UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
FACULDADE DE LETRAS

MARIA LUIZA DE ALMEIDA MATTOS WEINSTEIN

ANÁLISE DA PRODUÇÃO E DA PERCEPÇÃO PROSÓDICA MULTIMODAL NA DUBLAGEM DO FILME *FRAGMENTADO* (2016)

RIO DE JANEIRO

2025

Maria Luiza de Almeida Mattos Weinstein

ANÁLISE DA PRODUÇÃO E DA PERCEPÇÃO PROSÓDICA MULTIMODAL NA DUBLAGEM DO FILME *FRAGMENTADO* (2016)

Monografia submetida à Faculdade de Letras, da Universidade Federal do Rio de Janeiro, como requisito para obtenção do título de Licenciada em Letras na habilitação Português/Inglês.

Orientadora: Profª. Drª. Manuella Carnaval

Coorientador: Prof. Dr. Vitor Gabriel Caldas

RIO DE JANEIRO

2025

FICHA CATALOGRÁFICA

MARIA LUIZA DE ALMEIDA MATTOS WEINSTEIN

DRE: 122055462

ANÁLISE DA PRODUÇÃO E PERCEPÇÃO PROSÓDICA MULTIMODAL NA DUBLAGEM DO FILME *FRAGMENTADO* (2016)

Monografia submetida à Faculdade de Letras da Universidade Federal do Rio de Janeiro, como requisito para obtenção do título de Licenciada em Letras na habilitação Português/Inglês.

Data da avaliação: 13/01/2026

Banca examinadora:

gov.br Documento assinado digitalmente
MANUELLA CARNAVAL
Data: 13/01/2026 18:11:05-0300
Verifique em https://validar.iti.gov.br

NOTA: 10,0

Orientadora: Profª. Drª. Manuella Carnaval

gov.br Documento assinado digitalmente
VITOR GABRIEL CALDAS
Data: 13/01/2026 21:19:56-0300
Verifique em https://validar.iti.gov.br

Coorientador: Prof. Dr. Vitor Gabriel Caldas NOTA: 10,0

gov.br Documento assinado digitalmente
JOAO ANTONIO DE MORAES
Data: 13/01/2026 17:05:55-0300
Verifique em https://validar.iti.gov.br

NOTA: 10,0

Leitor Crítico: Prof. Dr. João Antônio de Moraes

MÉDIA: 10,0

Agradecimentos

Ao meu Deus, dono de toda ciência, sabedoria e poder, agradeço por me carregar, me acolher e me inspirar quando eu achava não ser mais possível seguir com minhas próprias pernas. Ele, que me fortalece do respirar ao escrever, me tirou o medo de continuar por me mostrar que é tudo sobre Ele e não sobre mim.

À minha melhor amiga, minha estrela além do tempo, minha mãe Elizabeth, sou grata pela vida e pela parceria imensurável. Eu não seria nada sem o seu amor e o seu suporte. Saiba que eu te escolheria em todas as vidas.

Aos meus padrinhos mágicos, por todos os ensinamentos, amor, incentivo e por me mostrarem o valor de uma educação pública de qualidade.

Às minhas grandes famílias, por me moldarem como ser humano e por acreditarem em mim desde sempre. Em especial, às minhas avós e às minhas tias, pelas orações a cada dia e pelo acolhimento em todas as situações.

Aos amigos que sempre estiveram e estão lá pela minha mãe e por mim, agradeço por me darem infinitos lugares para chamar de casa.

Aos meus amigos, por todos os desabafos, risadas e incentivos. Obrigada por serem luz na minha caminhada e por transformarem o triste em suportável, a alegria em felicidade e o simples em extraordinário.

Ao meu Robin capiano, por dividir comigo a experiência do estágio e tantas memórias especiais.

Às minhas mosqueteiras, Anny, Clarinha, Gisi, Lina, Livs, Let e Madá, por simplesmente tudo. Vocês são minhas músicas preferidas, minhas cores prediletas, meus números da sorte, meus animais peculiares, minhas melhores fanfics, minhas docinhas. Poder crescer e me formar com uma ministra da educação, uma escritora, uma herdeira, uma avó historiadora, uma babá, uma professora universitária e uma CEO foi uma experiência sensacional.

Ao meu parceiro de iniciação científica, Daniel, pelos últimos dois anos de construção de todas as etapas dessa pesquisa.

Ao meu parceiro de laboratório, Victor Hugo, por toda ajuda, aprendizado e contribuição ao longo destes anos de faculdade.

À minha bússola, meu par mínimo, Prof. Manuella, por me escolher. Agradeço por me enxergar como pessoa e, a partir disso, me constituir como pesquisadora. Obrigada por buscar sempre o melhor em mim.

À nossa Odete Roitman, Prof. Vitor, pelas risadas, pelo experimento e pelas análises estatísticas que contribuíram imensamente para este trabalho.

Ao prof. João, por todos os ensinamentos na pós, pela gentileza, pela paciência e por aceitar contribuir criticamente para a consolidação desta pesquisa.

Aos meus mestres, por me inspirarem, desde cedo, a olhar o mundo através das lentes da educação. Em especial, à Carolina Lopes, por me acolher desde o primeiro dia, e à Monique Ridge, por me ensinar que “Ler é atribuir sentido”.

Aos meus aprendizes, por todos os dias me emprestarem novos óculos para enxergar a vida.

A todos os participantes de experimento e a todos que de alguma forma contribuíram para o crescimento desta pesquisa, obrigada de todo o coração!

Resumo

MATTOS WEINSTEIN, Maria Luiza de Almeida. **Análise da produção e percepção prosódica multimodal na dublagem do filme Fragmentado (2016).** (Graduação em Letras: Português/ Inglês), Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2025.

A prosódia estuda como algo é dito por um locutor e suas implicações no interlocutor. Considerando o processo de atuação e dublagem em obras audiovisuais como um diálogo entre ator/dublador e espectador, este trabalho busca compreender como ocorre a individualização prosódica das personas Dennis, Fera, Hedwig e Patrícia no filme *Fragmentado (2016),* uma vez que todas são interpretadas visualmente pelo mesmo ator e dubladas acusticamente pelo mesmo dublador. O estudo investiga de que modo essa individualização é construída a partir da unidade visual e vocal no processo de dublagem. Especificamente, o estudo objetiva: (1) caracterizar acusticamente, a partir do parâmetro de frequência fundamental (F0), as qualidades de voz performadas pelo dublador, (2) realizar a análise qualitativa dos contornos melódicos das personas, (3) coletar rótulos interpretativos de ouvintes a partir de sua percepção auditiva, por meio de um teste de *free labeling*, (4) validar esses rótulos em um experimento de escolha forçada com um novo grupo de ouvintes, (5) verificar a relevância das modalidades auditiva (AU), visual (VI) e audiovisual (AV) na identificação das personalidades em um experimento de escolha forçada multimodal, (6) analisar a consistência perceptiva entre a interpretação vocal do dublador e a informação visual do produto cinematográfico, observando a sincronia entre pistas prosódicas visuais do ator e acústicas do dublador e (7) analisar qualitativamente a prosódia visual das quatro personas. Para esses fins, foi adotada uma metodologia multidimensional e integrada. No eixo da percepção, foram aplicados três testes perceptivos: um teste de *free labeling*, um teste de escolha forçada e um teste multimodal. No eixo da produção, os áudios foram analisados no *software* Praat (Boersma; Weenink, 1993–2025) para extração de F0, enquanto os vídeos foram analisados no software ELAN, com auxílio do manual FACS – *Facial Action Coding System* (Ekman, Friesen & Hager, 2002) e do manual de rotulagem M3D (Rohrer et alii, 2025). Os testes perceptivos foram aplicados nas plataformas Microsoft Forms e PCIBex (Zehr; Schwarz, 2018). O teste de *free labeling* estabeleceu os rótulos “Calmo” para Dennis, “Raivoso” para Fera, “Brincalhão” para Hedwig e “Cuidadoso” para Patrícia, utilizados tanto no teste de escolha condicionada com áudios quanto no teste multimodal. Os resultados indicaram que as personas mais caricatas, Fera e Hedwig, foram bem identificadas pelos juízes, enquanto Dennis e Patrícia apresentaram índices de acerto não estatisticamente significativos na correspondência persona-rótulo. Na análise da produção acústica, o contorno melódico de Dennis foi caracterizado por *breathy voice*, o de Fera, por alto esforço vocal (Gussenhoven, 2004), o de Hedwig, por início com ataque alto, e o de Patrícia, por modulação *flat*. Na prosódia visual, identificou-se uma gama significativamente maior de unidades de ação por enunciado nas personas perceptualmente mais complexas, Dennis e Patrícia, em comparação a Fera e Hedwig. Os resultados indicam que os rótulos do *free labeling* não capturam plenamente a complexidade de Dennis e Patrícia, não houve padrão consistente na relevância das modalidades sensoriais no teste multimodal e, na produção, há correspondência entre prosódia visual e auditiva, o que não se confirma plenamente na percepção para todas as personas.

PALAVRAS-CHAVE: Prosódia; Expressividade; Percepção; Produção; Multimodalidade; Dublagem.

Abstract

Prosody studies how something is said by a speaker and its implications for the listener. Considering acting and dubbing in audiovisual works as a dialogic process between actor/dubber and spectator, this study seeks to understand how prosodic individualization occurs in the personas Dennis, Fera, Hedwig, and Patrícia in the film *Split (2016),* given that all personas are visually portrayed by the same actor and acoustically dubbed by the same voice actor. The study investigates how such individualization is constructed despite visual and vocal unity. Specifically, the study aims to: (1) acoustically characterize, based on the fundamental frequency (F0) parameter, the voice qualities performed by the voice actor; (2) conduct a qualitative analysis of the melodic contours of the personas; (3) collect interpretive labels from listeners based on their audio perception through a free labeling test; (4) validate these labels in a forced-choice experiment with a new group of listeners; (5) verify the relevance of the audio (AU), visual (VI), and audiovisual (AV) modalities in the identification of personalities in a multimodal forced-choice experiment; (6) analyze perceptual consistency between the voice actor's vocal performance and the visual information of the cinematic product, observing the synchrony between visual prosodic cues produced by the actor and acoustic cues produced by the voice actor; and (7) qualitatively analyze the visual prosody of the four selected personas. To achieve these objectives, a multidimensional and integrated methodology was adopted. In the perception domain, three perceptual tests were applied: a free labeling test, a forced-choice test, and a multimodal test. In the production domain, audio samples were analyzed using the software Praat (Boersma & Weenink, 1993–2025) for F0 extraction, while video samples were analyzed using the software ELAN, with support from the FACS manual – *Facial Action Coding System* (Ekman, Friesen & Hager, 2002) and the M3D labeling manual (Rohrer et alii, 2025). The perceptual tests were conducted using the platforms Microsoft Forms and PCIBex (Zehr & Schwarz, 2018). The free labeling test established the labels "Calm" for Dennis, "Angry" for Fera, "Playful" for Hedwig, and "Careful" for Patrícia, which were used both in the audio-based forced-choice test and in the multimodal test. Results indicated that the more caricatured personas, Fera and Hedwig, were accurately identified by judges, whereas Dennis and Patrícia showed non-statistically significant accuracy rates in persona–label correspondence. In the acoustic production analysis, Dennis's melodic contour was characterized by a breathy voice; Fera's by high vocal effort (Gussenhoven, 2004); Hedwig's by a high-onset attack; and Patrícia's by flat modulation. In visual prosody, a significantly broader range of action units per utterance was identified in the perceptually more complex personas, Dennis and Patrícia, compared to Fera and Hedwig. The results indicate that the free labeling labels do not fully capture the complexity of Dennis and Patrícia, that no consistent pattern emerged regarding the relevance of sensory modalities in the multimodal test, and that, in production, there is correspondence between visual and auditory prosody, which is not fully confirmed in perception for all personas.

KEYWORDS: Prosody; Expressivity; Perception; Production; Multimodality; Dubbing.

"Certamente nós falamos conosco mesmos; não há um ser pensante que não tenha experimentado isso. Alguém até poderia dizer que a palavra é um mistério ainda mais magnífico quando, dentro de um homem, ela viaja de seu pensamento até a consciência, e retorna da consciência ao pensamento."

(Les Misérables ,Victor Hugo, 1862)

Lista de figuras

Lista de gráficos

Lista de quadros

Sumário

1. Introdução

A prosódia apresenta relevância comunicativa tanto na constituição estrutural e sintática da língua quanto em sua construção pragmática do contexto discursivo e situacional. Portanto, a mesma não considera diretamente o conteúdo segmental dos enunciados, mas sim a sua forma sonora e funcionalidade a ela atribuída (Barbosa, 2019), atuando no nível suprassegmental da língua.

Ao abordar a entoação, fenômeno prosódico referente às modulações melódicas da fala (Moraes e Rilliard, 2022), percebe-se as diferentes funcionalidades que esta assume na construção do enunciado. Dessa forma, segundo a categorização de Fónagy (1993), este trabalho se propõe a observar a relevância das funções expressiva – emoções e atitudes intencionadas pelo falante – e identificadora – distinção entre os locutores da enunciação da fala – da prosódia como recursos de dublagem no Português do Brasil, a partir do filme *Fragmentado (2016)*.

A perspectiva da expressividade do discurso, aqui adotada, dialoga com Madureira, Fontes e Camargo (2019), que considera o simbolismo sonoro um recurso intrínseco à função expressiva da prosódia. Este conceito se refere às características atribuídas – por um interlocutor – a um falante a partir das propriedades acústicas e articulatórias de sua voz (Hinton, Nichols e Ohala 1994; Ohala 1997; Abelin, 1999).

Ademais, com base nos estudos de Ohala sobre os códigos de simbolismo sonoro, Gussenhoven (2004) propõe um código biológico baseado no conceito de "esforço vocal". Assim, o maior esforço vocal estaria relacionado a maiores oscilações melódicas, enquanto o menor esforço vocal estaria relacionado a padrões melódicos mais "planos", monótonos, com menor oscilação. Dessa forma, o conceito de simbolismo sonoro estabelece que características como "grande", "forte" e "dominante" são atribuídas, perceptivamente, a vozes mais graves, com menores valores de *pitch*, enquanto simbolismos como "pequeno", "fraco" e "submisso" são atribuídos a vozes mais agudas, com valores mais altos de *pitch*.

A caracterização de um locutor, contudo, não é percebida apenas por sua constituição vocal. A prosódia visual considera que os gestos e expressões faciais contribuem para a construção do significado e da intencionalidade no discurso. Gili Fivela (2018) destaca a união entre os canais auditivo e visual como inerente à situação comunicativa. Ainda, a autora evidencia a relevância da análise multimodal de fenômenos prosódicos para uma perspectiva ampla da produção e percepção da mensagem linguística.

No tangente à multimodalidade, adota-se a perspectiva de Rilliard (2020; *online*) em que "modalidade se refere ao jeito com que os diferentes sentidos participam da percepção da fala: a audição com a modalidade áudio e a visão com a modalidade visual são os dois sentidos principais, razão pela qual também se denomina fala audiovisual." Dessa maneira, a combinação das modalidades auditiva e visual se articulam para a transmissão da mensagem linguística em diálogo com a expressão emocional situacional do falante.

Sendo a dublagem uma produção audiovisual em que os canais auditivo e visual são mesclados de forma assíncrona, esta pesquisa possibilita uma abordagem abrangente dos elementos prosódicos constitutivos da obra *Fragmentado (2016)*. O filme narra o sequestro de três meninas por um protagonista com transtorno dissociativo de identidade e estas diferentes identidades aparecem ao longo da obra desempenhando papéis e personalidades distintas. Das 24 identidades intrínsecas ao personagem principal, aqui exploramos as figuras Dennis, Fera, Hedwig e Patrícia, visto que estas aparecem de maneira mais significativa no longa.

Considerando o filme como obra audiovisual e a proposta de sua transmissão em grandes telas, percebe-se o empenho na caracterização das identidades não somente através da voz mas também na produção dos gestos e expressões faciais, modalidade visual, que ocorre simultaneamente à prosódia acústica. Assim, para o espectador, "o processamento da fala ocorre, por um canal audiovisual, em que as pistas auditivas e visuais apresentam grande funcionalidade" (Carnaval, 2021, p.36). Ademais, destaca-se que as modalidades áudio e vídeo são interpretadas por diferentes locutores para condução de uma mesma mensagem em contextos culturais e situacionais distintos.

Portanto, o objetivo geral deste trabalho é analisar como o dublador modula sua fala a fim de imprimir as singularidades e personalidade de cada voz em diálogo com a modulação da interpretação visual do ator na obra cinematográfica. Mais especificamente, a pesquisa objetiva:

(1) Caracterizar acusticamente, a partir do parâmetro de frequência fundamental (F0), cada uma das vozes performadas pelo dublador;
(2) Realizar a análise qualitativa dos contornos melódicos das personas[1] interpretadas.
(3) Coletar rótulos sobre a interpretação de ouvintes a partir de sua percepção auditiva das vozes, em teste *Free Labeling.*

[1] A terminologia "persona" é aqui adotada como sinônimo de identidade e se refere aos indivíduos dotados de nome e representação específica ainda que os mesmos sejam parte de um único personagem.

(4) Validar os rótulos obtidos no primeiro teste com base em um experimento de escolha forçada junto a um novo grupo de ouvintes;
(5) Verificar a relevância das modalidades auditiva (AU), visual (VI) e audiovisual (AV) na identificação de personalidades[2] interpretadas no processo de dublagem através de um experimento de escolha forçada multimodal;
(6) Analisar a consistência perceptiva entre a interpretação vocal do dublador e a informação visual do produto cinematográfico, averiguando a sincronia entre pistas prosódicas visuais realizadas pelo ator e acústicas realizadas pelo dublador.
(7) Analisar qualitativamente a prosódia visual das quatro personas selecionadas.

