DOI: 10.17469/O2112AISV000003

ALICE CROCHIQUIA, ANDERS ERIKSSON, SANDRA MADUREIRA

# Dubbing in animated films: challenges and the impact of voice design

Listeners rely on speech vocal cues to judge speakers' age, size, personality, and other paralinguistic and extralinguistic features. In this paper, we focus on two perception experiments investigating how characters' voices in animated films provide indexical information about their individuality to listeners. In the first experiment, we examined how listeners perceived physical, psychological, and social features from voice characteristics. The stimuli consisted of speech utterances by five animated film characters. The second experiment aimed to investigate the potential influence of lexical content on the evaluation of speaker characteristics. This experiment was identical to the first experiment, except the stimuli were transcriptions of the utterances used in the first experiment. Our hypothesis was that the results from the listening test and the reading test would not be congruent, but they turned out to be the exact opposite. This finding is intriguing and claims for further research.

*Keywords*: dubbing, animated films, vocal stereotypes, voice quality, voice analysis.

## 1. *Introduction*

As most of us have experienced and as backed by extensive accounts on the subject (Kreiman & Sidtis, 2011), listeners frequently rely on vocal cues to judge a speaker's characteristics, such as physical appearance and demeanour. At the same time, we also create expectations for how someone will sound based on these features.

Over the last century, research on those links has shed light on the complex interplay between physiological factors and cultural expectations in voice perception. While some expectations, such as women generally having higher-pitched voices, appear universal due to physiological differences, other correlations, such as physical attractiveness and voice quality, seem to vary between languages and cultures (Weirich, 2008; Xu & Lee, 2018).

A common finding in many studies dealing with voice identification has been the high degree of agreement among listeners regarding particular impressions of speakers from their voices, even if they are often inaccurate. These findings of shared expectations and assumptions point towards the phenomenon of vocal stereotypes and their consistent influence on impressions based on voice (Aronovitch, 1976; Imhof, 2010; Mahrholz, Belin & McAleer, 2018).

These insights have significant implications for fields such as voice design and animation, where understanding the nuances of voice perception can enhance the creation of more authentic and engaging characters.

For several reasons, animated films and TV shows are a particularly interesting kind of media for investigating those vocal stereotypes and their nature. Firstly, they allow creators to match physical, social, and psychological features with the expected correlating vocal characteristics, presumably also shared with the intended audience, which works as an effective and quick way to help establish a character to the viewers. These productions also welcome exaggerated vocal performances, which would be out of place in most live-action productions, as a way to enhance the intended impressions of the characters. Lastly, such productions are usually dubbed for distribution in international markets, thus allowing for comparing how the same character sounds in many different languages.

Our research takes a unique approach by exploring how vocal cues are perceived in the context of animated films (Crochiquia, 2020; Crochiquia, Andersson, Fontes & Madureira, 2021; Crochiquia, Andersson, Barbosa & Madureira, 2022; Crochiquia, Andersson, Madureira & Barbosa, 2023), a field that has not been extensively studied. In this paper, we detail the steps in developing the methodological procedures to investigate the role of voice in the construction of animated characters and their features, the use of vocal stereotypes in these types of performances, and how they may match or differ in multiple languages.

Zootopia, the film animated chosen for our investigation, heavily relies on preconceived notions of how its characters are meant to behave, both to confirm and to subvert expectations. The investigation of the vocal performances in it allows us to explore how vocal features are used to reflect biological and social meanings as outlined by hypotheses such as the Frequency Code (Ohala, 1984) and the Sirenic Code (Gussenhoven, 2016) in the different languages in which the film is dubbed. The Frequency Code associates low $f_0$ values with bigness, power, force, hostility, and aggressiveness. In contrast, high $f_0$ values are associated with smallness, submission, and fragility. The Sirenic Code (Gussenhoven, 2016) is related to the voice qualities characterised by air escaping between the vocal folds (breathy and whispery voices). Paralinguistically, breathy and whispery voices can be associated with low excitment and seduction.

Our work investigates those correlations by combining perceptual experiments to assess listeners' impressions of the characters from speech samples and the analysis of their acoustic parameters, perceptual assessment of voice quality, as well as text-only perceptual experiments with transcriptions of the audio samples to investigate the influence of lexical content in those judgments.

## 2. *Dubbing*

Dubbing, as a type of language transfer, most commonly refers to a kind of revoicing where the original voice track of an audiovisual product is entirely removed and

replaced by a new voice track in the target language while keeping the original soundtrack and sound effects in place. This is also called "synchronisation", as this type of audiovisual translation (AVT) often requires the spoken text in the target language to follow the lip movements on screen as closely as possible (Díaz-Cintas, 1999). Dubbing is one of the two most widespread forms of AVT, along with interlingual subtitling.

Generally, national markets are broadly categorised according to which AVT format is preferred for the distribution of foreign productions, usually dubbing or subtitling. The reasons for such preferences are many; among them, we can find population size, rates of literacy, cultural and ideological factors, the potential audience and profit of the production, and foreign language proficiency (Danan, 1991; Díaz-Cintas, 2019). However, the boundaries of this division have become increasingly blurred in the age of digitalisation (Chaume, 2018).

Traditionally, Brazil is one of the markets categorised as a dubbing country, with most foreign media productions arriving on TV, movie theatres, home video and streaming services dubbed in Brazilian Portuguese, including productions from other lusophone countries. On the other hand, Nordic countries are firmly on the subtitling side of that division, where the use of dubbing is almost exclusively limited to productions aimed at younger audiences and computer-animated films (Tveit, 2009; Pedersen, 2010).

