sobre um problema familiar pelo qual está passando. Você quer fazê-lo se sentir acolhido e compreendido, e quando ele pergunta se você tem ideia da dor que ele está sentindo, você responde: <b>Eu sei.</b>