Nesse sentido, uma abordagem prosódica multimodal, que se propõe a descrever experimental e perceptivamente os recursos melódicos no processo de dublagem e a analisar experimental e qualitativamente a produção visual no processo de atuação, é aqui adotada.

[2] O termo "personalidade" será adotado para a descrição do conjunto de características referentes ao caráter, porte social e/ou comportamento das diferentes personas.

2. O filme

O filme *Fragmentado*, de 2016, é uma produção norte-americana escrita e dirigida por M. Night Shyamalan. Originalmente intitulado *Split*, o *thriller* integra uma sequência de três filmes, na qual ocupa a posição central.

A obra cinematográfica introduz o protagonista Kevin, que possui Transtorno Dissociativo de Identidade (TDI). Ele apresenta 23 personas distintas e o filme se desenvolve a partir da busca pela comprovação de uma vigésima quarta. Neste contexto, quatro personas são exploradas mais profundamente: Dennis, Fera, Hedwig e Patrícia.

Na individualização de cada uma das identidades, observa-se que Dennis é uma persona masculina soturna, obcecada com limpeza e organização. Hedwig, por sua vez, é uma personalidade infantilizada, com apenas 9 anos. Ademais, Fera é a vigésima quarta identidade encontrada e é uma personalidade animalesca e bestial. Por fim, Patrícia é a personalidade manipuladora e sorrateira da trama, a que arquiteta os planos que desenvolvem os acontecimentos do longa. Assim, percebe-se que Fera e Hedwig são personas com personalidades mais estereotipadas enquanto Dennis e Patrícia apresentam personalidades mais complexas.

Figura 1 - Personas do filme Fragmentado (2016): Patrícia, Fera, Hedwig e Dennis

Fonte: www. g1.globo.com.br

A relevância desta obra para a pesquisa desenvolvida é o interesse na caracterização prosódica – auditiva e visual – distinta de cada uma das personas. Tal aspecto torna-se significativo porque todas são interpretadas pelo mesmo ator, James McAvoy, na obra original e são dubladas pelo mesmo ator e dublador, McKeidy Lisita, em Português Brasileiro.

Figura 2 - Dublador McKeidy Lisita

Fonte: www.cadernopop.com.br

3. Pressupostos teóricos

A prosódia como estudo dos elementos suprassegmentais da fala se debruça sobre como algo é dito e suas implicações na produção e percepção do locutor e do interlocutor, respectivamente. Barbosa (2019) aponta que a abordagem fonética dos estudos prosódicos considera, tradicionalmente, três principais parâmetros prosódico-acústicos: a frequência fundamental (F0), a duração e a intensidade.

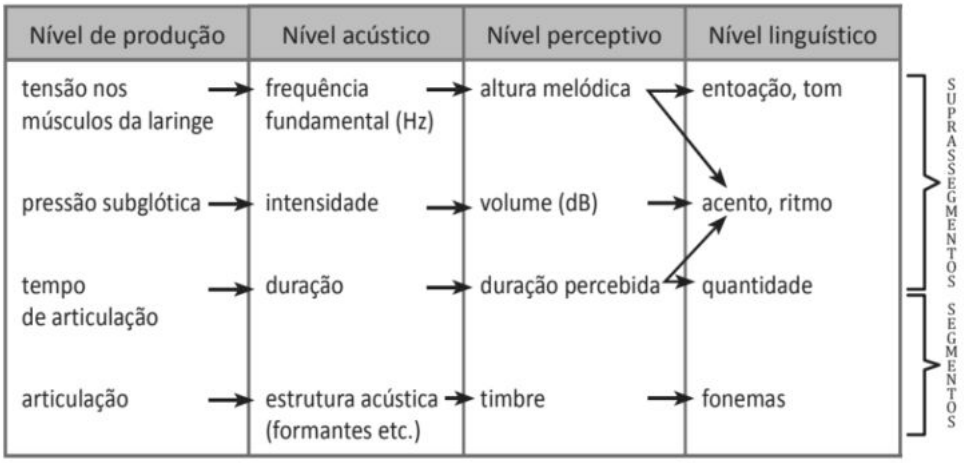
A frequência fundamental, no âmbito da produção, está relacionada ao número de vibrações das pregas vocais em determinado intervalo de tempo. Seu correlato perceptivo é o *pitch*, que se refere à percepção de um tom de voz mais agudo ou mais grave. A duração, por sua vez, está atrelada ao tempo relativo das unidades linguísticas que estruturam a informação suprassegmental dos enunciados (Barbosa, 2019), cujo correlato perceptivo é o ritmo. Por último, a intensidade diz respeito à amplitude do sinal acústico – perceptivamente, o volume da fala.

A estes parâmetros, pode-se adicionar a qualidade de voz. Esta concerne à vibração das pregas vocais e à configuração da laringe durante um enunciado. A partir das modificações dessa configuração, maior ou menor tensão, e à regularidade da vibração das pregas, podem-se demarcar diferentes estilos de fala, atitudes e emoções do falante (Moraes, 2024). Qualidades de voz delimitadas a partir de estudos da fisiologia do aparelho fonador (Laver, 1980) incluem a voz crepitante, a voz rouca e a voz soprosa, por exemplo.

Ao abordar a qualidade de voz, é possível realizá-lo a partir de uma perspectiva da produção – abordagem fisiológica que investiga, mapeia e delimita as diferentes disposições dos componentes do aparelho fonador e como estas se combinam para a produção de qualidades de voz distintas – ou da percepção. Para a presente pesquisa, adota-se uma abordagem perceptiva da qualidade de voz à luz da fonética experimental (Hayward, 2000).

De volta aos parâmetros prosódicos utilizados no estudo do nível suprassegmental da fala, menciona-se que a prosódia investiga mais intrinsecamente três fenômenos: o ritmo (constituído pela relação entre a duração silábica e a intensidade), o acento (proeminência de uma sílaba ou palavra em relação a um vocábulo ou enunciado) e a entoação (variações melódicas da fala; variações de F0). Abaixo, observa-se no quadro 1, de Moraes (2024), as relações entre os diferentes níveis fonéticos e o nível linguístico:

Quadro 1 – Correspondências entre os distintos níveis fonéticos (da produção, acústico e perceptivo) e o nível linguístico.



Fonte: Moraes (2024).

Neste estudo, o foco recairá sobre o fenômeno da entoação, o qual considera as modulações da altura melódica ao longo de um enunciado. Moraes (2024) menciona, ainda, que estas podem ser atravessadas pelo fenômeno da duração, inclusive de pausas, e do fenômeno da intensidade na percepção do volume do sinal acústico. Estes correlatos expressam, assim, uma multiplicidade de funções no contexto comunicativo que se organizam em um contínuo do nível mais linguístico para o nível mais extralinguístico.

De acordo com Fónagy (1993), são estabelecidas mais de 18 funções prosódicas que assumem papéis distintos como:

> "(i) demarcar unidades discursivas; (ii) segmentar a mensagem em partes; (iii) chamar a atenção do ouvinte para um ponto específico da mensagem; (iv) mostrar qual é o tipo gramatical do enunciado proferido; (v) tirar a ambiguidade de frases com dois (ou mais) significados; (vi) distinguir declarações declarativas de interrogativas; (vii) preparar a informação que a próxima frase vai trazer; (viii) expressar a atitude, emoção e/ou intenção do locutor; (ix) colaborar na identificação de quem está falando e (x) contribuir para o reconhecimento de diferentes gêneros discursivos." (Gomes da Silva e Carnaval, 2024, p.245).

Estas funções são colocadas resumidamente em Moraes (2024) como: função sintática (hierarquização e segmentação do enunciado), função informacional (organização e estruturação prosódica da transmissão do conteúdo lexical), função ilocutória (caracterização dos atos de fala) e funções paralinguísticas (das quais destacam-se a expressiva, relacionada às emoções e atitudes do falante, e a identificadora, atrelada à caracterização de uma

identidade pessoal, estilística e de diferentes gêneros discursivos). Nesta pesquisa, exploram-se as funções paralinguísticas expressiva e identificadora para a análise da produção e da percepção acústica das diferentes personas encenadas pelo dublador, visto que "a fala é caracterizada pela plasticidade e possibilidade infinita de gerar sonoridades com valor expressivo" (Madureira e Fontes, 2022, p.157).

Ao abordar a expressividade prosódica e a identificação de um locutor por um ouvinte, considera-se, aqui, o conceito de simbolismo sonoro, proposto por Ohala (1997), em que o interlocutor assume características físicas e psicológicas de um falante a partir de uma correlação direta com as propriedades de sua qualidade de voz. Tal conceito é embasado pela interpretação da entoação a partir dos códigos biológicos estabelecidos por Ohala (1994) e Gussenhoven (2004).

O código de frequência está relacionado às modulações de F0, que se reflete perceptivamente no correlato de *pitch*. Dessa maneira, valores mais altos de F0 (percepção de um som mais agudo) estariam atrelados a um posicionamento submisso enquanto valores mais baixos de F0 (percepção de um som mais grave) estariam relacionados a figuras dominantes. O código de produção, por outro lado, está atrelado à pressão de ar subglótica que é maior no início do enunciado e menor ao final. O código de esforço, por fim, diz respeito ao nível de tensão muscular da glote envolvido no processo de articulação sonora. Neste quesito, maiores valores de *pitch* refletem maior esforço vocal, enquanto menores valores de *pitch* estariam atrelados a um menor esforço vocal.

A proposta de análise da percepção e produção prosódica deste estudo, portanto, considera a perspectiva de Madureira e Fontes (2022; p.157), segundo a qual "minha qualidade de voz me identifica como locutor - produto de ajustes fonatórios, articulatórios, de tensão e do suporte respiratório".

Porém, a prosódia acústica não trabalha isoladamente no processo de caracterização e identificação de um falante. A prosódia visual atua representativa e intrinsecamente no emprego das funções prosódicas mencionadas:

> "Kendon (2004) distingue duas funções principais dos gestos coocorrentes à fala (*co-speech gestures*), a saber: gestos substanciais e gestos pragmáticos. Os primeiros contribuem para o conteúdo do enunciado, enquanto os segundos auxiliam na negociação de aspectos do enquadramento situacional" (Wagner, 2014).

Os gestos assumem protagonismo muitas vezes na segmentação do enunciado, na entoação de uma ênfase, na expressividade do discurso, na comunicação de emoção, atitude, atenção e engajamento (Ekman, 1979; Cafaro et al., 2012; Ishii & Nakano, 2008 *apud* Wagner, 2014). Logo, as expressões faciais, foco deste estudo, endossam a intencionalidade do discurso e reiteram a fala do locutor. Neste processo, gesto e prosódia se reforçam mutuamente e se coordenam temporalmente, sincronia, e estruturalmente, formas e intensidades similares (Wagner, Malisz e Kopp, 2014).

Assim, a prosódia audiovisual estabelece funções de maneira multimodal. Dessa forma, o presente trabalho aborda as diferentes personas interpretadas a partir da análise da produção e da percepção das diferentes modalidades abarcadas pela prosódia multimodal (auditiva, visual e audiovisual) intrínsecas a uma obra cinematográfica.

4. Metodologia

A abordagem metodológica deste trabalho é de natureza multidimensional devido ao seu objetivo de analisar a percepção prosódica das diferentes personas, por parte dos interlocutores, e a produção do ator e do dublador envolvidos na obra. Esses objetivos são entremeados, ainda, pela perspectiva multimodal da análise realizada. Inicialmente, a abordagem da pesquisa foi exclusivamente acústica, contudo, em seguida, esta foi desenvolvida a partir da consideração das modalidades visual e audiovisual. Assim, serão averiguadas sete hipóteses acerca da caracterização prosódica do produto audiovisual no português brasileiro:

(1) As diferentes personas analisadas apresentarão caracterização acústica distinta, a partir do parâmetro de F0.
(2) Padrões melódicos específicos são realizados para caracterizar cada uma das diferentes personas interpretadas pelo dublador.
(3) Haverá distinção perceptiva entre as quatro personas apresentadas.
(4) Os rótulos atribuídos às diferentes personas, em teste perceptivo, sinalizarão diferentes traços de personalidade, em diálogo com o conceito de simbolismo sonoro, e essas correspondências se confirmarão a partir da percepção de outros juízes.
(5) Haverá uma gradação na relevância das modalidades para a identificação das personas: audiovisual > visual > auditivo.
(6) As pistas prosódicas visuais serão um facilitador para o reconhecimento das personalidades menos caricatas.
(7) A análise qualitativa da prosódia visual dos enunciados selecionados confirmará a complexidade das personalidades com menor índice de identificação nos testes perceptivos.

Desse modo, os procedimentos desenvolvidos para a investigação dessas questões compreendem a preparação do *corpus*, a análise acústica da média de frequência fundamental, a descrição dos contornos melódicos, a análise qualitativa da produção visual, assim como os testes de percepção aplicados.

4.1 *Corpus*

O *corpus* desta pesquisa consiste em enunciados extraídos do filme *Fragmentado* (2016), disponível para livre acesso em plataformas digitais. O longa-metragem foi extraído integralmente do *Youtube* a partir da plataforma de inteligência artificial Turbo Scribe (2025), uma ferramenta online de transcrição automática que também permite o download de recursos audiovisuais em diferentes formatos e modalidades. Assim, o filme foi exportado em modalidade auditiva (AU) – .wav – e audiovisual (AV) – .mp4.

Em seguida, o áudio completo do filme foi segmentado em enunciados através do programa Audacity (2024), um software de gravação e edição de áudios, e exportado em quatro diferentes arquivos – um referente a cada uma das personas analisadas. Dessa forma, cada um dos arquivos foi composto por todas as falas de exclusivamente uma persona. Posteriormente, em cada arquivo foram eliminados os enunciados com muito ruído de fundo.

Por fim, cada um dos quatro arquivos de áudio foi analisado no Praat (Boersma e Weenink, 1992-2025), um software gratuito utilizado para análise e síntese de fala. Esses 4 arquivos de falas foram reservados para a análise das médias de F0. Contudo, para a descrição dos contornos melódicos, apenas um enunciado de cada persona foi selecionado, inicialmente, com base em sua qualidade acústica. Na análise multimodal final, mais cinco contornos foram analisados – dois da persona Dennis e três da persona Patrícia.

Para a composição dos testes de percepção em modalidade exclusivamente auditiva, foram selecionados e extraídos quatro enunciados de cada persona, totalizando os dezesseis estímulos apresentados no quadro 2 abaixo:

Quadro 2 - Enunciados utilizados nos testes de percepção de modalidade exclusivamente auditiva

Persona	Enunciados
Dennis	1. “Prometo que não vou incomodar vocês de novo” 2. “Não deviam enganar crianças” 3. “Vou trazer um copo d’água” 4. “Eu me lembro de tudo”
Fera	5. “O seu coração é puro” 6. “Eles estão adormecidos”

	7. “Obrigado…por nos ajudar até agora” 8. Os afligidos são os mais influídos”
Hedwig	9. “Acabei de comer um cachorro-quente” 10.“Meu nome é Hedwig” 11.“Eu tenho uma meia vermelha” 12.“Vocês não vão gostar”
Patrícia	13.“Eu converso com ele” 14.“Ele sabe disso” 15.“Coloquei um pouquinho de páprica” 16.“Kevin está dormindo agora”

Ademais, para a preparação do *corpus* multimodal, o arquivo do filme completo em formato AV foi inserido no editor de vídeos CapCut (2024), através do qual foram recortados novos 16 enunciados – 4 referentes a cada uma das personas. Desses enunciados, muitos coincidem com os utilizados nos primeiros dois testes. Contudo, 7 precisaram ser substituídos (3 da persona Dennis, 2 da persona Fera e 2 da persona Patrícia), como destacado em verde no quadro 3, devido ao enquadramento de cena ser um critério relevante para as modalidades VI e AV.

Quadro 3 - Enunciados utilizados no teste de percepção multimodal (em verde os distintos dos primeiros testes)

Persona	Enunciados
Dennis	1. “Que isso tranquilize você” 2. “Não deviam enganar crianças” 3. “Na verdade eu quero ser sincero com você” 4. “Ficam nos chamando de horda…os outros, sabe?”

Fera	5. “O seu coração é puro” 6. “Você é tão diferente das outras meninas” 7. “Alegre-se” 8. Os afligidos são os mais influídos”
Hedwig	9. “Acabei de comer um cachorro-quente” 10.“Meu nome é Hedwig” 11.“Eu tenho uma meia vermelha” 12.“Vocês não vão gostar”
Patrícia	13.“Eu converso com ele” 14.“Ele sabe disso” 15.“Coloquei um pouquinho de páprica” 16. “Obrigada, Dennis”

Enfim, os dezesseis enunciados selecionados foram exportados em 3 modalidades (AU, VI e AV), totalizando 48 arquivos para a composição do teste multimodal. Destes, apenas um enunciado de cada persona, em formato VI, foi utilizado para a análise qualitativa visual.

4.2 Procedimentos de análise acústica

No processo de análise acústica da pesquisa, foram realizados dois procedimentos, no *software* Praat: a aferição da média de F0 por persona e a descrição qualitativa de um contorno melódico de cada uma das personas. É importante ressaltar que, devido à natureza não-controlada do *corpus*, a abordagem acústica do mesmo encontrou muitas limitações em virtude de elementos como música de fundo, ruídos, além da qualidade inferior da versão do filme obtida *online*. Esses elementos são interpretados de maneira equivocada pelo *software* e interferem nas medidas e contornos gerados pelo programa.

Inicialmente, os quatro arquivos de enunciados, – 1 referente a cada persona – em formato de áudio, foram submetidos à segmentação no Praat. Em seguida, realizou-se a aferição de medidas de F0 dos trechos em cada uma das vogais em posição tônica, devido à maior estabilidade acústica neste contexto. Cada uma das medidas foi lançada em uma planilha no Excel e assim foram calculadas as médias de F0 por persona.