Dubbing is a collaborative process involving translators, dialogue writers, dubbing directors, voice actors and sound engineers. Dubbing directors are perhaps the most crucial link in this chain, as they are responsible not only for the initial selection of voice actors and guiding them in their performances, providing them with information about the narrative, (which is often not available to actors beyond the scenes they perform in) but they can also make changes to dialogues to ensure a better fit (Bosseaux, 2018). In choosing a dubber for a character in a film, directors have to consider not only how they match the original voice and the character on the screen but also their ability to work within the highly constrained context of dubbing and to deliver a vocal performance that fits well with the onscreen gestures (Brisola, 2024).

Unsurprisingly, given the greater number of steps and professionals involved in the process of dubbing a film, this type of AVT is both more time-consuming and more expensive than subtitling. Keeping in mind that the quality of a dub can affect how the audiences engage with the production and impact its reception (Bosseaux, 2019), the dubbing of an audiovisual production is an investment that demands careful consideration from studios and distribution companies, and often the final choices for voice actors falls to them (Brisola, 2024).

The choice of voice actors and directions given to them during their performance is often made so with the intent to support the character's features and attitudes displayed within the narrative (Chaume, 2014) as to "enhance the audience's recognition of the character's image" (Liu, Zhang & Liang, 2022: 02) and preserve the so-called 'character synchrony' (Whitman-Linsen, 1992).

However, 'mismatched' voices can also be used in original voice acting and dubbing, usually for comedic purposes, like a physically attractive character having an unpleasant and shrill voice.

As such, dubbing directors and voice actors often rely on vocal stereotypes to portray these characters (Bosseaux, 2015), both to confirm and to subvert expectations (Liu et al., 2022) Such stereotypes are often based on particular voice qualities and prosodic elements, like a grave voice belonging to a large-sized character, but they can also rely on impressions associated with dialects and accents, like the use of the Bergen dialect for tough, brash characters in Norwegian dubbed versions of animated films (Nybakk, 2014).

Despite the complexity of its process, its widespread use and its commercial and cultural impact (especially in dubbing strongholds like Brazil, Germany, Italy and Spain), dubbing has lacked attention and recognition from both academic and professional perspectives for much of its history, especially in comparison with other types of AVT (Sánchez-Mompeán, 2020; Díaz-Cintas, 2015). This landscape has suffered an important, if still slow, shift in the past decade or so.

Recent research works on dubbing, especially within the so-called dubbing countries, have explored many facets of this mode of AVT (Sánchez-Mompéan, 2016; Fasoli, Mazzurega & Sulpizio, 2017), demonstrating that the field is fertile ground for exploration and showing its potential for intersection with many disciplines. Considering the growing interest in dubbing and how it can readily welcome different perspectives, it would be expected that non-lexical elements in speech, like prosodic features and voice quality, would emerge as one of the topics of concern, given how these features carry linguistic, paralinguistic and extralinguistic information and can affect the expression and comprehension of a speaker's identity and attitudes (Kreiman & Sidtis, 2011). However, the subject is still 'virtually unexplored' in AVT studies, with the modest number of academic works on the subject mostly coming from research in phonetics and speech sciences (Sánchez-Mompeán, 2020) and often focused on single-parameter analysis and correlations with character traits (Liu et al., 2022) and within a single language.

## 3. *Our experiments with dubbing data*

A listening and a reading experiment are reported in this paper. These two experiments are part of a project that focuses on the analysis and comparison of the vocal performances of the four main characters in the 2016 computer-animated film Zootopia (also titled Zootropolis in European markets), from Walt Disney Studios, in its original version in English and its dubbing in Brazilian Portuguese and Swedish. Zootopia tells the story of a metropolis of the same name, inhabited by anthropomorphic mammals, and it serves as an allegory about human prejudices and biases. This paper details and discusses the methodological procedures applied to the investigation of the Brazilian Portuguese dubbing of the film.

The film first presents the city as a utopia (as referenced by its name), as we see it through the eyes of Judy Hopps, an idealistic bunny who becomes the city's first prey police officer, a position exclusively occupied by predators until then. However, that utopian view quickly washes away as Judy (and the audience) uncovers the complex dynamics of Zootopia, where prey, despite making up the large majority of the city's population, are often underestimated. At the same time, predators, due to their nature, are often regarded with suspicion by prey.

The film focuses on anthropomorphic mammals whose characterisations and roles in the narrative are directly influenced by the behaviour displayed by those animals in nature or associated with them in popular culture. Among the large cast of characters in the film, we find a lion as the mayor of Zootopia and a sloth as a worker at an overly bureaucratic and slow-moving government agency.

We have chosen these characters for our corpus for two reasons: as the main characters, more material from their speech productions is available for investigation, especially samples that do not feature music or sound effects in the background, as the isolated track with the dialogue lines of the characters was not available, and removing the music and effects would possibly result in some acoustic information from the characters' voices being removed as well; the second reason for our choice of characters was the fact that their physical, psychological and social features were discussed by the film's writers, directors and voice actors in interviews and promotional material for the film, which gave us insight into the choices that guided the actors and dubbers in their performances and the intended impressions these characters are supposed to have on the audience.

Alongside the lead character of Judy Hopps (JH), we have Assistant Mayor Bellwether, who is first presented as a meek but friendly little sheep (B1) who then reveals herself as bitter and conniving (B2). The film's deuteragonist is a red fox named Nick Wilde (NW), who is an easygoing con artist. Chief Bogo (CB), Zootopia's police chief, is a gruff African buffalo. Figure 1 shows the four characters.

As Assistant Mayor Bellwether is revealed as the film's villain, we consider the samples from her productions after that to belong to a different character from the one in the first portion of the film. As such, five different characters are part of our experiments and analyses.