No que tange à análise qualitativa dos contornos melódicos, foram analisados 10 enunciados, no total (Quadro 4). Preliminarmente, foram selecionados quatro enunciados, um de cada persona, com a menor incidência de ruídos. Posteriormente, a partir da análise visual, foram selecionados mais seis enunciados para a realização da descrição qualitativa dos contornos, a fim de complementar a abordagem multimodal da produção. Destes, dois são referentes à persona Dennis, 1 à persona Fera e três à persona Patrícia:

Quadro 4 - Enunciados dos contornos melódicos descritos qualitativamente.

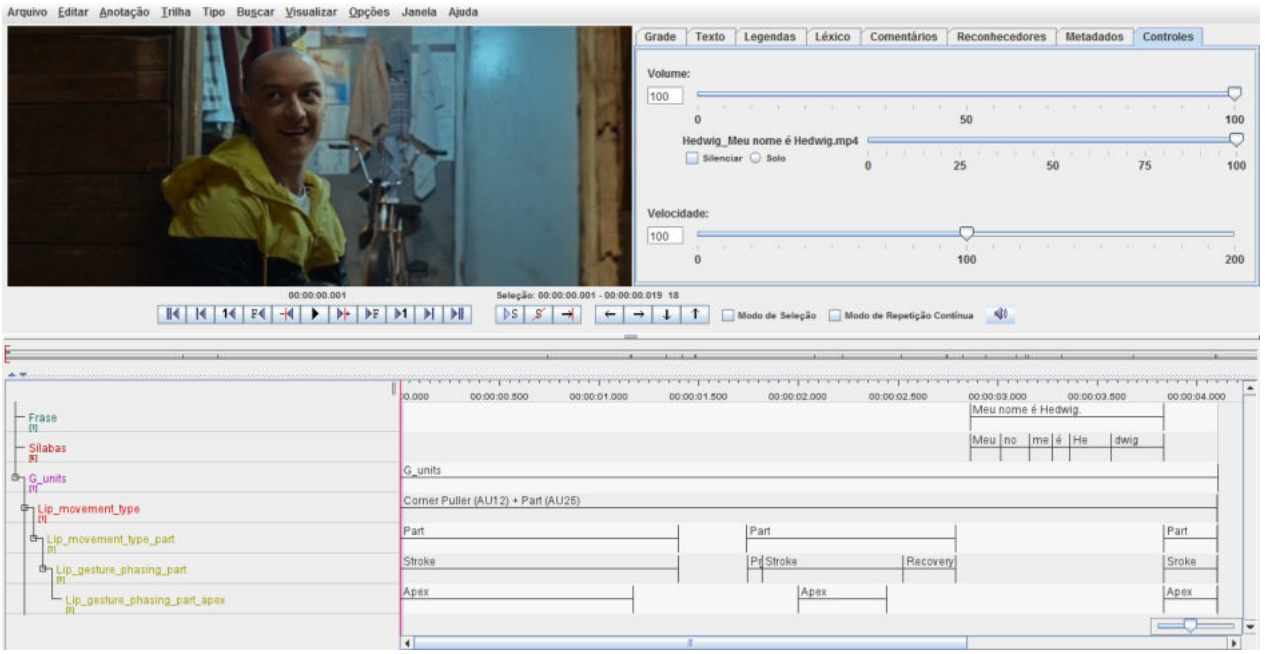
Persona	**Enunciados**
Dennis	1. “Não deviam enganar crianças” 2. “Ficam nos chamando de horda” 3. “os outros, sabe?”
Fera	4. “O seu coração é puro” 5. “Alegre-se”
Hedwig	6. “Meu nome é Hedwig”
Patrícia	7.“Kevin está dormindo” 8. “Ele não pode tocar em vocês” 9.“Ele sabe disso” 10.“Hum hum”

Esse procedimento metodológico se justifica pelo critério de enquadramento de cena, necessário à descrição visual, que motivou a alteração dos enunciados utilizados para a análise das referidas personas.

4.3 Procedimentos de análise visual

Para a análise visual, foi utilizado o *software EUDICO Linguistic Annotator* (ELAN), uma ferramenta que possibilita anotar, documentar e analisar a comunicação, incluindo linguagem de sinais e gestos. Desenvolvido pelo *Max Planck Institute for Psycholinguistics*, esse programa permite a anotação de maneira sistemática (em estruturas hierárquicas de trilhas) simultaneamente ao acompanhamento do arquivo multimídia. Assim, o *software* viabiliza o alinhamento entre notações e canal auditivo, visual ou audiovisual.

Figura 3 - Interface do *software EUDICO Linguistic Annotator* (ELAN)



Fonte: Captura de tela da autora

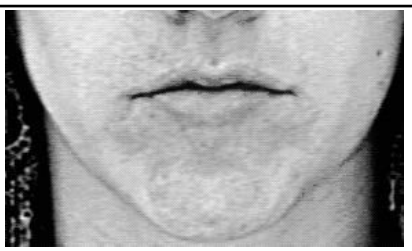
Ademais, para o processo de descrição e sistematização da prosódia visual dos trechos selecionados, foram utilizados o manual FACS - *Facial Action Coding System* (Ekman, Friesen & Hager, 2002) e o manual de rotulagem M3D - *MultiModal MultiDimensional* (Rohrer et alii, 2025).

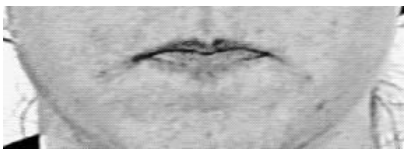
O sistema FACS consiste em um manual que mapeia as diferentes configurações da face a partir das atividades musculares do rosto. Ele, portanto, não se dispõe a discutir emoções ou reações, mas se atém a medir e descrever minuciosamente as Unidades de Ação

(AUs) referentes a músculos específicos – quadro 5. As AUs são categorizadas pelas regiões do rosto em que ocorrem, face superior ou face inferior, e pela escala de evidência do movimento em relação aos seus níveis de intensidade, variando de A (quase imperceptível) a E (máxima evidência). Além disso, o manual aborda as AUs que não podem ocorrer simultâneamente e as que ocorrem como consequência de outras. Dessa forma, o sistema é aqui utilizado para descrição dos gestos e expressões faciais.

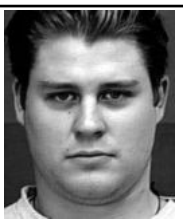
Quadro 5 – Descrição traduzida dos movimentos de face (Ekman & Friesen, 1978).

AU	Descrição	Músculo Facial	Exemplo em imagem
1	Levantamento da parte interna da sobrancelha	*Frontalis, pars medialis*	
2	Levantamento da parte externa da sobrancelha	*Frontalis, pars lateralis*	
4	Franzimento da sobrancelha	*Corrugator supercilii, Depressor supercilii*	
5	Levantamento da pálpebra superior	*Levator palpebrae superioris*	
6	Bochechas para cima	*Orbicularis oculi, pars orbitalis*	
7	Pálpebras apertadas	*Orbicularis oculi, pars palpebralis*	
9	Nariz enrugado	*Levator labii superioris alaquae nasi*	
10	Elevação do lábio superior	*Levator labii superioris*	

11	Nariz e lábios aumentados	*Zygomaticus minor*	
12	Levantamento dos lábios	*Zygomaticus major*	
13	Covinhas na bochecha	*Levator anguli oris (a.k.a. Caninus)*	
14	Estiramento do canto dos lábios	*Buccinator*	
15	Canto dos lábios abaixados	*Depressor anguli oris (a.k.a. Triangularis)*	
16	Abaixamento dos lábios	*Depressor labii inferioris*	
17	Franzimento do queixo	*Mentalis*	
18	Fazendo “biquinho”	*Incisivii labii superioris and Incisivii labii inferioris*	

20	Lábios esticados	*Risorius w/ platysma*	
22	Lábios afunilados	*Orbicularis oris*	
23	Estreitamento dos lábios	*Orbicularis oris*	
24	Pressionamento dos lábios	*Orbicularis oris*	
25	Afastamento dos lábios	*Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris*	
26	Caimento da mandíbula	*Masseter, relaxed Temporalis and internal Pterygoid*	
27	Abertura estendida da boca	*Pterygoids, Digastric*	
28	Contração dos lábios	*Orbicularis oris*	

41	Caimento das pálpebras	*Relaxation of Levator palpebrae superioris*	
42	Olhos quase fechados	*Orbicularis oculi*	
43	Olhos fechados	*Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*	
44	Olhadinha	*Orbicularis oculi, pars palpebralis*	
45	Pestanejar	*Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*	–
46	Piscadinha	*Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*	–
51	Movimento de cabeça para a esquerda		
52	Movimento de cabeça para a direita		
53	Movimento de cabeça para cima		

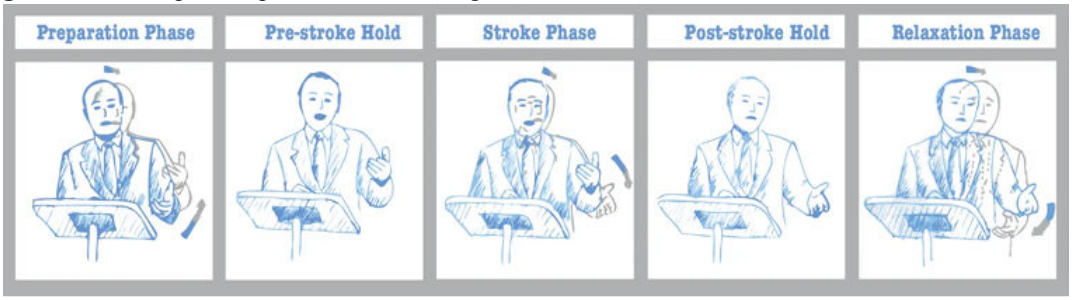
54	Movimento de cabeça para baixo		
55	Inclinação para a esquerda		
56	Inclinação para a direita		
57	Movimento de cabeça para a frente		
58	Movimento de cabeça para trás		
61	Olhar para a esquerda		
62	Olhar para a direita		
63	Olhar para cima		
64	Olhar para baixo		

Fonte: Extraído de Figueiredo (2018, pp. 70-75)

O manual M3D, por sua vez, aborda as AUs como unidades gestuais (*G-units*). Apesar de apresentar enfoque nos gestos de mãos, irrelevantes à nossa descrição, o manual oferece instrução e treinamento para o uso do ELAN no processo de notação para descrição prosódica visual, com uma metodologia de aplicação bastante flexível e adaptável. Outrossim, o M3D é organizado de maneira tridimensional e se estrutura a partir das dimensões da forma, da prosódia e do significado. No tangente à dimensão prosódica, há a segmentação das fases gestuais de uma *G-unit*, o que reflete a organização estrutural hierárquica dos constituintes do movimento, assim como a sua relação com a prosódia da fala. O manual especifica as seguintes fases do movimento: preparação, hesitação pré-*stroke*, *stroke* (o movimento intencionado), hesitação pós-*stroke* e relaxamento, como apresentado na figura 5. Destaca-se que nem todas as *G-units* serão compostas por, necessariamente, todas as fases.

Dessa forma, as fases do gesto podem ser analisadas no software ELAN a partir da estrutura hierarquizada das trilhas. Partindo desse princípio, é estabelecida uma trilha-mãe para o tempo de duração da unidade gestual por completo e uma trilha-filha para a segmentação das fases constituintes da ação. Uma vez segmentada a ação, estabelece-se uma trilha-filha (em relação à trilha de segmentação gestual) que conterá o *apex* dos *strokes* delimitados e descritos. O *apex* é o "ápice do movimento", ou seja, o trecho de maior intensidade da realização da unidade gestual intencionada.

Figura 4 – Fases que compõem uma unidade gestual.



Fonte: Retirado do Manual M3D. segmentação do movimento

Por fim, para a descrição da prosódia visual deste trabalho, foram selecionados 4 vídeos, um por persona, dos estímulos utilizados no teste multimodal, que será apresentado na próxima seção, ou presentes no mesmo enquadramento de cena. Os critérios de seleção de tais enunciados consideraram os vídeos com maior extensão e maior densidade de unidades de ação (AUs).

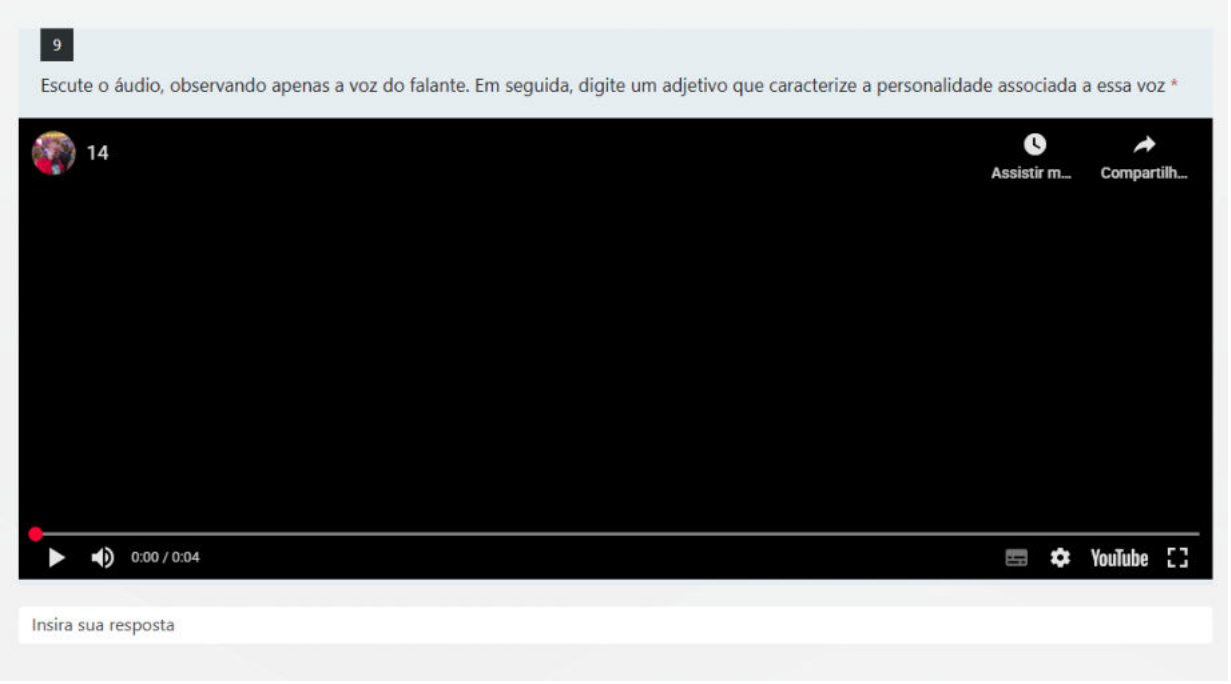
4.4 Testes perceptivos

As três atividades perceptivas desenvolvidas ao longo desta pesquisa adotam uma abordagem metodológica sequencial e integrada. Nessa perspectiva, o primeiro teste gerou os dados que estruturaram o segundo e o terceiro teste foi proposto por justificativas oriundas dos resultados do teste anterior.

O primeiro experimento foi realizado em uma modelagem *Free labeling*. Nesta tarefa, os juízes deveriam ouvir os estímulos, exclusivamente auditivos, e rotular a personalidade da voz escutada com um adjetivo que melhor a caracterizasse. A motivação para a realização desta atividade consistiu na busca pela rotulação das personalidades das identidades estudadas sem o enviesamento da perspectiva dos envolvidos com o estudo.

Para a composição do teste, foram selecionados 16 estímulos (vide quadro 2), quatro por persona, sob os critérios de (1) extensão lexical semelhante, (2) menos ruído de fundo e (3) conteúdo lexical com maior neutralidade possível. O mesmo foi elaborado na plataforma *Microsoft Forms,* que permite a aleatorização da ordem dos estímulos, e cada um desses foi introduzido pelo mesmo comando, seguido por uma aba de resposta livre (Figura 5). Observa-se que foi realizado o upload dos áudios para o *Youtube* de modo que estes pudessem ser inseridos no teste.

Figura 5 – *Layout* do teste de *free labeling.*

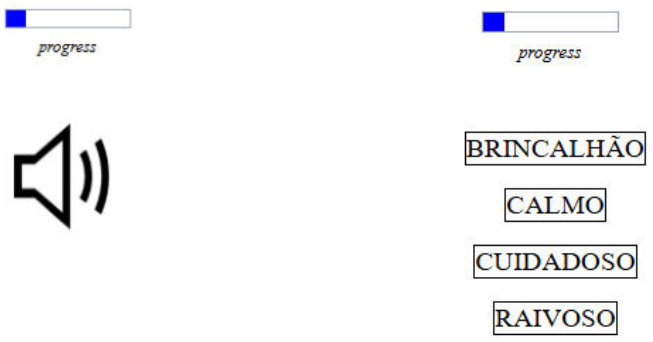


Fonte: Captura de tela da plataforma *Microsoft Forms*

Participaram desta atividade 24 juízes, de maneira assíncrona, selecionados sob o critério geral de ter nível superior completo ou em curso. Além disso, eles foram abordados a partir da sua distribuição homogênea em 2 grupos: 12 compuseram o grupo controle (pessoas que viram o filme previamente) e 12 integraram o grupo experimental (pessoas que nunca haviam visto o filme). A partir dos adjetivos propostos pelos participantes, foi realizada a seleção do rótulo de maior incidência, para cada persona, na interseção entre as respostas obtidas dos dois grupos. Assim, foram selecionados 4 rótulos para a elaboração do teste seguinte.

Estabelecidos os rótulos para a caracterização das personas, foi elaborado um teste de escolha condicionada a partir do *software* PCIbex - *PennController for Internet Based Experiments* (Zehr; Schwarz, 2018) – plataforma que permite a elaboração e o compartilhamento de experimentos online. Nesta tarefa, participantes distintos dos que realizaram o teste anterior deveriam ouvir os estímulos e selecionar, entre quatro alternativas, qual rótulo melhor caracterizava a personalidade da voz escutada (Figuras 6 e 7 abaixo).

Figuras 6 e 7 – *Layout* do teste de escolha condicionada.



Fonte: Capturas de tela do software PCIbex.

Para a estruturação deste segundo teste, foram utilizados os mesmos 16 estímulos do experimento anterior. Em seguida, foram selecionados novos 24 juízes (12 para o grupo experimental e 12 para o grupo controle) para realização da tarefa perceptiva. Este teste também foi aplicado assincronamente. Contudo, para esse havia a obrigatoriedade da sua execução em um computador.

Os resultados do segundo experimento, de escolha condicionada, foram submetidos à análise estatística realizada no software R (R Core Team, 2021), através de um Modelo de Regressão Logística de Efeitos Mistos, considerando o índice de correspondência correta

entre persona e rótulo de maneira geral e, em seguida, por grupo. Este modelo estatístico combina efeitos fixos e aleatórios no processamento dos dados. Neste estudo, considera-se as variáveis independentes (preditoras): personas e grupos; e as variáveis de efeitos aleatórios: participantes e conteúdo léxico dos enunciados.

Além disso, os resultados da média de tempo de reação dos participantes para a realização da tarefa perceptiva foram processados através de um Teste T. Essa é uma ferramenta estatística usada para determinar se há diferença significativa entre as médias de dois grupos de dados. Dessa forma, os resultados das médias de tempo obtidas por persona foram submetidos a uma análise combinatória dois a dois. Contudo, duas identidades não foram significativamente associadas a seus devidos rótulos. Com isso, foi desenvolvido o terceiro e último experimento, de caráter multimodal.

O teste perceptivo multimodal consistiu de uma tarefa de escolha condicionada, em que, igualmente ao segundo experimento, os juízes precisaram atribuir o rótulo que melhor descrevia a personalidade da identidade apresentada, porém, agora, em 3 modalidades distintas.

Para essa tarefa foram utilizados 16 enunciados apresentados nas modalidades AU, VI e AV, o que totalizou 48 estímulos. Dos enunciados empregados nos experimentos anteriores, alguns precisaram ser revisados e substituídos devido ao novo critério estabelecido a partir da modalidade visual: o enquadramento da câmera nas cenas para que as expressões faciais das personas estivessem evidentes nos vídeos (em quadro 3).

Similarmente ao primeiro experimento, a tarefa perceptiva foi elaborada na plataforma *Microsoft Forms*. Contudo, foi estruturado um formulário para cada modalidade, visto que a randomização dos estímulos não pode ocorrer quando há mais de uma seção no *forms*. Assim, foi impossível a aleatorização dos estímulos exclusivamente em cada modalidade. Desse modo, para que a aleatorização dos estímulos fosse realizada, os 3 formulários foram enviados separadamente a cada um dos participantes, que respondiam os mesmos em sequência.

Participaram desta fase experimental novos 24 juízes, dos quais nenhum havia assistido à obra cinematográfica. Os sujeitos, divididos em dois grupos, foram apresentados aos estímulos nas seguintes ordens: primeiro grupo – (1) áudio isolado, (2) vídeo isolado e (3) áudio e vídeo combinados; segundo grupo – (1) vídeo isolado, (2) áudio isolado e (3) áudio e vídeo combinados. Essa abordagem aos participantes consistiu apenas em uma opção metodológica para evitar o enviesamento das respostas; este não constituiu um fator de análise dos dados.

Os resultados obtidos nesta etapa perceptiva também foram submetidos à análise estatística através do modelo de regressão logística de efeitos mistos na plataforma R. Nesta ocasião, foram consideradas (1) a porcentagem de reconhecimento do rótulo atribuído a cada persona, independente da modalidade (2) a porcentagem de correta atribuição de rótulos por modalidade e (3) a análise da porcentagem do reconhecimento de cada persona por modalidade.

5. Análise de dados e discussão de resultados

Nesta seção, será apresentada a análise dos dados da pesquisa e serão discutidos os resultados ao estabelecer a relação entre os dados obtidos através de diferentes etapas metodológicas. Inicialmente, será detalhada a abordagem perceptiva do estudo, obtida a partir dos resultados dos testes de percepção – teste *free labeling*, teste de escolha condicionada e teste multimodal. Em seguida, será descrita a consistência entre a produção acústica do dublador, a partir das médias de frequência fundamental de cada persona e das considerações tecidas sobre a análise de seus contornos melódicos, e a produção visual do ator, a partir da análise qualitativa da prosódia visual dos enunciados selecionados.

5. 1 Análise dos testes de percepção

5.1.1 Teste Free labeling

Na abordagem da percepção prosódica do produto audiovisual, o primeiro passo adotado foi a realização do teste *free labeling*, reunindo, assim, as categorizações – propostas por 24 juízes, 12 do grupo controle e 12 do grupo experimental – das vozes das personas a partir de 16 enunciados, consistindo em 4 estímulos de cada persona, todos apresentados na modalidade exclusivamente auditiva.

Nesta etapa, contabilizamos os rótulos propostos a fim de selecionar aqueles apresentados com maior incidência por ambos os grupos de participantes, como pode ser observado nas nuvens de palavras a seguir. Observa-se que cada nuvem é constituída de um grupo de palavras em cor azul, composto pelas respostas do grupo controle, e um conjunto de palavras em cor amarela, composto pelas respostas do grupo experimental. Em verde, estão as respostas que foram observadas mutuamente em ambos e em vermelho, encontra-se o rótulo com maior ocorrência entre os dois núcleos de juízes.

Em primeiro plano, a persona “Dennis” apresentou a segunda maior gama de diferentes rótulos propostos (Figura 8). Contudo, ela foi a persona com o maior número de adjetivos sugeridos comumente por ambos os grupos controle – representado em azul – e experimental – representado em amarelo. Ademais, foi verificado que os rótulos "calmo" e "sério" apresentaram o mesmo número de ocorrências na intercessão entre os dois grupos. Enquanto o rótulo “calmo” aparece igualmente distribuído entre os dois grupos, o rótulo “sério” aparece com maior incidência no grupo experimental – 6 vezes. Por isso, o rótulo “calmo” foi o escolhido para caracterizar a persona Dennis.

Figura 8 - Rótulos obtidos, a partir do teste *free labeling*, para a persona Dennis.

Fonte: Nuvem gerada no *site WordArt.*

Em seguida, a persona "Fera" apresentou o menor espectro de rótulos propostos durante o teste (Figura 9). Além de "Raivoso" ter sido mais incidentemente apresentado pelos juízes para a caracterização da personalidade desta voz, ele também foi o rótulo selecionado com maior número de ocorrências (14), quando comparado a todas as outras personas. Demonstra-se, assim, uma aparente maior facilidade dos participantes em descrever esta voz, o que dialoga com o emprego, por parte do dublador, da estratégia melódica para a particularização da voz bestial.

Figura 9 - Rótulos obtidos, a partir do teste *free labeling*, para a persona Fera.

Fonte: Nuvem gerada no *site WordArt.*

Ao analisar o resultado dos rótulos sugeridos para a caracterização da personalidade da persona "Hedwig", percebe-se a maior ocorrência do adjetivo "Brincalhão" (Figura 10).

Destaca-se, também, que os rótulos “Infantil” e “Infantilizado” aparecem em quantidades consideráveis e que, se somados, ultrapassam a recorrência do rótulo “brincalhão”. Contudo, metodologicamente optou-se por contabilizá-los separadamente, visto que considera-se distintas uma personalidade intrinsecamente infantil e uma personalidade infantilizada, pois esta última demonstra intencionalidade e desconexão com uma idade biológica.

Figura 10 - Rótulos obtidos, a partir do teste *free labeling*, para a persona Hedwig.

Fonte: Nuvem gerada no *site WordArt.*

Por fim, a persona “Patrícia” apresentou a maior quantidade de rótulos distintos propostos pelos juízes (Figura 11). Observa-se que o adjetivo de maior incidência no teste foi “Cuidadosa”. Contudo, notamos que “Calmo”, curiosamente, foi o segundo rótulo mais recorrente na interseção entre os grupos controle e experimental, tendo apenas uma ocorrência a menos que “Cuidadosa”.

Figura 11 - Rótulos obtidos, a partir do teste *free labeling*, para a persona Patrícia.

Fonte: Nuvem gerada no *site WordArt.*

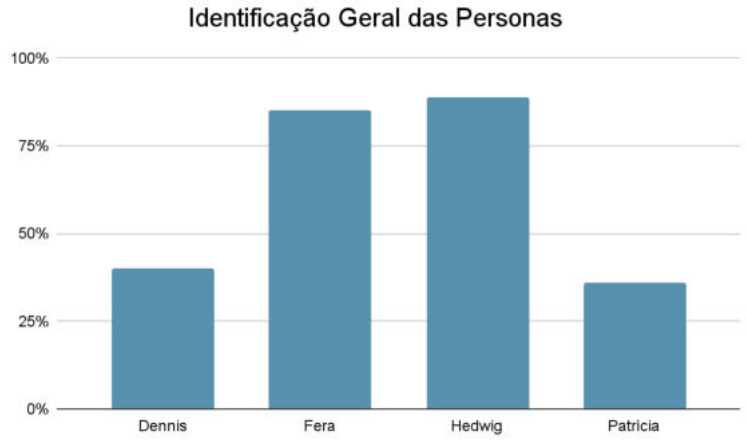
Ao longo deste primeiro teste, rótulos como "mocinho" – proposto por um juiz ao estímulo "Não deveriam enganar crianças" – ou "prendado" – proposto ao enunciado "Coloquei um pouquinho de páprica" – foram exemplos, dentre outras ocorrências, da provável influência do conteúdo lexical dos enunciados na atribuição de rótulos distintos para a mesma persona. Percebemos que, ainda que os rótulos selecionados dialoguem bem com as personalidades propostas dentro da própria obra audiovisual, alguns dos adjetivos atribuídos se relacionavam com o conteúdo lexical dos estímulos propostos.

5.1.2 Teste de escolha condicionada

Em etapa perceptiva posterior, foi realizado um teste de escolha condicionada, no *software* PCIbex, a partir dos rótulos de maior incidência para cada persona no teste anterior – "calmo" para a persona Dennis, "raivoso" para a persona Fera, "brincalhão" para a persona Hedwig e "cuidadosa" para a persona Patrícia. É importante ressaltar que o rótulo para a persona Patrícia foi inserido no novo teste como "cuidadoso" a fim de evitar o enviesamento dos participantes a partir da marcação morfológica de gênero.

Os mesmos 16 estímulos do teste *free labeling* foram submetidos a novos 24 juízes novamente distribuídos em 2 grupos de análise – grupo experimental (juízes que não haviam visto o filme) e grupo controle (juízes que viram o filme). A seguir, descreve-se as análises estatísticas dos dados de identificação, geral (Gráfico 1; Figura 12) e por grupo (Gráfico 2; Figura 13) realizadas através do modelo estatístico e a análise das médias dos tempos de resposta, por persona (Quadro 6), comparadas através de um Teste T. Todas as análises foram processadas no *software* R.

Gráfico 1 – Índice de identificação geral das personas no teste de escolha condicionada, a partir do modelo estatístico de regressão logística de efeitos mistos.



Fonte: Elaborado pela autora.

Figura 12 – Análise estatística da identificação geral das personas no teste de escolha condicionada, a partir do modelo de regressão logística de efeitos mistos.

	Correct		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.64	0.29 – 1.41	0.269
Persona [Fera]	10.11	3.01 – 33.98	**<0.001**
Persona [Hedwig]	14.50	4.11 – 51.14	**<0.001**
Persona [Patricia]	0.83	0.28 – 2.52	0.746

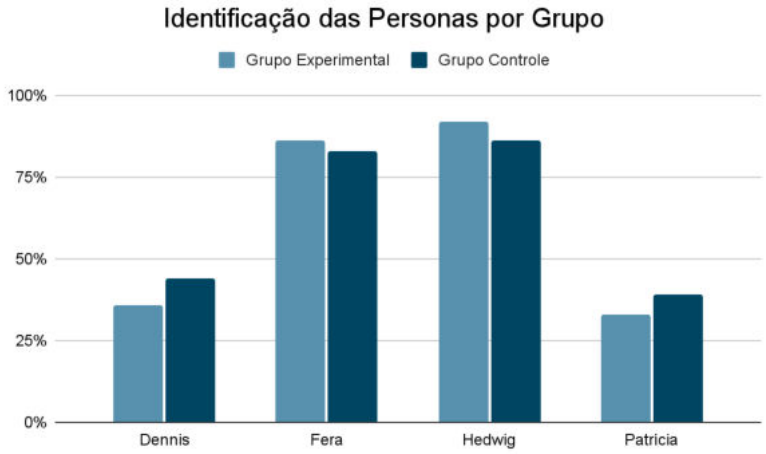
Fonte: *Software* R

Descritivamente (Gráfico 1) percebe-se que os índices de acerto na correspondência entre rótulo e persona ultrapassaram 75% na identificação das personas Fera e Hedwig. Contudo, isso não se aplica às personas Dennis e Patrícia, cujas porcentagens de identificação com o adjetivo esperado ficaram abaixo de 50%.

A partir da estatística inferencial, o modelo revelou que o percentual geral de identificação das personas Fera (p-valor <0.00) e Hedwig (p-valor <0.001) foi significativo (Figura 12). Em contrapartida, os índices de identificação de Dennis e Patrícia não tiveram relevância estatística, devido à grande influência do acaso.

Assim, foi realizada a análise do percentual de identificação das personas por grupo, como demonstrado no gráfico 2 a seguir.

Gráfico 2 – Índice de identificação das personas por grupo no teste de escolha condicionada, a partir do modelo estatístico de regressão logística de efeitos mistos.



Fonte: Elaborado pela autora.

Figura 13 – Análise estatística da identificação geral das personas no teste de escolha condicionada, a partir do modelo de regressão logística de efeitos mistos.

		Correct	
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.52	0.20 – 1.37	0.188
Persona [Fera]	13.89	3.09 – 62.41	**0.001**
Persona [Hedwig]	24.59	4.71 – 128.53	**<0.001**
Persona [Patricia]	0.89	0.23 – 3.36	0.862
Fragmentado [SIM]	1.48	0.52 – 4.16	0.460
Persona [Fera] × Fragmentado [SIM]	0.54	0.10 – 2.80	0.463
Persona [Hedwig] × Fragmentado [SIM]	0.37	0.06 – 2.30	0.289
Persona [Patricia] × Fragmentado [SIM]	0.88	0.21 – 3.63	0.859

Fonte: *Software* R

A partir do gráfico dos percentuais de acerto na identificação das personas, destaca-se a permanência do padrão de identificação sem grande distinção entre as percentagens de acerto dos grupos experimental e controle.

Inferencialmente, o modelo de regressão logística de efeitos mistos revelou que o percentual de identificação por grupos das personas Fera (p-valor =0.001) e Hedwig (p-valor

<0.001) foi significativo, enquanto para as personas Dennis e Patrícia não houve relevância estatística em ambos os grupos. Assim, a análise indicou que a variável grupo não apresentou relevância estatística no índice de acerto na correspondência entre rótulo e persona.

Por fim, intencionou-se analisar como a média dos tempos de resposta para a identificação das personas se relaciona com os baixos índices de acerto na rotulação esperada para as personas Dennis e Patrícia em oposição aos altos percentuais de identificação das personalidades das personas Fera e Hedwig (Quadro 6):

Quadro 6 - Média dos tempos de resposta, por persona, no teste de escolha condicionada

Personas	Dennis	Fera	Hedwig	Patrícia
Médias dos Tempos de Resposta	3574 ms	3237 ms	3242 ms	4163 ms

Fonte: Elaborado pela autora.

Em consonância com a dificuldade de rotulação das personas Dennis e Patrícia, por parte dos juízes, as médias do tempo de resposta para essas identidades foram as mais altas, independentemente do rótulo atribuído estar de acordo com o esperado ou não: Dennis (3574 ms) e Patrícia (4163 ms). Percebe-se ainda que a persona Patrícia foi a que apresentou a maior média de tempo de resposta, no geral. Isso se reflete nos testes T aplicados entre todas as possibilidades combinatórias, em que este se revelou significativo apenas para diferença dos tempos de resposta entre Fera ~ Patrícia (p-valor = 0.011) e Hedwig ~ Patrícia (p-valor = 0.015).

5.1.3 Teste Multimodal

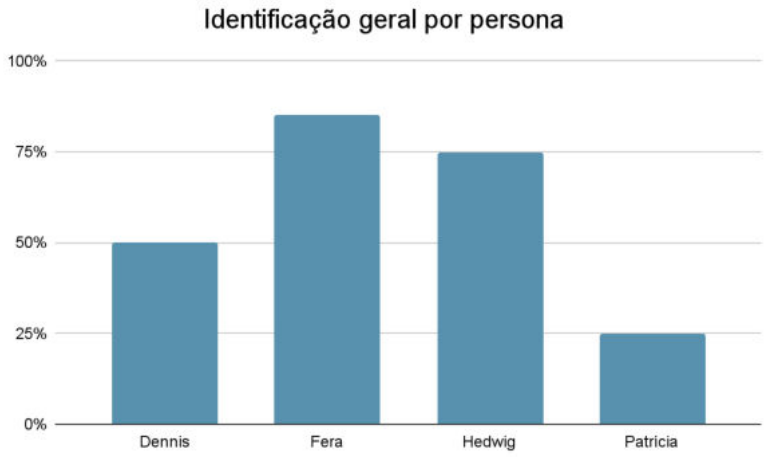
A partir dos testes perceptivos anteriores, foi possível perceber uma fragmentação entre as personas aqui analisadas. Ao passo que Fera e Hedwig foram mais facilmente categorizadas no teste *free labeling* e tiveram seus rótulos validados no teste de escolha condicionada, as personas Dennis e Patrícia apresentaram um número significativamente maior de rótulos propostos no primeiro teste e não tiveram os adjetivos utilizados para a sua caracterização validados no teste seguinte, apresentando baixo índice de identificação.