Figure 1 - *The main characters of Zootopia: Judy Hopps & Nick Wilde (top row), Chief Bogo & Assistant Mayor Bellwether (bottom row)*

3.1 Material, data collection, and stimulus preparation procedures of the experiments

As the first step for the experiments and analyses, the video file from the Blu-ray disc of the Brazilian version of Zootopia was extracted using *Format Factory* (FORMATZ, version 4.3.0.0, 2018) to a .mp4 file. The film's English and Brazilian Portuguese tracks were then extracted to PCM files using Adobe Premiere. The same procedures were later used to extract the Swedish audio from the Nordic Blu-ray disc.

The files were then imported into *ELAN* (Max Planck Institute for Psycholinguistics, version 5.4, 2018), where the entire film was transcribed into English and Brazilian Portuguese. Figure 2 shows the transcription process of the dialogue lines in Brazilian Portuguese in *ELAN*.

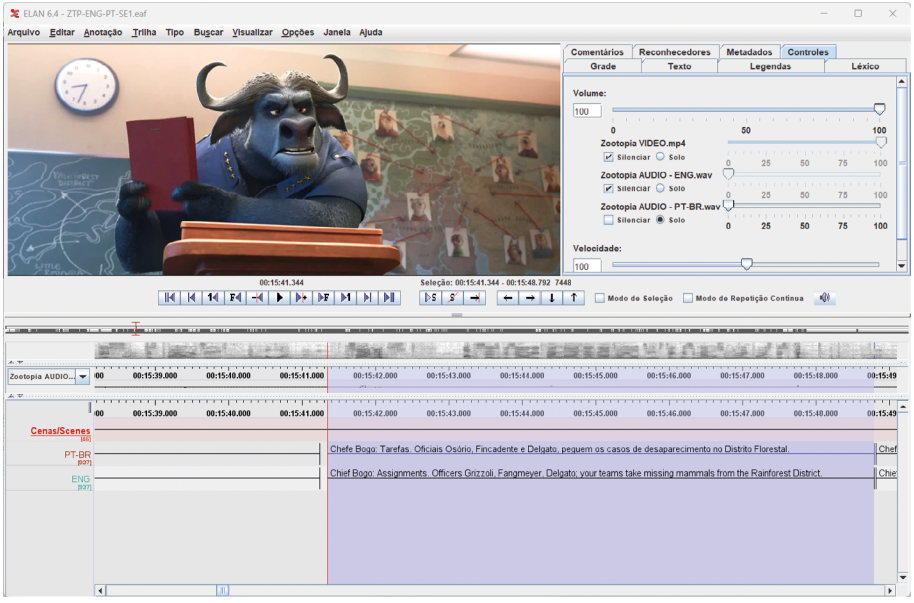
The audio files and annotation tiers were then exported to *PRAAT* (Boersma & Weenink, 2019) for further editing and the selection of samples. Later, the transcription of the Swedish audio was done on *PRAAT*.

For the experiments, five audio samples (and later the five corresponding text samples) were produced, one for each character, around 27 seconds in length. Each sample featured three dialogue lines from different scenes in the film.

No masking techniques were used to isolate voice from linguistic information in the samples as we believed the audio would feel odd to the listeners and affect their answers. In section 3.4, we will discuss our solution to investigate the influence of semantic and syntactic content in the audio samples.

The audio samples were then uploaded to the audio streaming platform SoundCloud, as the tool chosen for the experiments, SurveyMonkey.com, does not allow audio to be directly uploaded to the platform.

Figure 2 - *Transcription of dialogue lines in ELAN*




## 3.2 Listening Experiment

The Listening Experiment comprised a perceptual test, a phonetic description of voice quality settings, and acoustic data analysis.

### 3.2.1 The perceptual test

The test was conducted entirely online on SurveyMonkey.com. Listeners were initially recruited from the Graduate Program of Applied Linguistics and Language Studies at PUC-SP and were emailed a link to the test. Participants were also told they were welcome to share the link with family, friends, and whoever else they believed would be interested in participating in the test.

Participation in the test was voluntary. Participants were informed of their right to withdraw at any given time and that any data collected in the test would be anonymous past the collection stage. They received no compensation for their participation. Seventy-seven Brazilian Portuguese native speakers, 46 women and 31 men, aged between 20 and 50, participated in the test.

The test was divided into eight pages. The first page contained a brief description of the test, the study, its aims, and a questionnaire to collect background information on the respondents, including whether they had any hearing impairments or had seen the film before.

The instructions for the test and an example of the slide feature they would use for the evaluations were presented on the second page. Respondents were instructed to base their evaluations solely on the voices in the samples and their impressions of them. They were also told to use headphones to complete the test, and to indicate if that had not been the case.

The following five pages each presented a speech sample from one of the characters, using the embedded player from SoundCloud on the page, and 14 sliding scales for rating.

The respondents were asked to evaluate each of the five samples according to the bipolar scales regarding the speakers' size, age, temper, attitudes, character, and social and vocal features. They had to do that by using the sliding tool to indicate the point of the scale they felt best fit the character, which was translated into a score from 0 to 100; respondents were not made aware of the score during the test.

The descriptors chosen for the test were based on features displayed by the characters in the film and on descriptions given by the writers, directors, and voice actors in interviews, discussions, and promotional material for the film.

The test did not include pictures or videos of the characters. Table 1 lists the scales and pairs of descriptors in Brazilian Portuguese used in the test, and their translations to English, and Figure 3 shows a few of the sliding scales used in the Experiment, translated to English.

Respondents could only move to the next page once they had made a choice for all 14 scales; however, they were able to return to previous pages if they wished to review or change their answers. The final page featured a message warning the respondents that their answers were final and could not be changed once they had finished the test.