Esse panorama motivou a realização de uma terceira tarefa perceptiva de abordagem multimodal sob a hipótese de que as pistas prosódicas visuais seriam um facilitador para o reconhecimento das personalidades com menor índice de identificação.

O experimento foi composto por 16 enunciados, 4 de cada persona, apresentados em 3 modalidades diferentes, totalizando 48 estímulos. Esses estímulos foram dispostos em 3 formulários distintos, 1 por modalidade (AU, VI e AV) e foram submetidos à avaliação de novos 24 juízes em duas ordens de apresentação: AU - VI - AV e VI - AU -AV.

Os resultados deste teste também foram analisados a partir do modelo de regressão logística de efeitos mistos no *software* R (Gráfico 3; Figura 14).

Gráfico 3 – Índice de identificação geral das personas no teste multimodal a partir do modelo estatístico de regressão logística de efeitos mistos.



Fonte: Elaborado pela autora.

Figura 14 – Análise estatística da identificação geral das personas no teste multimodal, a partir do modelo de regressão logística de efeitos mistos.

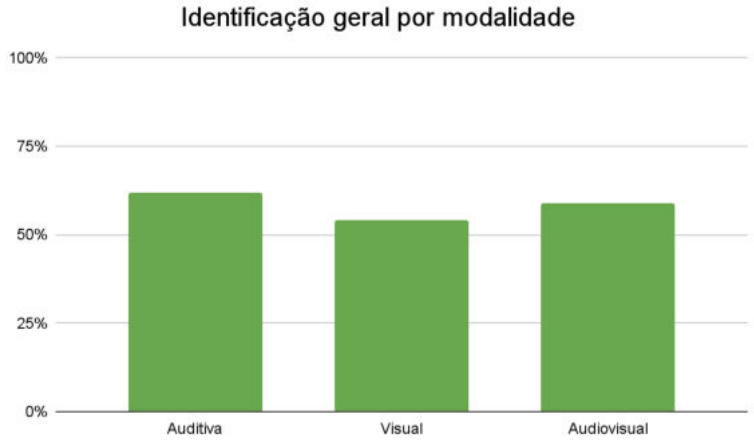
		Correct	
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.99	0.62 – 1.57	0.954
Persona [Fera]	6.49	3.35 – 12.59	**<0.001**
Persona [Hedwig]	3.24	1.71 – 6.12	**<0.001**
Persona [Patrícia]	0.31	0.16 – 0.59	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Nome}$	0.17		
$\tau_{00\ Item}$	0.14		
ICC	0.09		
N_{Nome}	24		
N_{Item}	16		
Observations	1152		
Marginal R^2 / Conditional R^2	0.272 / 0.335		

Fonte: *Software* R

Inicialmente, foi analisado o índice de identificação geral por persona (Gráfico 3). No tangente à identidade Dennis, houve um percentual de 50% de acerto na atribuição do rótulo esperado para sua caracterização. Para a persona Fera, este percentual foi de 85%. Para a persona Hedwig, 74%. Por fim, para a persona Patrícia o percentual de acerto foi de 25%. Novamente, as personas Dennis e Patrícia apresentaram baixos índices de identificação a partir dos rótulos propostos. A partir destes percentuais, o modelo de regressão logística de efeitos mistos revelou que a diferença na identificação entre as personas, independente da modalidade, foi significativa.

Em seguida, foi realizada a estatística descritiva e inferencial do índice de identificação por modalidade, independentemente das personas (Gráfico 4; Figura 15).

Gráfico 4 – Índice de identificação geral por modalidade, no teste multimodal, a partir do modelo estatístico de regressão logística de efeitos mistos.



Fonte: Elaborado pela autora.

Figura 15 – Análise estatística da identificação geral por modalidade no teste multimodal, a partir do modelo de regressão logística de efeitos mistos.

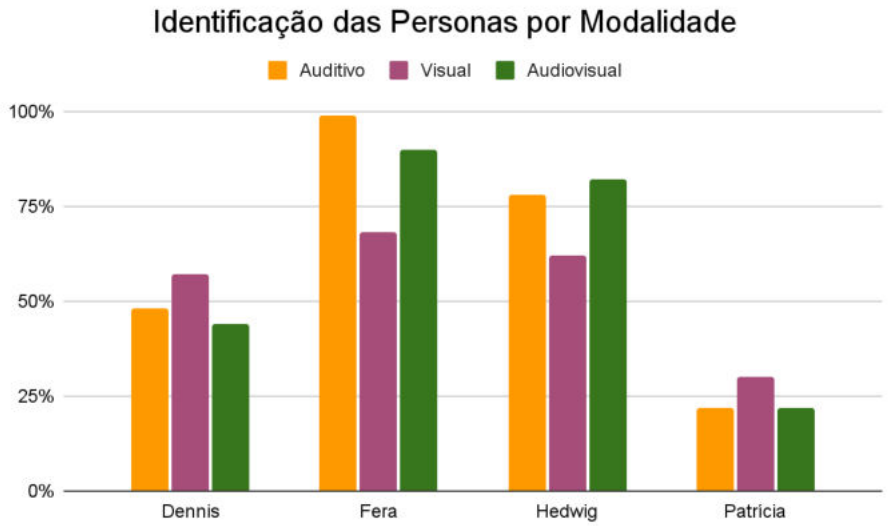
	Correct		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.91	0.97 – 3.75	0.060
Modalidade [AV]	0.87	0.62 – 1.22	0.434
Modalidade [V]	0.66	0.47 – 0.93	**0.016**
Random Effects			
σ^2	3.29		
$\tau_{00\ Nome}$	0.18		
$\tau_{00\ Item}$	1.53		
ICC	0.34		
$N_{\ Nome}$	24		
$N_{\ Item}$	16		
Observations	1152		
Marginal R^2 / Conditional R^2	0.006 / 0.346		

Fonte: *Software* R

Percebe-se, qualitativamente, que os índices de acerto por modalidade foram todos acima de 50%, porém abaixo de 70% – auditivo (62%), visual (54%) e audiovisual (59%). A análise dos p-valores de cada modalidade indicou que apenas a diferença na identificação na modalidade visual foi significativa em relação à modalidade auditiva.

Por fim, foi conduzida a análise dos dados do experimento multimodal a partir da confluência das variáveis preditoras persona e modalidade (Gráfico 5).

Gráfico 5 – Índice de identificação das personas por modalidade, no teste multimodal, a partir do modelo estatístico de regressão logística de efeitos mistos.



Fonte: Elaborado pela autora.

Descritivamente, percebe-se a predominância da modalidade visual na identificação da persona Dennis (57%), seguida pelo percentual de identificação na modalidade auditiva (48%) e por fim na modalidade AV (44%). Similarmente ocorre com a identidade Patrícia, que apresenta o percentual de 30% para a acurada relação entre rótulo e persona na modalidade VI, seguido pelos percentuais de identificação nas modalidades A e AV (ambos 22%).

Em contraponto, observa-se que a persona Fera, com maior índice de identificação geral, foi mais facilmente identificada na modalidade AU (99%), seguida da modalidade AV (90%), e obteve menor percentual de correta rotulação na modalidade exclusivamente visual (68%). Esses resultados aparentam dialogar coerentemente com os estímulos utilizados, visto que nestes a persona esboça um sorriso e uma risada, que serão descritos visualmente na seção 5.2.2.

A persona Hedwig, por fim, foi melhor identificada na modalidade AV (82%) e, em sequência, na modalidade AU. A modalidade VI (62%), contudo, se revelou como o contexto de menor índice de identificação desta persona, assim como o observado com a persona Fera.

Todavia, inferencialmente, as análises estatísticas revelaram que somente os índices de identificação da Fera na modalidade visual (p-valor <0.001), da Fera na modalidade audiovisual (p-valor = 0.04) e do Hedwig na modalidade visual (p-valor = 0.006) apresentaram relevância estatística.

Em suma, nota-se que não houve padrão na gradação da relevância das modalidades no processo de identificação das personas. Destaca-se, ainda, que foi reiterada uma

dissociação entre os percentuais de identificação das personas mais facilmente identificadas, Hedwig e Fera, e das personas percebidas em dissonância aos rótulos propostos, Dennis e Patrícia.

5.2 Análise da produção audiovisual

Ao abordar a produção no *corpus* deste estudo, foram analisadas a prosódia acústica impressa pelas diferentes vozes interpretadas pelo dublador, a partir das médias de F0 de cada persona e da descrição qualitativa de seus contornos, e a prosódia visual interpretada pelo ator na caracterização das diferentes personas, a partir da descrição dos gestos faciais e de cabeça de 1 enunciado de cada persona.

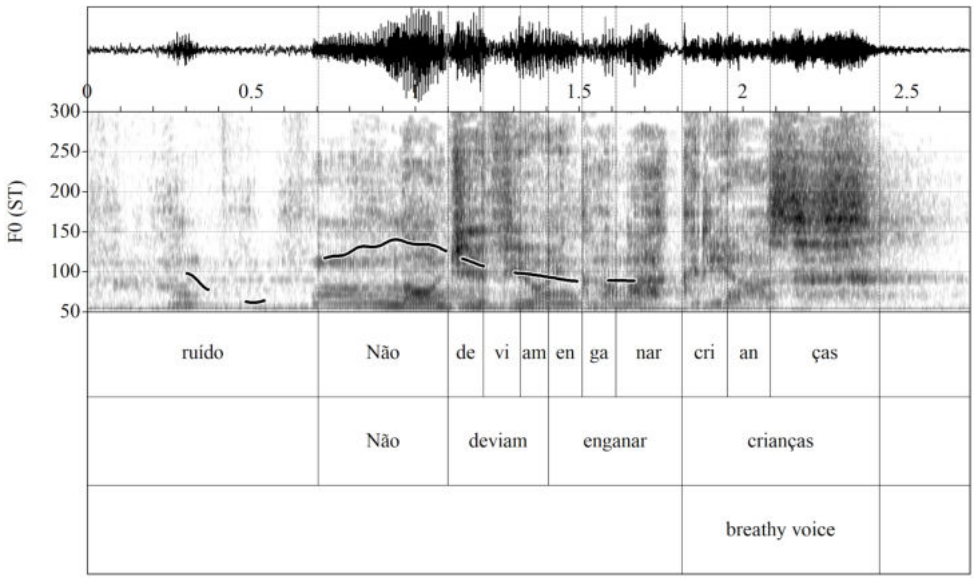
5.2.1 Análise acústica e melódica

5.2.1.1 Dennis

Em primeiro plano, no tangente à caracterização acústica das qualidades de voz, a persona Dennis apresentou um valor médio de F0 de 102 Hz.

Em seguida, a descrição qualitativa do contorno melódico da qualidade de voz dessa persona foi realizada a partir do enunciado “Não deviam enganar crianças” (Figura 16). A progressão melódica da curva de F0 do enunciado apresenta um movimento ascendente inicial, seguido de um alongamento da primeira tônica vocabular e um movimento descendente até a sílaba “-am”, onde a linha melódica se mantém até o fim do vocábulo “enganar”. Na última palavra do enunciado, percebe-se uma qualidade de voz soprosa, *breathy voice*, não captada pelo programa Praat. A partir do trabalho com o *corpus,* destaca-se que esta soprosidade ocorre recorrentemente nos enunciados desta persona.

Figura 16 – Contorno melódico do enunciado "Não deviam enganar crianças", produzido pela persona Dennis.

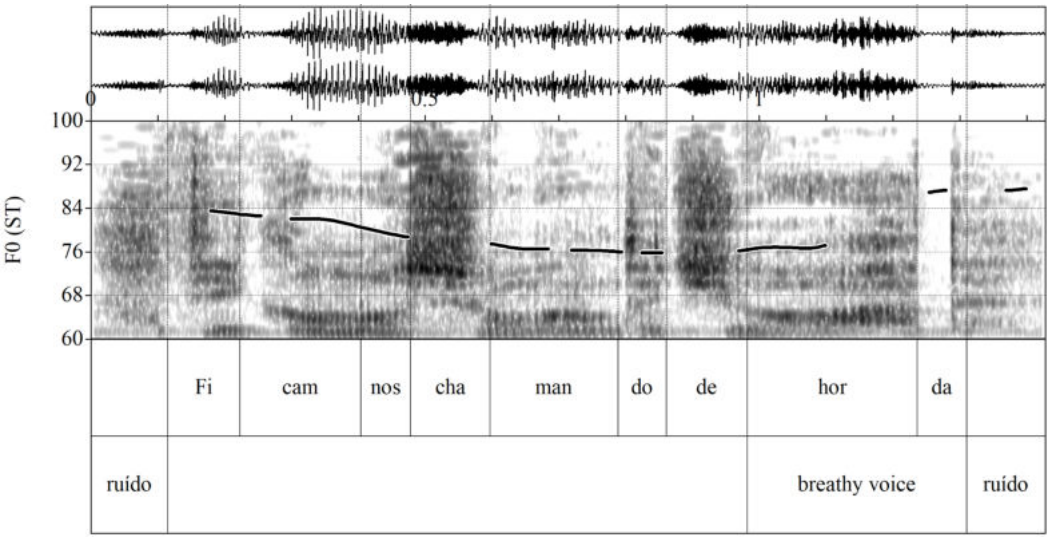


Fonte: *Software* Praat

Ao seguir para a análise multimodal da obra cinematográfica, foram selecionados outros dois enunciados (devido ao critério de enquadramento de câmera): (1) "Ficam nos chamando de horda" – figura 17 – e (2) "Os outros, sabe?" – figura 18.

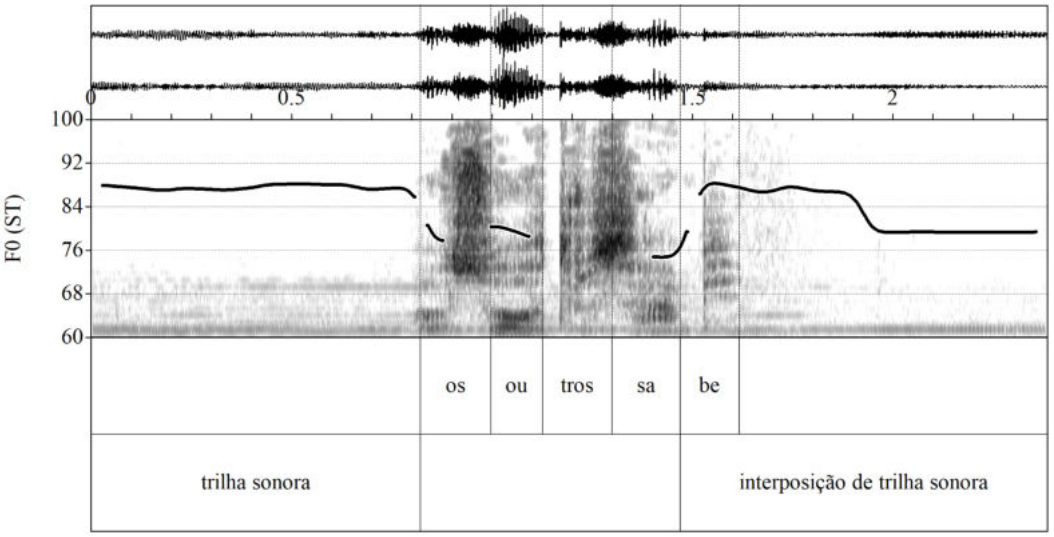
Em (1), exemplifica-se novamente a voz soprosa, característica de final de enunciado, como recorrente nos vocábulos finais da entoação melódica desta persona, sendo este elemento abordado, assim, como uma característica da sua qualidade de voz. Em (2), por sua vez, destaca-se a subida da curva na primeira tônica do enunciado, que será descrito interligado à prosódia visual em 5.2.2.1. Também observa-se o movimento ascendente na tônica final devido à interposição de ruído e à entoação do enunciado interrogativo.

Figura 17 – Contorno melódico do enunciado "Ficam nos chamando de horda", produzido pela persona Dennis.



Fonte: *Software* Praat

Figura 18 – Contorno melódico do enunciado "Os outros, sabe?", produzido pela persona Dennis.