Table 1 - *Scales and Descriptors in Listening Experiment*

| *Scale* | *Descriptors* |
|---|---|
| *Size* | *Pequeno - Grande* (Small - Big) |
| *Age* | *Jovem - Velho* (Young - Old) |
| *Temper* | *Agressivo - Dócil* (Aggressive - Docile) |
| *Attitudes* | *Dominador - Submisso* (Dominant - Submissive) |
| | *Preguiçoso - Ativo* (Lazy - Energetic) |
| | *Durão - Gentil* (Tough - Gentle) |
| | *Charmoso - Desinteressante* (Charming - Uninteresting) |
| | *Bondoso - Maldoso* (Kind - Mean) |
| | *Cauteloso - Audacioso* (Cautious - Bold) |
| | *Alegre - Triste* (Happy - Sad) |
| *Character* | *Honesto - Desonesto* (Honest - Dishonest) |
| *Social features* | *Sofisticado - Simples* (Sophisticated - Simple) |
| *Voice* | *Animada - Monótona* (Lively - Monotonous)<br>*Agradável - Desagradável* (Pleasant - Unpleasant) |

Figure 3 - *Sliding scales in the Listening Experiment*


3.2.2 Description of the voice quality settings

For the perceptual evaluation of voice quality, we have used the Voice Profile Analysis, or VPA, developed by Laver and Mackenzie Beck (2007) to outline the characters' vocal profiles.

The VPA is a comprehensive tool for voice assessment as it considers phonatory, supralaryngeal and prosodic features of a speaker's voice. The analytic unit of the protocol is the setting. The vocal features are evaluated based on their deviation from the neutral setting. This condition indicates no contraction or expansion, no shortening or lengthening of the vocal tract articulators, no extreme variation in muscular tension, no nasality beyond the necessary for linguistic purposes, periodical vibration of the vocal folds under moderate tension and compression and no audible friction (Laver, 1980; San Segundo & Mompean, 2017).

The protocol is divided into six parts, with the first three parts comprising features related to voice quality and the last three relating to vocal dynamics. The features are organised according to the location of indidence of the settings. Assessments are made primarily by listening to audio samples, which can be combined with corresponding video samples.

The assessment with the protocol is done in two stages. The first step is to identify the presence of neutral and non-neutral settings in a speech sample. In the second step, the marked non-neutral settings are rated in a scalar degree, from lesser (1) to more significant (6). Degrees 1 to 3 are moderate, and 4 to 6 are considered extreme; however, some settings are marked only for their absence or presence. Figure 4 shows the voice quality settings, and Figure 5 shows the prosodic elements.

Two phoneticians with extensive experience with VPA described the voice quality characteristics of the characters chosen for our analyses.

Figure 4 - *Vocal Tract, Muscular Tension and Phonation settings in VPA (Laver & Mackenzie Beck, 2007)*

| | FIRST PASS | | SECOND PASS | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Neutral | Non-neutral | SETTING | moderate | | | extreme | | |
| | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| **A. VOCAL TRACT FEATURES** | | | | | | | | | |
| 1. Labial | | | Lip rounding | | | | | | |
| | | | Lip spreading | | | | | | |
| | | | Labiodentalization | | | | | | |
| | | | Extensive range | | | | | | |
| | | | Minimised range | | | | | | |
| 2. Mandibular | | | Close jaw | | | | | | |
| | | | Open jaw | | | | | | |
| | | | Protruded jaw | | | | | | |
| | | | Extensive range | | | | | | |
| | | | Minimised range | | | | | | |
| 3. Lingual tip/blade | | | Advanced tip/blade | | | | | | |
| | | | Retracted tip/blade | | | | | | |
| 4. Lingual body | | | Fronted tongue body | | | | | | |
| | | | Backed tongue body | | | | | | |
| | | | Raised tongue body | | | | | | |
| | | | Lowered tongue body | | | | | | |
| | | | Extensive range | | | | | | |
| | | | Minimised range | | | | | | |
| 5. Pharyngeal | | | Pharyngeal constriction | | | | | | |
| | | | Pharyngeal expansion | | | | | | |
| 6. Velopharyngeal | | | Audible nasal escape | | | | | | |
| | | | Nasal | | | | | | |
| | | | Denasal | | | | | | |
| 7. Larynx height | | | Raised larynx | | | | | | |
| | | | Lowered larynx | | | | | | |
| **B. OVERALL MUSCULAR TENSION** | | | | | | | | | |
| 8. Vocal tract tension | | | Tense vocal tract | | | | | | |
| | | | Lax vocal tract | | | | | | |
| 9. Laryngeal tension | | | Tense larynx | | | | | | |
| | | | Lax larynx | | | | | | |

**C. PHONATION FEATURES**

| | SETTING | Present | | Scalar degree | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Neutral | Non-neutral | Moderate | | | Extreme | | |
| | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| 10.Voicing type | Voice | | | | | | | | |
| | Falsetto | | | | | | | | |
| | Creak | | | | | | | | |
| | Creaky | | | | | | | | |
| 11. Laryngeal frication | Whisper | | | | | | | | |
| | Whispery | | | | | | | | |
| 12. Laryngeal irregularity | Harsh | | | | | | | | |
| | Tremor | | | | | | | | |

Figure 5 - *Prosodic, Temporal Organization and Other Features in VPA (Laver & Mackenzie Beck, 2007)*

| | | | moderate | | | extreme | | |
|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 |
| **D. PROSODIC FEATURES** | | | | | | | | |
| 13. Pitch | Mean | High | | | | | | |
| | | Low | | | | | | |
| | Range | Extensive range | | | | | | |
| | | Minimised range | | | | | | |
| | Variability | High | | | | | | |
| | | Low | | | | | | |
| 14. Loudness | Mean | High | | | | | | |
| | | Low | | | | | | |
| | Range | Extensive range | | | | | | |
| | | Minimised range | | | | | | |
| | Variability | High | | | | | | |
| | | Low | | | | | | |
| **E. TEMPORAL ORGANIZATION** | | | | | | | | |
| 15. Continuity | | Interrupted | | | | | | |
| 16. Rate | | Fast | | | | | | |
| | | Slow | | | | | | |
| **F. OTHER FEATURES** | | | | | | | | |
| 17. Respiratory support | | Adequate | | | | | | |
| | | Inadequate | | | | | | |
| 18. Diplophonia | | Absent | | | | | | |
| | | Present | | | | | | |