Fonte: *Software* Praat

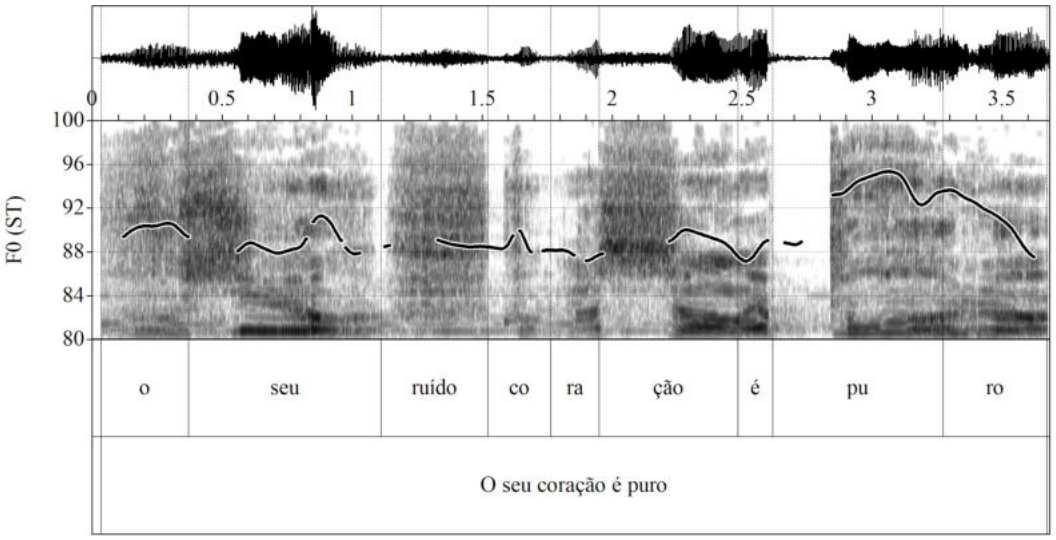
5.2.1.2 Fera

Inicialmente, quanto ao parâmetro acústico de frequência fundamental, a persona Fera apresentou a média de 231 Hz. Esse maior valor, em contraste com a voz da persona Dennis (considerada a voz de registro de referência do dublador), não foi associado a uma percepção mais aguda da entoação mas a um maior esforço vocal empregado (Gussenhoven, 2004).

A partir do enunciado "O seu coração é puro", o alto esforço vocal que caracteriza esta qualidade de voz é percebido nas oscilações melódicas do contorno analisado. Nota-se,

ainda, o salto melódico que ocorre na tônica do vocábulo “puro”, marcando ênfase (vide figura 19).

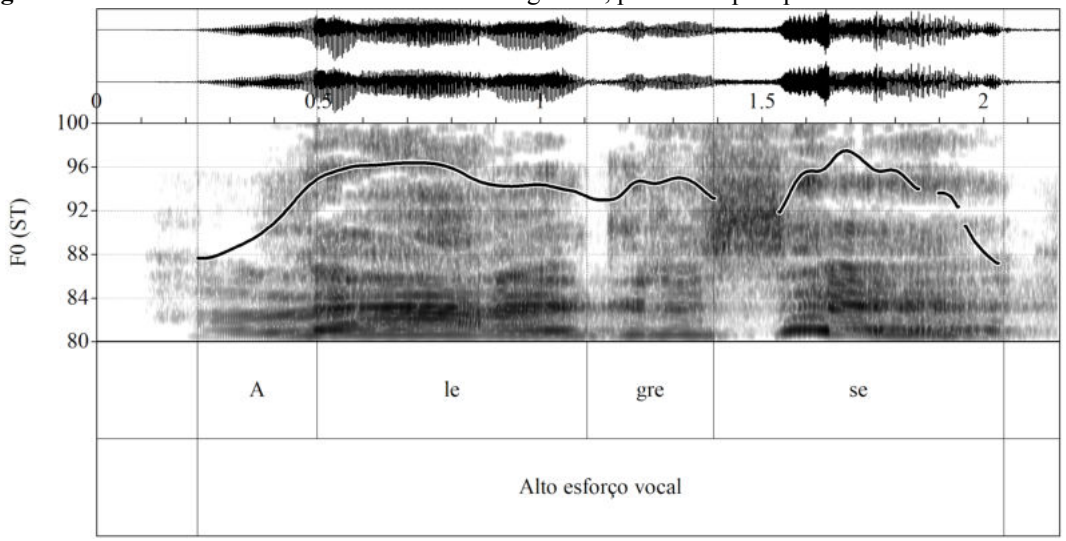
Figura 19 – Contorno melódico do enunciado “O seu coração é puro”, produzido pela persona Fera.



Fonte: *Software* Praat

Ademais, analisa-se o enunciado “Alegre-se” – fig. 20. Neste contexto, destaca-se, novamente, o maior esforço vocal empregado na enunciação, indicado pelo maior valor de F0. Também, observa-se o prolongamento das sílabas “-le” e “-se”, o que será melhor descrito na seção *5.2.2.2,* em diálogo com a prosódia visual do recorte de cena.

Figura 20 – Contorno melódico do enunciado “Alegre-se”, produzido pela persona Fera.



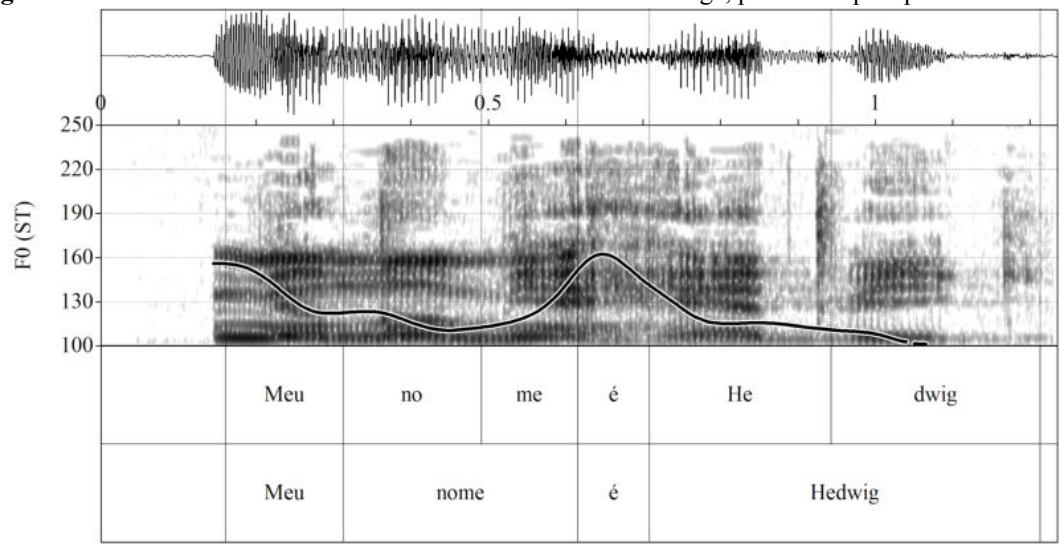
Fonte: *Software* Praat

5.2.1.3 Hedwig

Primeiramente, a persona Hedwig apresentou média de F0, a partir de todas as tônicas de seus enunciados menos ruidosos, de 131 Hz.

Ao considerar o enunciado "Meu nome é Hedwig" para a descrição qualitativa do contorno melódico, percebe-se uma curva iniciada por ataque alto e com grande variação melódica como ilustrado, mais especificamente, pelo pico melódico no vocábulo "é" (Figura 21). Há, aqui, a mudança de núcleo do enunciado da sílaba "He", última tônica do enunciado, para a sílaba "é", devido à ênfase realizada pelo falante.

Figura 21 – Contorno melódico do enunciado "Meu nome é Hedwig", produzido pela persona Hedwig.



Fonte: *Software* Praat

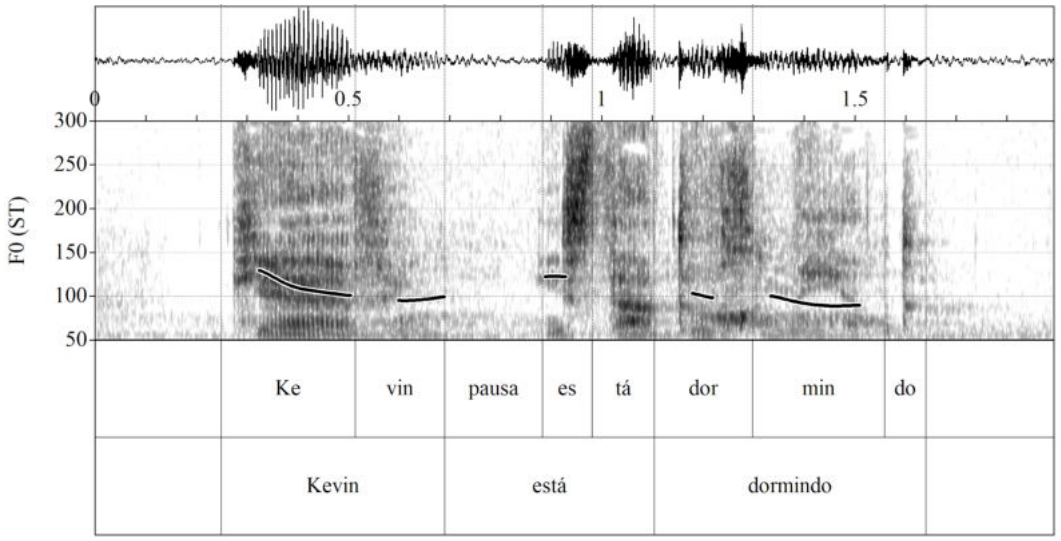
5.2.1.4 Patrícia

Para a persona Patrícia, por sua vez, foi encontrado o valor de 102 Hz referente à média de F0 de seus enunciados. Devido ao valor igualmente atribuído à persona Dennis, a análise aqui conduzida considerou a não intencionalidade do dublador em representar uma voz feminina, visto que, por se tratar de uma obra audiovisual, outros recursos como figurino, gestos e trejeitos também constituem a caracterização das personas.

Em primeiro momento, a descrição melódica qualitativa da qualidade de voz desta persona foi realizada a partir do enunciado "Kevin está dormindo", devido a menor incidência de ruídos de fundo – figura 22. Nesse, enunciado, é possível

observar uma curva de F0 com pouquíssima oscilação melódica, apresentando uma entoação achatada descrita como modulação *flat*.

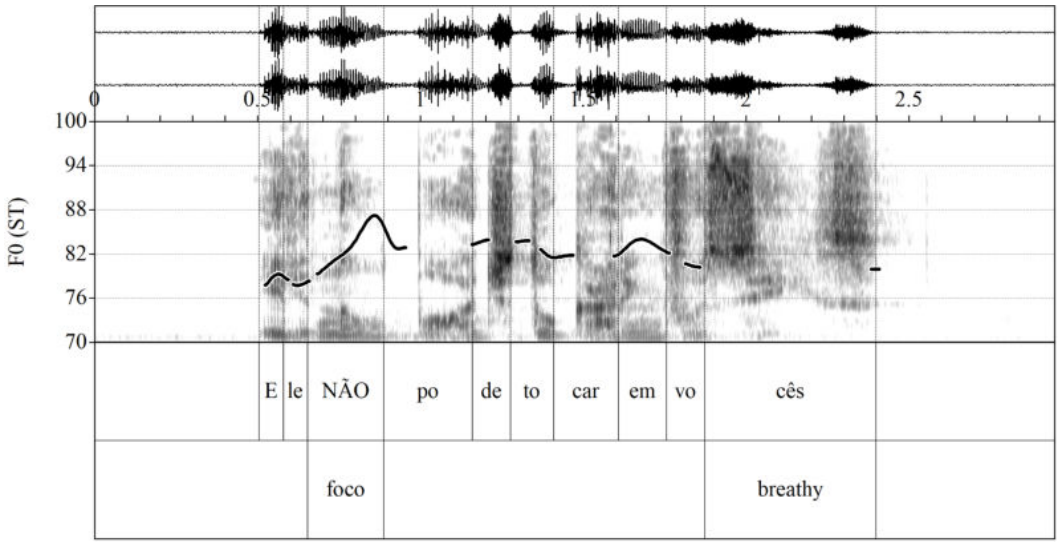
Figura 22 – Contorno melódico do enunciado "Kevin está dormindo", produzido pela persona Patrícia.



Fonte: *Software* Praat

Posteriormente, para a análise multimodal, foram descritos, ainda, os enunciados (1) "Ele não pode tocar em vocês" , (2) "Ele sabe disso" e (3) "Hum hum" – em entoação expressivamente negativa. Na curva melódica de (1), percebe-se uma ênfase no vocábulo "não" e, em seguida, novamente uma modulação com pouca variabilidade tonal (Fig. 24).

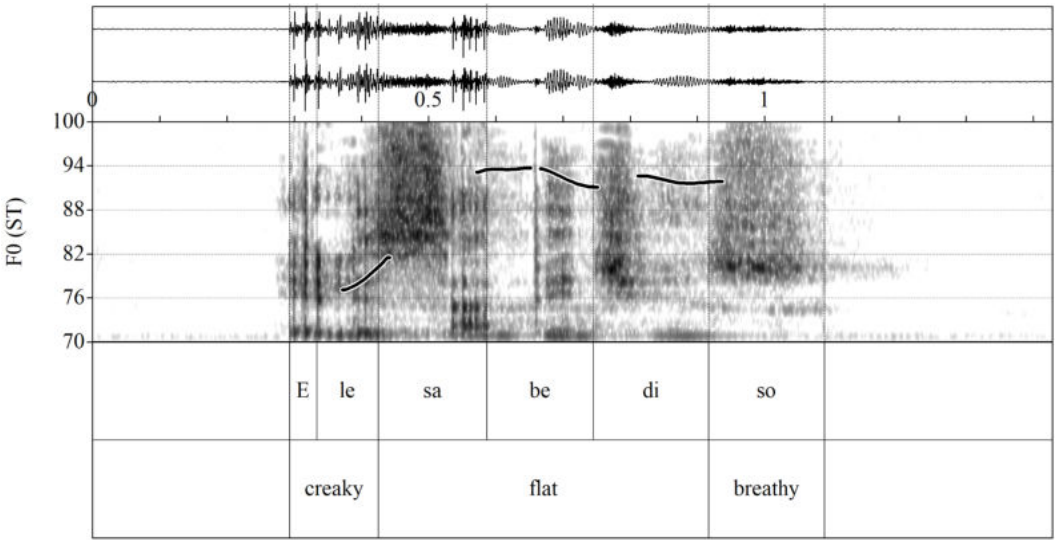
Figura 23 – Contorno melódico do enunciado "Ele não pode tocar em vocês", produzido pela persona Patrícia.



Fonte: *Software* Praat

Em (2), o enunciado é introduzido por uma qualidade de voz crepitada, *creaky*. Logo, após, há um movimento ascendente da linha melódica na extensão sílaba "sa-", sem que haja um movimento descendente posterior, o enunciado se mantém em uma F0 mais alta (Figura 23). É importante ressaltar que, apesar do salto melódico no início da curva, a modulação *flat* se mantém em toda a extensão do enunciado após a subida.

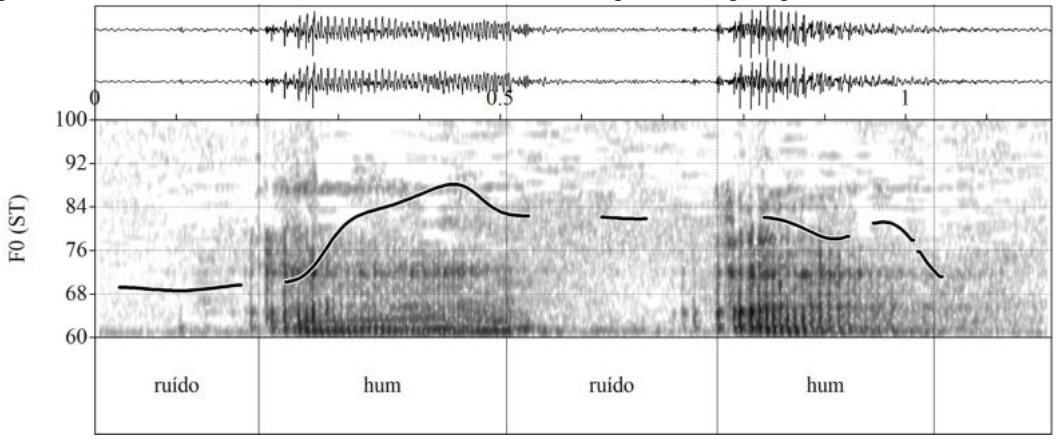
Figura 24 – Contorno melódico do enunciado "Ele sabe disso", produzido pela persona Patrícia.



Fonte: *Software* Praat

Por fim, em (3), há um movimento ascendente no começo do enunciado que se estende por toda a primeira sílaba, a tônica (Figura 25). A curva entoacional inicia então um movimento descendente na sílaba seguinte, final, que decai progressivamente até o fim do enunciado.

Figura 25 – Contorno melódico do enunciado "Hum hum", produzido pela persona Patrícia.



Fonte: *Software* Praat

5.2.2 Análise visual

A descrição qualitativa da produção visual deste trabalho aborda a enumeração de todas as AUs percebidas na análise dos quatro vídeos selecionados, um por persona, seguida da pormenorização das unidades de ação que estabelecem uma relação direta com a entoação prosódica. Assim, serão descritas as AUs que, no processo de alinhamento entre vídeo, áudio e notações de gestos e expressões faciais, apresentaram sincronia entre *apex* do *stroke* e a orientação da linha melódica no mesmo instante do enunciado.

5.2.2.1 Dennis

Para a descrição visual da persona Dennis, foi selecionado o trecho que contém os enunciados (1) “Ficam nos chamando de horda” e (2) “os outros, sabe?”. Ao longo da cena foram descritas as seguintes unidades de ação: *head tilt* (FACS, 2002; M3D, 2025), AU 1, AU 2, AU 4, AU 12, AU 17, AU 25, AU 43, AU 45, AU 54 e AU 62 (FACS, 2002).