### 3.2.3 Acoustic analysis

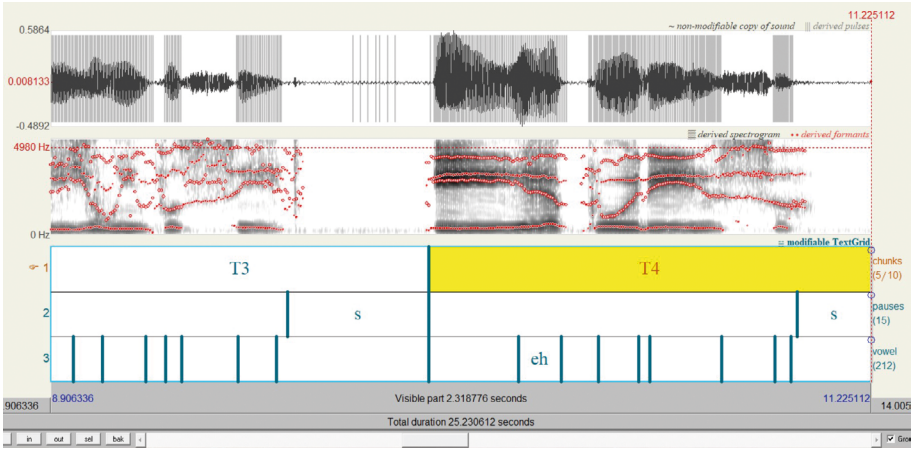
A modified version of the Prosody Descriptor Extractor script for PRAAT (Barbosa, 2021) was used for the acoustic analysis of the speech samples. The original script computes 30 prosodic parameters related to statistical descriptors $f_0$, intensity, long-term spectrum, voice quality, duration of vowels and silence. The modified version used in this study comprised 11 parameters: the difference between harmonics H1 and H2 in the vowel intervals (H1-H2 dB); Harmonic-to-Noise ratio in dB (HNR dB); Spectral emphasis in dB (emph dB); $f_0$ minimum in semitones re 1 Hz and in Hertz ($f_0$ min s.t.); $f_0$ maximum in semitones re 1 Hz and in Hertz ($f_0$ max s.t.); $f0$ median in semitones re 1 Hz and in Hertz ($f_0$ med s.t.); LTAS slope between bands 0-1000 Hz and 1000/4000 Hz (slLTASmed); $f_0$ baseline in semitones (f base s.t.); $f_0$ standard-deviation in semitones/Hertz (fSD s.t.); Local shimmer in per cent (shimmer %); Local jitter in percentual values (jitter %).

Many acoustic parameters selected here are among the most used for voice assessment and often appear in the literature investigating vocal expressivity. In addition, a few of these parameters have also been correlated in previous research to some of the traits we selected for our perceptual experiments; for example, impressions of size and dominance are often attributed to parameters related to $f_0$, as described in Ohala's Frequency Code (1984),

We chose to use the scale of semitones for most of the parameters related to $f_0$, as we expected the variation in frequency for female and male speakers to be roughly the same and to reduce skewness as they are more representative of a speaker's perceptual variability.

To run the script, a TextGrid file was created for each of the audio files, segmenting and annotating chunks of speech (Tier 1), their following silent pauses (Tier 2), and the stressed oral vowels (Tier 3), using an ASCII code in direct correspondence with the IPA symbols for Brazilian Portuguese. Figure 6 shows this process in PRAAT.

Figure 6 - *Annotation for extraction of acoustic parameters in PRAAT*



### 3.3 Reading Experiment

The main concern of our research on vocal stereotypes is to investigate the influence of vocal features on the impressions caused by the characters on the listeners. However, as no masking techniques were used in our experiments, we cannot ignore the semantic, syntactic, and vocal interacting factors potentially influencing listeners' judgments of the speech stimuli. As such, it became necessary to investigate how lexical content affects these judgments.

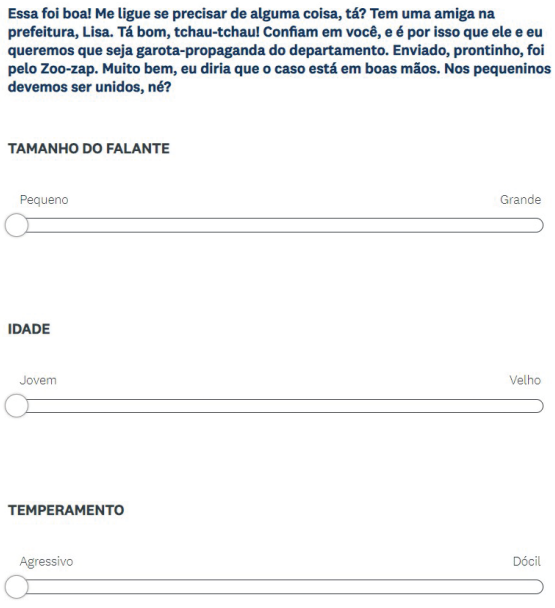
The design for the Reading Experiment was nearly identical to the Listening Experiment, with questions and instructions related to the aural aspect of the first test being removed. The experiment was also conducted online on SurveyMonkey.com.