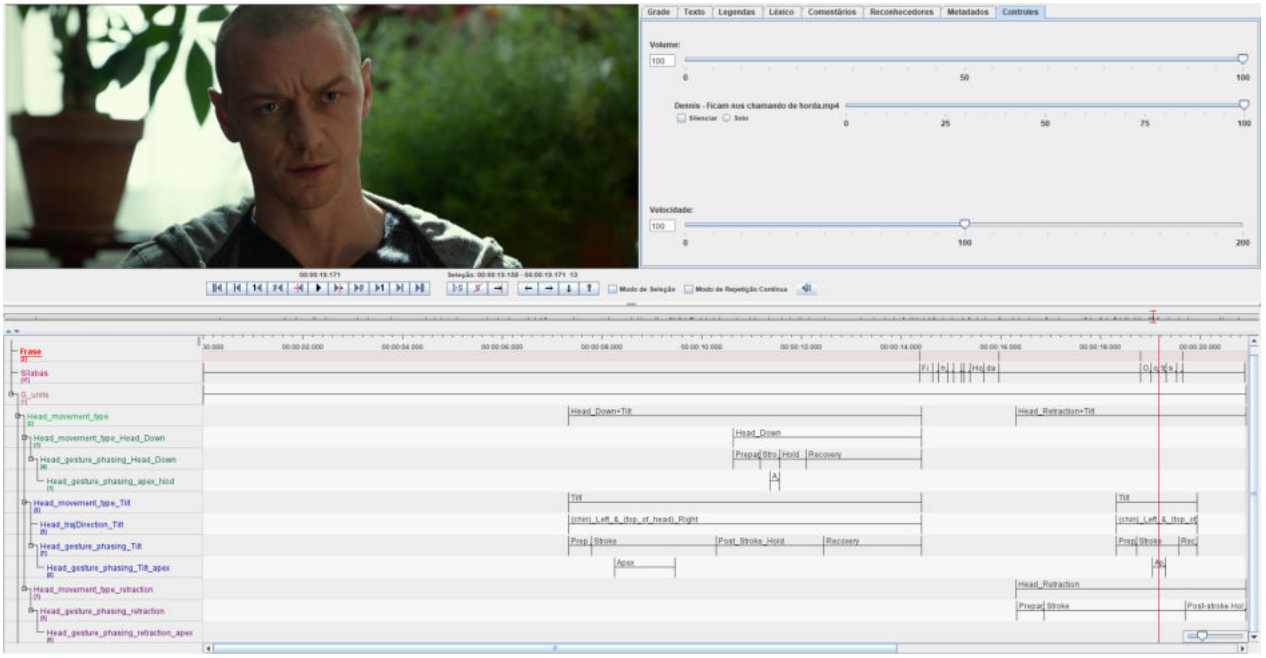
No momento pré-fala e no silêncio entre os enunciados foram observados dois movimentos de lábio: *corner puller* (AU 12), referente à extensão do canto dos lábios em direção à região superior do rosto, e *lip part* (AU 25), que demarca a separação do lábio inferior do lábio superior.

Ainda nos trechos de silêncio da cena, há duas ocorrências da *chin raiser* (AU 17), movimento de contração do queixo de baixo para cima, antes da fala e entre enunciados.

No tangente à movimentação de cabeça, foi utilizada a notação do do sistema FACS combinada ao manual M3D. Foi demarcada um abaixamento de cabeça, *head down* (AU 54), e uma inclinação translacional do topo da cabeça para direita e o queixo para a esquerda, *tilt* (AU 56), antes da fala dos enunciados.

Depois, novamente ocorre o movimento de *head down*. Este começa no silêncio entre enunciados e se estende até o fim da cena. O movimento de *head tilt*, por sua vez, também tem uma segunda ocorrência no trecho. Ele é iniciado pouco antes do sintagma “os outros” e é finalizado simultaneamente ao enunciado. É importante destacar que o *apex* desta unidade de ação, momento de pico do movimento, ocorre no vocábulo “outros” (Figura 26).

Figura 26 – Apex do movimento de *Tilt*, com a cabeça, realizado pela persona Dennis.

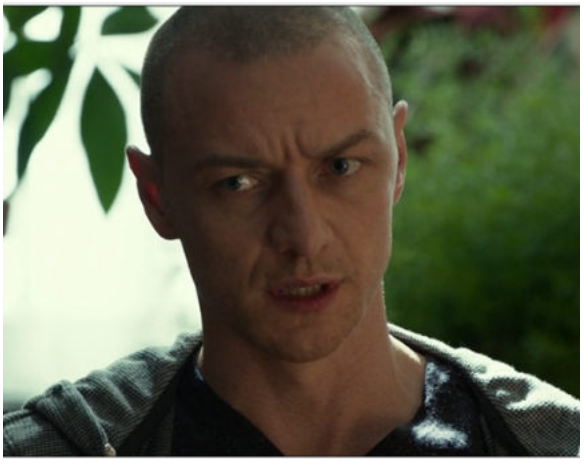


Fonte: Captura de tela do *software* ELAN.

Ao abordar as unidades de ação referentes aos olhos, foi observada a ocorrência de três AUs diferentes. Inicialmente, destaca-se o piscar de ambos os olhos, *eye blink* (AU 45), que ocorre de maneira repetida e sequencial nos primeiros segundos da cena, previamente à enunciação da fala. Em seguida, há o fechamento dos olhos por um tempo mais extenso que o de objetivamente piscar, *eye closure* (AU 43). Destaca-se que a AU 43 ocorre simultaneamente ao movimento de *head down*, antes dos enunciados.

A última AU relacionada aos olhos foi a virada de olho para a direita, *eye turn right* (AU 62), observada na figura 27. Referente a este movimento, foi demarcada o *apex* da sua ocorrência no vocábulo "outros".

Figura 27 – Apex do movimento de *Eye turn right*, com os olhos, realizado pela persona Dennis.



Fonte: Captura de tela do *software* ELAN, editada pela autora.

Em último plano, foram descritos três movimentos de sobrancelha no estímulo visual referente à persona Dennis. Previamente à análise, é importante ressaltar que o sistema FACS segmenta a movimentação muscular de sobrancelha em duas partes: interna e externa. Assim, a primeira AU demarcada foi o levantamento da parte externa da sobrancelha esquerda, *outer brow raiser* (AU 2), que se mantém por toda a extensão dos enunciados. Simultaneamente, há a ocorrência do abaixamento da parte interna da sobrancelha, *inner brow lowerer* (AU 4), até a metade do primeiro enunciado e depois de uma parte do intervalo de silêncio até o fim do último enunciado. Essa configuração se mantém por quase toda a cena , imprimindo uma imagem de arqueamento constante da sobrancelha esquerda.

Por fim, foi observado o levantamento de ambas as sobrancelhas, decorrente da sustentação da AU 2, por parte do ator, na sobrancelha esquerda, em combinação com a realização da mesma AU na sobrancelha direita e com o levantamento da parte interna de ambas as sobrancelhas, *inner brow raiser* (AU 1). Foi identificado, ainda, que o apex dessa co-ocorrência de unidades de ação recai sobre a entoação do vocábulo "horda" – fig. 28.

Figura 28 – Apex do movimento de *Inner brow raiser*, com as sobrancelhas, realizado pela persona Dennis.

Fonte: Captura de tela do *software* ELAN, editada pela autora.

5.2.2.2 Fera

Objetivando a análise qualitativa da prosódia visual da persona fera, foi selecionado o enunciado "Alegre-se", seguido de uma risada. Nesta cena, foram observadas as seguintes unidades de ação: *head nod* (M3D, 2025), AU 4, AU, 7, AU 9, AU 10, AU 12, AU 27, AU 31 e AU 43 (FACS, 2002).

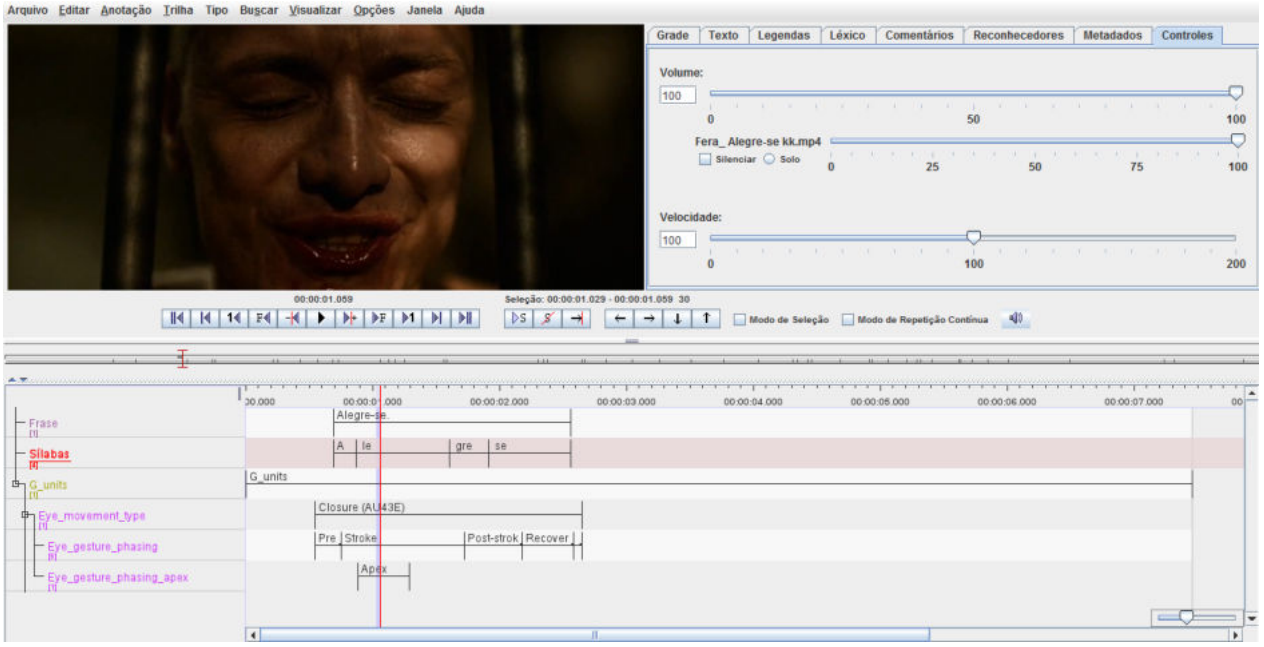
Para esta persona, foi destacado como diferencial a quantidade de ações múltiplas ocorrendo simultaneamente e o grau de intensidade da realização das unidades de ação. Após a fala enunciada, por exemplo, a persona expressa uma risada. Nessa ocasião, há o franzimento do nariz, *nose wrinkle* (AU 9), que consiste na contração da musculatura do nariz próxima aos olhos e sobrancelhas. Contudo, esta ocorre de maneira muito acentuada sendo categorizada no nível E de intensidade sob o critério de *máxima evidência* na escala de evidência do movimento.

Devido a essa intensidade, outras AUs ocorrem concomitantemente à AU 9. Foram descritos o rebaixamento das sobrancelhas, – *brow lowerer* (AU 4) – o aperto das pálpebras – *lid tightener* (AU 7) – e a suspensão do lábio superior – *upper lip raiser* (AU 10). É válido destacar que a AU 10 também foi categorizada no nível E de intensidade.

No tangente à região da boca – lábios e mandíbula – foram identificados 3 AUs para além da AU 10. Primeiramente, observa-se que, durante toda a extensão do trecho selecionado, o ator está sorrindo, movimento que é conduzido pela suspensão dos cantos da boca, *corner puller* (AU 12). O *apex* deste movimento ocorre na risada/gargalhada final. Ao analisar a mandíbula, foi descrito uma contração ou trincamento da mesma, *jaw clencher* (AU 31), desde o começo da fala até a preparação para a risada. Em seguida, já há a ocorrência da abertura da mandíbula, *jaw opening* (AU 27), atrelada à intensificação da risada.

Partindo para a abordagem da face superior, foi analisada a ação de fechar os olhos, *eye closure* (AU 43) – mais uma ação que é categorizada no nível E. O movimento começa e termina na mesma extensão da duração da fala. Ainda, seu *apex* ocorre na sílaba tônica "-le" e dura, aproximadamente, metade do tempo de duração da sílaba falada, como pode ser observado na figura 29.

Figura 29 – Apex do movimento de *Eye closure*, com os olhos, realizado pela persona Fera.



Fonte: Captura de tela do *software* ELAN.

Enfim, em relação aos movimentos de cabeça, foi realizada a descrição da unidade de ação *head nod* (Figuras 30 e 31). A mesma começa no meio da tônica "-le" e a cabeça segue retornando à posição inicial até o fim do vocábulo, não apresentando fase de relaxamento. Além disso, esta ação é iniciada simultaneamente a realização da AU 43E, em que o fechamento do olhos e o impulso da cabeça para trás ocorrem concomitantemente.

Figuras 30 e 31 – Movimento de *Head nod*, com a cabeça, realizado pela persona Fera.

Fonte: Captura de tela do *software* ELAN, editada pela autora.

5.2.2.3 Hedwig

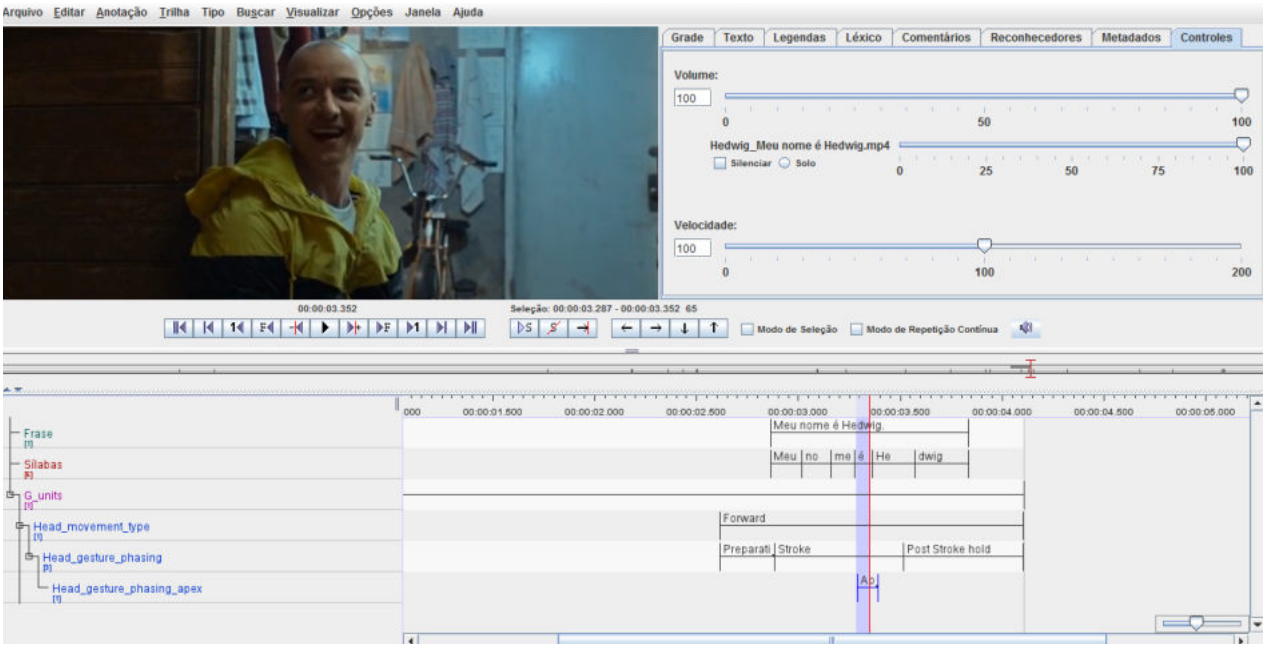
Para a descrição qualitativa da prosódia visual da persona Hedwig, foi utilizada a cena de introdução da persona na obra cinematográfica. Neste contexto, foi selecionada a fala "Meu nome é Hedwig" e ao longo da análise, foram detectadas as seguintes unidades de ação: AU 5, AU 12, AU 25, AU 26 e AU 57 (FACS, 2002). Esta persona foi caracterizada pela proeminência de AUs que se entendem por todo o trecho.

Inicialmente, no tangente à face superior, foi observado o movimento de suspensão das pálpebras, *upper lid raiser* (AU 5), que se estende por todo o enunciado e está atrelada a percepção de olhos meio arregalados.

Em seguida, em relação à face inferior, foram descritas 3 unidades de ação. Primeiramente, percebe-se a suspensão dos cantos do lábio, *lip corner puller* (AU 12), que tem duração equivalente à extensão do enunciado. Há, ainda, três ocorrências da separação dos lábios que não antecedem nenhuma fala, *lip part* (AU 25). Esta AU, por fim, ocorre simultâneamente às três ocorrências do relaxamento de mandíbula, *jaw drop* (AU 26).

Enfim, foi analisada a movimentação de cabeça realizada para frente, pelo ator. Esta unidade de ação é descrita pelo M3D como protrusão, e abordada pelo FACS como *head forward*, referente a AU 57. A AU é preparada antes do enunciado e o *stroke* ocorre do início da fala até a tônica "He". Percebe-se que na realização da pós-tônica a AU já está na fase de hesitação pós-movimento e não há fase de relaxamento. O *apex* desta AU ocorre na pré-tônica "é", como pode ser observado na figura 32.

Figura 32 – Apex do movimento de *Head forward*, com a cabeça, realizado pela persona Hedwig.



Fonte: Captura de tela do *software* ELAN.