The subjects were recruited in the same way as for the Listening Experiment. Like with the first experiment, participation was voluntary. Participants were informed of their right to withdraw at any point if they wished, and that any data collected in the experiment would be anonymised. They received no compensation for their participation. The respondents in the experiment were 73 Brazilian Portuguese-native speakers, 43 women and 30 men, aged between 19 and 58.

For this experiment, respondents were instructed to read the transcriptions of the five speech samples presented in the previous test and rate the characters in the 14 bipolar scales presented. The majority of the scales remained the same as in the first experiment, and the only changes were made to the last two scales, which were renamed from *Voz* (Voice) *to Maneira de expressão* (Manner of expression). The descriptors all remained the same.

Figure 7 shows a page of the Reading Experiment with the text sample and the same sliding scales shown in Figure 2.

Figure 7 - *Sliding scales in the Reading Experiment*



## 4. *Results*

### 4.1 Results of the interrater reliability tests in both the Listening and Reading Experiments

We used Cronbach's Alpha with both perceptual experiments to measure the reliability of the interrater agreement. With this method, values can range from 0 to 1, indicating complete disagreement and perfect agreement, respectively. Values above 0.7 are considered good, and scores above 0.9 are considered excellent. Interrater agreement scores were excellent for all characters except one in both the Listening Experiment (Table 2) and the Reading Experiment (Table 3), regardless of whether or not the respondents had seen the film. Nick Wilde's scores were slightly below excellent in some cases but still good.

Table 2 - *Interrater reliability scores for all judges in the Listening Experiment*

| *Voices* | *Male* | *Female* | *All* |
|---|---|---|---|
| *Bellwether 1* | 0.984 | 0.988 | 0.993 |
| *Bellwether 2* | 0.957 | 0.976 | 0.985 |
| *Chief Bogo* | 0.981 | 0.986 | 0.992 |
| *Judy Hopps* | 0.939 | 0.977 | 0.982 |
| *Nick Wilde* | 0.815 | 0.848 | 0.908 |

Table 3 - *Interrater reliability scores for all judges in the Reading Experiment*

| *Voices* | | | *Male* | *Female* |
|---|---|---|---|---|
| | *z-norm* | *M/F* | *S/nS* | *S/nS* |
| *Bellwether 1* | 0.987 | 0.990 | 0.980 | 0.988 |
| *Bellwether 2* | 0.934 | 0.984 | 0.973 | 0.986 |
| *Chief Bogo* | 0.951 | 0.965 | 0.916 | 0.968 |
| *Judy Hopps* | 0.955 | 0.967 | 0.873 | 0.911 |
| *Nick Wilde* | 0.940 | 0.931 | 0.890 | 0.927 |
| *All* | 0.961 | 0.976 | 0.953 | 0.965 |

### 4.2 Results of the perception tests of the Listening Experiment

Table 4 presents the mean values for the characters' perception scores. Higher values indicate an overall score closer to the descriptor listed, while lower values indicate a rating closer to the descriptor on the opposite end of the scale (See Table 1).

We have highlighted in bold the features with the strongest scores of each character, higher than 75 and lower than 25, as we consider them significant for the description of their profiles. For Age and Size, however, all scores were considered, as the values around the mid-points of those scales do not represent a neutral reading

of the character on those scales but rather an impression of the speaker as mid-sized or middle-aged.

Chief Bogo and the nice version of Bellwether usually appear on opposite ends of the scales, which is expected considering their physical appearance and demeanour in the film. The scores for the villainous Bellwether are similar to those of her façade in physical appearance but closer to Chief Bogo's in temperament and attitude.

Judy has one strong score, for Energetic, while Nick falls somewhere around the middle for most of the scales. Charming, Dominant and a Pleasant voice were the only features for which Nick received moderately high scores.

Table 4 - *Mean values for perception scores for the characters' features*

| | *B1* | *B2* | *CB* | *JH* | *NW* |
|---|---|---|---|---|---|
| *Big* | 30 | **24** | **87** | 34 | 53 |
| *Old* | **23** | **22** | 65 | 28 | 60 |
| *Aggressive* | **13** | 69 | **86** | 49 | 43 |
| *Dominant* | 44 | **88** | **91** | 69 | 66 |
| *Energetic* | **88** | **84** | **78** | **83** | 44 |
| *Tough* | **13** | 69 | **91** | 58 | 45 |
| *Charming* | **78** | 63 | 51 | 68 | 66 |
| *Kind* | **82** | 37 | 72 | 62 | 52 |
| *Bold* | 50 | 74 | 72 | 68 | 46 |
| *Happy* | **89** | 65 | 45 | 70 | 51 |
| *Honest* | 73 | 39 | 41 | 73 | 46 |
| *Sophisticated* | 45 | 64 | 50 | 59 | 55 |
| *Lively* | **93** | 74 | 48 | 78 | 42 |
| *Pleasant* | 74 | 56 | 51 | 71 | 67 |

4.3 Acoustic analysis results of the Listening Experiment

Table 5 shows the mean values of the acoustic parameters that presented the highest correlation with the perceptual descriptors. To help the reader get a feel for the $f_0$ levels for the different characters at a glance, we have added a translation of the original results extracted in semitones to Hz.

Among the highest-valued correlations were $f_0$ median and Size, $f_0$ baseline, spectral emphasis and Age, spectral emphasis and Toughness, and jitter and Honesty.