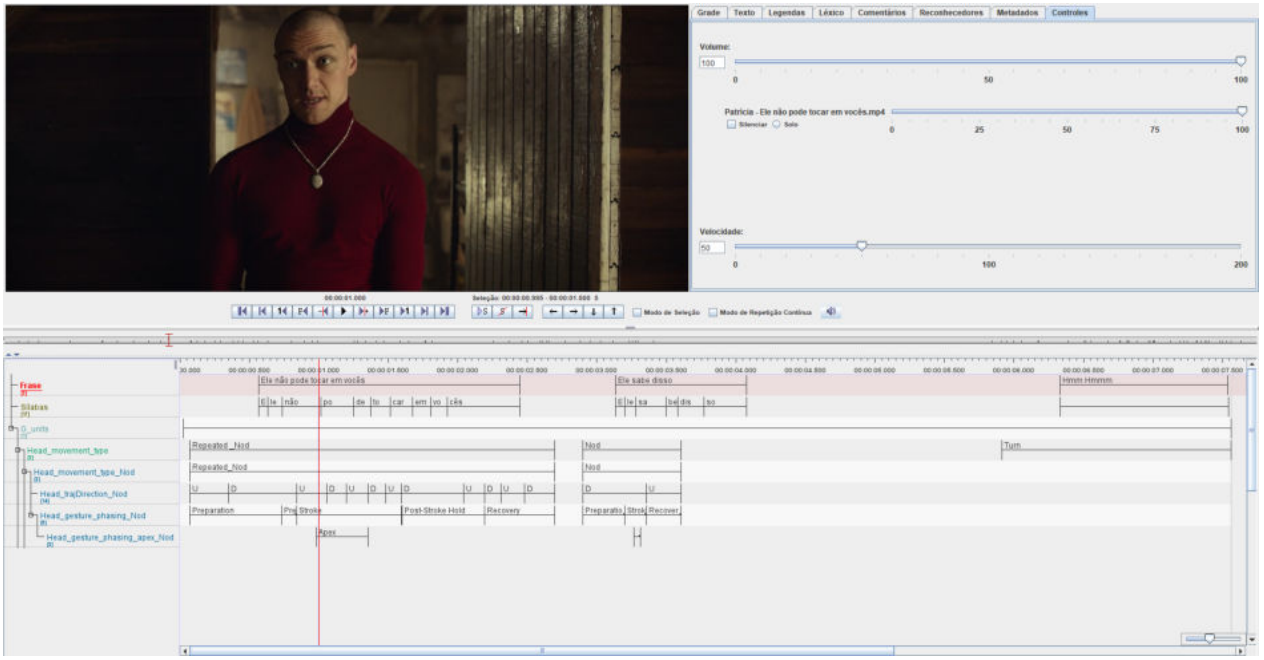
5.2.2.4 Patrícia

Em último plano, para a descrição da persona Patrícia foi utilizada a cena em que são apresentados os enunciados (1) “Ele não pode tocar em vocês”, (2) “Ele sabe disso” e (3) “Hum hum” – (3) em entoação que indica negação. Neste trecho, foram descritas as seguintes unidades de ação: *head nod, head turn* (M3D, 2025), AU 1, AU 2, AU 15, AU 17, AU 45 e AU 52 (FACS, 2002).

Acerca das AUs realizadas na face superior, foi observado o piscar, AU 45, repetitivo ao final do último enunciado. Essa ação se torna saliente pelo fato da persona não piscar em nenhum momento anterior a este, no trecho selecionado. Além desta, foi descrita a ocorrência simultânea do levantamento das partes internas e externas das duas sobrancelhas, AU 1 e AU 2, que se mantém pela extensão completa da cena.

Em sequência, ao abordar os movimentos de cabeça, foram delimitadas 3 AUs. A primeira unidade descrita é o aceno, *head nod*, que ocorre duas vezes. A sua ocorrência inicial foi delimitada como anterior ao enunciado (1), se estendendo até o fim do mesmo. Destaca-se que, neste trecho, a ação foi caracterizada como repetitiva, visto que é realizada mais de uma vez sequencialmente. O segundo *nod* ocorre isoladamente e se estende da metade do intervalo entre (1) e (2) ao fim do vocábulo “sabe”, parte do segundo enunciado. Quanto a esta AU, por fim, foi observado o *apex* do primeiro *nod* no vocábulo “pode” e o *apex* do segundo *nod* no início da frase (2), recaindo sobre “Ele”, como pode ser observado na figura 33.

Figura 33 – Apex do movimento de *Head nod* com a cabeça, realizado pela persona Patrícia.



Fonte: Captura de tela do *software* ELAN.

Ainda na descrição das AUs de cabeça, destaca-se a ocorrência articulada das AUs 52 e *head turn* no desenvolvimento da prosódia visual de maneira integrada à entoação da prosódia acústica. A AU 52 é realizada simultaneamente ao primeiro "Hum" do enunciado três, e seu apex se encontra no início do movimento (Fig. 34). Imediatamente após a recuperação do movimento, sem que haja uma fase de relaxamento, a unidade de ação *head turn* é iniciada e se estende até o fim da frase entoacional.

Figura 34 – Apex da combinação dos movimentos de *Head tilt* e *Head turn,* com a cabeça, realizados pela persona Patrícia.

Fonte: Captura de tela do *software* ELAN, editada pela autora.

Por último, foram observadas duas AUs referentes à região inferior da face. Foi descrita a depressão dos cantos da boca que forma um "u" invertido com os lábios, *lip corner depressor* (AU 15) se estendendo por todo o enunciado (3). A ocorrência do seu *apex*, contudo, recai apenas sobre o primeiro "hum". A AU 17, por sua vez, se refere à contração do queixo de baixo para cima, *chin raiser*, e foi registrada em duas ocorrências: uma simultânea à AU 15, apresentando mesma extensão e posição de *apex*, e outra no começo no enunciado (1), em coocorrência com o *nod* e com o *apex* recaindo sobre o vocábulo "vocês" – Fig. 35.

Figura 35 – Apex da combinação dos movimentos de *Chin raiser* e *Lip depressor,* com o queixo e os lábios, realizados pela persona Patrícia.

Fonte: Captura de tela do *software* ELAN, editada pela autora.

6. Considerações finais

A fim de investigar a correspondência entre a interpretação acústica do dublador e a interpretação visual do ator, este trabalho se desenvolveu a partir de sete objetivos específicos. Inicialmente, buscamos caracterizar acusticamente, a partir do parâmetro de F0, cada uma das quatro vozes performadas pelo dublador. Nossa hipótese era de que as diferentes personas apresentariam caracterização acústica distinta.

Realizamos essa análise a partir da segmentação dos arquivos de áudio com todos os enunciados de cada persona (no *software* Praat), após exclusão de enunciados com muito ruído de fundo, e do cálculo da média de F0 por persona. Para as personas Fera e Hedwig, foram encontrados diferentes valores de média de frequência fundamental, 231 Hz (maior esforço vocal) e 131 Hz (segundo maior valor obtido), respectivamente. O mesmo não ocorreu para as personas Dennis e Patrícia, visto que ambas apontaram equivalente valor, 102 Hz, de média. A partir desses resultados, observamos que a voz da persona Patrícia é entoada no mesmo *pitch* em que o registro de referência do dublador, persona Dennis. Desse modo, concluímos que o dublador manteve o enfoque da interpretação na caracterização da personalidade da persona Patricia, sendo o quesito "feminilidade" construído a partir dos elementos visuais da obra.

Em segundo plano buscamos realizar a análise qualitativa dos contornos melódicos das personas interpretadas sob a hipótese de que estas seriam caracterizadas por padrões melódicos específicos e distintos. Neste quesito, utilizamos o *software* Praat para gerar a imagem dos contornos melódicos selecionados e pudemos concluir que cada persona é interpretada pelo dublador com uma diferente propriedade vocal e entoacional. A persona Dennis foi caracterizada por uma *breathy voice* no final dos enunciados. A persona Fera, por sua vez, teve uma entoação caracterizada pelo alto esforço vocal. A persona Herdwig teve seu contorno melódico individualizado por um ataque alto no início do enunciado. Ademais, a persona Patrícia teve sua entoação caracterizada por uma modulação *flat.* Assim, confirmamos a hipótese do emprego de diferentes padrões melódicos para a interpretação de cada persona pelo dublador.

No tangente à percepção, primeiramente objetivamos a coleta de rótulos sobre a interpretação de ouvintes a partir da sua percepção auditiva das vozes em um teste *free labeling* sob a hipótese de que os rótulos atribuídos às diferentes personas sinalizariam diferentes traços de personalidade. Isso se confirmou visto que para a persona Dennis o rótulo

mais recorrente foi “Calmo”, para a persona Fera, “Raivoso”, para a persona Hedwig, “Brincalhão” e para a persona Patrícia, “Cuidadoso”.

A fim de validar estes rótulos propostos no teste de resposta livre, elaboramos e aplicamos um teste de escolha condicionada, com os mesmos estímulos auditivos do teste anterior, em que os juízes deveriam escolher entre os 4 rótulos – calmo, raivoso, brincalhão ou cuidadoso – qual melhor caracterizava a personalidade das vozes escutadas. Nesta etapa, validamos os rótulos “raivoso” e “brincalhão” para as personas Fera e Hedwig, respectivamente, e confirmamos a nossa hipótese de que há distinção perceptiva entre as personas mencionadas. Contudo, o mesmo não ocorreu para as personas Dennis e Patrícia, visto que o índice de acerto na correspondência rótulo-persona para estas figuras não foi significativo estatisticamente. Isso não somente não validou os rótulos propostos como apontou uma não distinção clara entre estas duas personas. Interpretamos, assim, que Dennis e Patrícia são personas com personalidades mais complexas e menos caricatas, enquanto Fera é uma persona com personalidade basicamente bestial e Hedwig é uma persona simplesmente infantilizada.

Na busca de entender a não validação destes 2 rótulos, propusemos um teste multimodal em que 16 estímulos foram apresentados a novos juízes em 3 modalidades: audio, vídeo e audiovisual. Aqui, objetivamos verificar a relevância das modalidades na identificação das personas sob a hipótese de que haveria uma gradação na relevância das modalidades para a identificação das personas: audiovisual > visual > auditivo, já que a modalidade auditiva foi percebida como insuficiente para a distinção entre as personas Dennis e Patrícia. Ademais, objetivamos investigar a consistência perceptiva entre as pistas prosódicas visuais e acústicas realizadas pelo ator e pelo dublador, respectivamente, sob a hipótese de que as pistas prosódicas visuais seriam um facilitador para o reconhecimento das personalidades menos caricatas.

Os índices de identificação destas duas personas, porém, se mantiveram não significativos estatisticamente, independentemente das modalidades. Além disso, não houve padrão na gradação da relevância das modalidades no processo de identificação das personas. Para a persona Dennis, a relevância das modalidades se estabeleceu da seguinte maneira: VI > AU > AV. Para a persona Patrícia, houve grande similaridade mas as modalidades AU e AV representaram um mesmo percentual de identificação (VI > AU = AV). Para a persona Fera houve proeminência da modalidade auditiva, seguida da modalidade audiovisual e, por fim, a modalidade visual (AU > AV > VI) e para a persona Hedwig houve proeminência da

modalidade audiovisual, seguida da auditiva e, por último, a modalidade visual (AV > AU > VI).

Por fim, decidimos analisar qualitativamente a produção visual das personas selecionadas a fim de verficar se este processo confirmaria a complexidade das personalidades com menor índice de identificação nos testes perceptivos. Além de investigar a correspondência entre prosódia auditiva e visual qualitativamente. Esta análise foi realizada no *software* ELAN com auxílio dos manuais FACS e M3D. A partir da descrição visual de 4 recortes de cena, 1 de cada persona, pudemos perceber uma maior gama de unidades de ação por cena para as personas Dennis e Patrícia, com personalidades mais complexas. Além disso, para estas personas, as AUs se distribuem de maneira equilibrada entre ações com valor prosódico relativo ao enunciado acústico e ações de caracterização da persona, enquanto para Fera e Hedwig, personas mais estereotipadas a maioria das AUs era relacionada à caracterização da persona em si.

Desse modo, concluímos que os gestos e expressões, realizados pelo ator, destacam a relevância da análise do processo de produção, ainda que as pistas visuais não tenham sido significativas estatisticamente na percepção e identificação de todas as personas. Na produção, portanto, há correspondência entre os gestos e expressões da prosódia visual, realizada pelo ator, e da prosódia auditiva, realizada pelo dublador, ainda que esta não tenha ocorrido na percepção dos juízes em relação a todas as personas. Por fim, pudemos concluir que os rótulos estabelecidos no teste *free labeling* não foram complexos o suficiente para caracterizar as personas menos caricatas.

Assim, esperamos que o presente trabalho contribua para o desenvolvimento das pesquisas relativas aos recursos acústicos, na dublagem do Português brasileiro, e visuais da prosódia empregados a fim de constituir a expressividade e individualização de personagens em obras cinematográficas. Igualmente, esperamos que a abordagem metodológica multidimensional adotada possa ser aprimorada e flexibilizada para abordagem de estudos de percepção e produção audiovisual multimodal, no geral.

As análises acústicas e visuais desenvolvidas nesta pesquisa estabeleceram um contraste relevante para o aprofundamento da investigação da prosódia em uso no cinema a fim de debater o que uma obra audiovisual intenciona expressivamente e como isso é percebido pelos espectadores. Dessa forma, estudos sobre o contraste entre a prosódia acústica do longa em Inglês original e a consistência entre a interpretação vocal do dublador e a atuação visual do ator em outras línguas são próximos passos de grande significância para

o desenvolvimento das contribuições deste corpus para os estudos da prosódia como recurso expressivo do processo de dublagem.

Referências bibliográficas

ABELIN, A. (1999). Studies in Sound Symbolism. Doctoral dissertation. Gothenburg: Göteborg University.

AUDACITY TEAM. Audacity: versão 3.5.1. Software de edição de áudio. 2024. Disponível em: https://www.audacityteam.org/.

BARBOSA, P. A. Prosódia. São Paulo: Parábola, 2019.

BOERSMA, P.; WEENINK, D. Praat: doing phonetics by computer. 1992-2025. [Computer program]. Disponível em: <http://www.praat.org/>.

BYTEDANCE PTE. LTD. CapCut: versão 2.2.0. Aplicativo de edição de vídeo. 2024. Disponível em: https://www.capcut.com/.

CARNAVAL, M. Focalização no português do Brasil: um estudo multimodal. Rio de Janeiro, 2021. 303 f. Tese (Doutorado em Letras Vernáculas – Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2021.

EKMAN, P.; FRIESEN, W. V.; HAGER, J. C. (2002). The Facial Action Coding System. (2nd ed.) Salt Lake City, UT: research Nexus ebook.

ELAN (Version 7.0) [Computer software]. (2025). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan.

FIGUEIREDO, N. S. Variação pragmática e ecologia das línguas: análise multimodal de atos de fala no espanhol do Paraguai e da Argentina. Tese de Doutorado em Língua Espanhola. Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

FÓNAGY, I. As funções modais da entoação. Tradução de João Antônio de Moraes. Cadernos de estudos linguísticos, Campinas, jul/dez, 1993, pp. 25-65.

FRAGMENTADO; Direção: M. Night Shyamalan. Produção: Blinding Edge Pictures; Blumhouse Pictures. Netflix. 2016. 1h 57min.

GILI FIVELA, B. Multimodal analyses of audio-visual information: Some methods and issues in prosody research. In: Feldhausen I, Fliessbach J, Vanrell MM (Eds.). Methods in prosody: A Romance language perspective (Studies in Laboratory Phonology 4). Berlin: Language Science Press, pp. 83-122, 2018.

GOMES DA SILVA, C. & CARNAVAL, M. Quando a prosódia e a pragmática se encontram: possibilidades de pesquisa e desafios para o ensino. *In: A ciência da linguagem: explorações*

analíticas em aquisição, processamento, significação e diversidade linguística (pp.243). Pontes Editores, 2024.

GUSSENHOVEN, C. Meanings of intonation. In: The phonology of tone and intonations. Cambridge University Press, 2004.

HAYWARD, K. Experimental Phonetics. Harlow, Eng: Longman, 2000.

HINTON, L.; NICHOLS, J.; OHALA, J. J. (eds.). Sound Symbolism. Cambridge: Cambridge University Press. 1994.

LAVER J. The phonetic description of voice quality. Cambridge: Cambridge University Press, 1980.

MADUREIRA, S. ; FONTES, M. A. S. ; CAMARGO, Z. . Sound symbolism, speech expressivity and crossmodality. Signifians (Signifying) , v. 3, p. 98-113, 2019.

MADUREIRA, S.; FONTES, M. A. S. A prosódia da fala expressiva. Miguel Oliveira Jr. Prosódia, Prosódias: uma introdução, Editora Contexto, pp.157-172, 2022.

MORAES, J. A. Fonética. São Paulo: Parábola, 2024.

MORAES, J. A.; RILLIARD, A. Entoação. Miguel Oliveira Jr. Prosódia, Prosódias: uma introdução, Editora Contexto, pp.45-66, 2022.

OHALA, J. J. Sound symbolism, in Proceedings of the 4th Seoul International Conference on Linguistics [SICOL] 11-15 Aug , 98-103. 1997.

R CORE TEAM. R: The R Foundation for Statistical Computing Platform, 2021. <http://www.r-project.org/index.html>.

RILLIARD, A. O. B. 2020. Fala e multimodalidade. In: Verbetes LBASS. Disponível em: http://www.letras.ufmg.br/lbass/.

ROHRER, P. L., TÜTÜNCÜBASI, U., VILÀ-GIMÉNEZ, I., FLORIT-PONS, J., ESTEVE-GIBERT, N., REN-MITCHELL, A., SHATTUCK-HUFNAGEL, S. & PRIETO, P. (2025). Multidimensional Labeling of Gesture in Communication: the M3D Proposal.

TACCHETTI, Madalena. User guide for ELAN Linguistic Anotator version 5.0.0. The latest version can be downloaded from: http://tla.mpi.nl/tools/tla-tools/elan/

TURBOSCRIBE. *TurboScribe*. Ferramenta online de transcrição automática. Plataforma em versão atual. 2025. Disponível em: https://turboscribe.ai/.

WAGNER, P.; MALISZ, Z.; KOPP, S. Gesture and speech in interaction: An overview. Speech Communication. (2014). 57. 209-232. 10.1016/j.specom.2013.09.008.

WORDART. AI Word Cloud Generator | WordArt.com. Disponível em: https://wordart.com/.

ZEHR, J.; SCHWARZ, F. PennController for Internet Based Experiments (IBEX), 2018.<https://doi.org/10.17605/OSF.IO/MD832>.