Table 5 - *Mean values for acoustic parameters*

| | *B1* | *B2* | *CB* | *JH* | *NW* |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| *emph dB* | 5.52 | 6.05 | 8.75 | 5.63 | 5.48 |
| *f0 med Hz* | 222 | 173 | 127 | 203 | 134 |
| *f0 base Hz* | 180 | 111 | 85 | 165 | 95 |
| *f med s.t.* | 83.50 | 89.25 | 83.83 | 92 | 84.75 |
| *f base s.t.* | 89.87 | 81.50 | 77 | 88.42 | 78.75 |
| *fSD s.t.* | 2.49 | 5.50 | 4.63 | 2.64 | 4.20 |
| *shimmer %* | 10.63 | 16.58 | 14.90 | 14.04 | 14.30 |
| *jitter %* | 2.40 | 3.93 | 4.08 | 2.67 | 3.76 |

For the first parameter, Chief Bogo stands out with a value about 2 dB higher than the other characters, whose values are typical for normal speech. We can look at the role Chief Bogo plays and the scenes he usually appears in for clues to explain that discrepancy. His lines usually involve him giving orders to the police officers, and in that situation, speaking louder is quite normal and expected.

Looking at the group in general, we find that spectral emphasis is positively correlated with the ratings for Aggressive (0.77), Dominant (0.67) and Tough (0.76), which were the features where Chief Bogo stood out, and negatively correlated with Charming (-0.86), Kind (-0.72) and Pleasant (-0.81), which were among Bellwether 1's strong scores.

Meanwhile, $f_0$ base value is negatively correlated with Size (-0.71), Aggressive (-0.75), Dominant (-0.73) and Tough (-0.70) and positively correlated with Charming (0.84), Kind (0.87), Happy (0.94), Honest 0.93), Lively (0.91) and Pleasant (0.81).

### 4.4 Results of the description of the VPA

Below are described the non-neutral voice quality settings and prosodic elements that characterise the characters' voice profiles in the audio samples included in the Listening Experiment.

Bellwether 1: Lip Spreading, Extensive Labial Range, Raised Larynx, Whispery Voice, High Mean Pitch, Extensive Pitch Range, High Pitch Variability, High Mean Loudness, Fast Speaking Rate.

Bellwether 2: Extensive Labial Range, Extensive Mandibular Range, Raised Larynx, Tense Vocal Tract, Whispery Voice, Harsh Voice, High Mean Pitch, Extensive Pitch Range.

Chief Bogo: Lip Rounding, Lowered Larynx, Tense Vocal Tract, Creaky Voice, Low Mean Pitch, Extensive Pitch Range, High Pitch Variability, High Mean Loudness.

Judy Hopps: Extensive Labial Range, Advanced Lingual Tip/Blade, Fronted Tongue Body, Raised Larynx, Whispery Voice, High Mean Pitch, Extensive Pitch Range, High Pitch Variability, High Loudness Variability, Fast Speaking Rate.

Nick Wilde: Lax Vocal Tract, Whispery Voice, Extensive Pitch Range, High Pitch Variability, Slow Speaking Rate.

### 4.5 Results of the Reading Experiment

Overall, the results from the Reading Experiment were quite similar to the results from the Listening Experiment. Table 6 shows a comparison between the scores for the Listening Experiment (**L**) and the scores from respondents of the Reading Experiment who had not seen the film (**RnS**), who, presumably, had no other contact with the characters from the film beyond the text samples presented to them in the Experiment.

Table 6 - *Comparison of perception scores between Listeners and Readers who did not see Zootopia*

| *Scales* | *B1* | | *B2* | | *CB* | | *JH* | | *NW* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *L* | *RnS* | *L* | *RnS* | *L* | *RnS* | *L* | *RnS* | *L* | *RnS* |
| *Big* | 30 | 23 | 24 | 27 | 87 | 56 | 34 | 38 | 53 | 47 |
| *Old* | 23 | 33 | 23 | 36 | 65 | 54 | 28 | 38 | 60 | 31 |
| *Aggressive* | 13 | 19 | 69 | 68 | 86 | 74 | 49 | 55 | 43 | 68 |
| *Dominant* | 44 | 54 | 88 | 78 | 91 | 83 | 69 | 68 | 66 | 80 |
| *Energetic* | 88 | 81 | 84 | 84 | 78 | 66 | 83 | 73 | 44 | 64 |
| *Tough* | 13 | 23 | 69 | 76 | 91 | 80 | 58 | 53 | 45 | 71 |
| *Charming* | 78 | 65 | 63 | 60 | 51 | 36 | 68 | 62 | 66 | 55 |
| *Kind* | 82 | 78 | 37 | 48 | 72 | 40 | 62 | 62 | 52 | 41 |
| *Bold* | 50 | 38 | 74 | 71 | 72 | 62 | 68 | 50 | 46 | 67 |
| *Happy* | 89 | 82 | 65 | 64 | 45 | 54 | 70 | 59 | 51 | 62 |
| *Honest* | 73 | 72 | 39 | 58 | 41 | 46 | 73 | 70 | 46 | 44 |
| *Sophisticated* | 45 | 47 | 64 | 54 | 50 | 45 | 59 | 60 | 55 | 45 |
| *Lively* | 93 | 83 | 74 | 70 | 48 | 56 | 78 | 58 | 58 | 65 |
| *Pleasant* | 74 | 79 | 56 | 55 | 51 | 33 | 71 | 60 | 67 | 42 |
| *Correlation* | *0.966* | | *0.921* | | *0.767* | | *0.893* | | *-0.215* | |

The correlations show that even in the absence of their voices and without previous knowledge of the characters and the narrative, the reading profiles are similar to the listening profiles, indicating that the textual information is enough to give the audience an idea of what these characters are like. However, the slightly stronger scores for most of the descriptors in the Listening Experiment indicate that vocal performance does help intensify the impressions given by the characters' lines, especially for Chief Bogo.

## 5. *Discussion*

Overall, the results of the Listening and the Reading Experiments showed congruence. However, the results have not been enough to explain in detail how the lexical content may have been a deciding factor in the perception of these characters, as only one of the samples made an explicit reference to a feature being investigated in the study. We hope to gain more insight into the influence of the lexical content by expanding the contexts of analysis.

The results of the Listening Experiment attest to the relevant role of voice characteristics in the attribution of physical and personality characteristics by the listeners.

Chief Bogo is a big character, Judy and Bellwether are small, and Nick is midsize. The scores obtained are according to their physical types. For the descriptor Big, Chief Bogo obtained a high score (87), and Nick an intermediate score (53). For the same scale, Bellwether 1, Bellwether 2 and Judy got, respectively, the following scores (30), (24) and (34). These results show the strong explanatory power of the Frequency Code. Looking at the acoustic parameters, we find a Low base $f_0$ value (85Hz) for Chief Bogo's voice, Mid $f_0$ values for Nick (95Hz) and high $f_0$ values for Bellwether 1 (180 Hz), Bellwether 2 (111 Hz) and Judy (165 Hz). Furthermore, Bogo's low $f_0$ voice was considered Aggressive and Tough, whereas Bellwether's High $f_0$ voice was considered Docile. Low Mean Pitch might also have influenced judgments of Charm. Chief Bogo, whose voice profile was characterised by Low Mean Pitch, Lowered Larynx and Lip Rounding voice quality settings, got the lowest score regarding Charm. Lowered Larynx and Lip Rounding expand the vocal tract and the acoustic result is lowered $f_0$.

Judgments of Dominance and Boldness were not related to high or low $f_0$ values, but the ones regarding Happiness, Pleasantness, and Liveliness were. The voices characterised by higher $f_0$ values were considered happier, more pleasant, and more lively. Bellwether 1 got the highest score for Happiness. This judgment might have been influenced by the Lip Spreading voice quality setting that the listeners frequently perceive as a smile (Emond, Rilliard & Trouvain, 2016).

Judgments regarding the characters' ages were also made according to their age profiles. Results show that the listeners of the perception tests could identify age differences. From older to younger, we find Bogo, Nick, Judy, and Bellwether 1 and 2.

Even in Nick's case, where none of his perception ratings met the threshold for strong scores in either the Listening Experiment or the Reading Experiment, voice features may have provided cues for personality judgements as he is portrayed in the film as a hard-to-read character. His vocal profile included Lax Vocal Tract, combined with Whispery Voice and Slow Speaking Rate. These characteristics might have influenced the low scores he obtained in judgments concerning Liveliness, Energy, Aggressiveness, Toughness and Boldness. Nick was the only character with a Slow Speaking Rate.

Judgments on Sophistication did not vary as much from character to character and from listening to reading. Associations between vocal profiles or lexical

characteristics and judgments of Honesty and Kindness are not clear-cut and may require investigation of contextual presuppositions.

## 6. *Conclusions*

The results of both experiments were compared and found to be congruent. The Listening Experiment performed better than the Reading Experiment in terms of revealing the characters' physical and psychological features. The vocal performances matched the characters' personality features displayed in the film and, in most cases, amplified them.

The $f_0$ baseline and spectral emphasis measures were helpful in considering the higher intensity of characteristics attributed to Size, Age, and attitudinal features such as Aggressiveness and Toughness.

Regarding the acoustic analysis of the material, some measurements aligned with the impressionistic patterns discussed and evidenced previously in the literature on sound symbolism (Hinton, Nichols & Ohala, 1995). At the same time, other correlations found in the study are not as efficiently explained by findings in previous studies and require further inspection, like whispery voices and higher scores for Charm (Gussenvhoven, 2004).

Some refinement concerning the analysis of the acoustic measurements may be necessary to eliminate measurements that do not offer any significant clues regarding the correlations between acoustic parameters and the physical and psychological features investigated here.

We believe that applying the experiments to data from different languages would also offer some insight into these issues and help formulate solutions to them. The Listening Experiment in Swedish is already underway. As correlations between vocal characteristics of a speaker and impressions of physical, psychological, and social features may vary depending on language and culture, we foresee an expansion of our research project by integrating researchers exploring other languages.

## *Acknowledgements*

The authors wish to thank Plinio A. Barbosa for his contributions to the study. The first author thanks the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* – Brasil (CAPES) – Finance Code 001 [grant #88887.630904/2021-00, grant #88881.689915/2022-01]. The second author acknowledges a grant from The Swedish Foundation for International Cooperation in Research and Higher Education (STINT) (IB2015-6488). The third author thanks PIPEq-PUCSP [grant #31279, 2024]

## *References*


Aronovitch, C.D. (1976). The voice of personality: stereotyped judgments and their relation to voice quality and sex of speaker. In *The Journal of Social Psychology*, 99, 207–220. https://doi.org/10.1080/00224545.1976.9924774

Bosseaux, C. (2015). Dubbing. In Bosseaux, C. *Dubbing, Film and Performance: Uncanny Encounters*. Oxford: Peter Lang, 55–84. https://doi.org/10.3726/978-3-0353-0737-5

Bosseaux, C. (2018). Investigating dubbing: Learning from the past, looking to the future. In, Pérez-González, L. (Ed.), *The Routledge Handbook of Audiovisual Translation.* New York: Routledge, 48–63. https://doi.org/10.4324/9781315717166

Bosseaux, C. (2019). Voice in French dubbing: the case of Julianne Moore. In *Perspectives,* 27(2)*,* 218–234. https://doi.org/10.1080/0907676X.2018.1452275

Brisola, E. (2024). *Um estudo experimental sobre o estilo de fala em dublagem.* Master's thesis. Pontifical Catholic University of São Paulo.

Campos Plaza, N. (Eds.), *Interdisciplinarity in translation studies. Theoretical models, creative approaches and applied methods*. Bern: Peter Lang, 259–276. https://10.3726/978-3-0351-0954-2

Chaume, F. (2014). *Audiovisual translation: dubbing*. New York: Routledge.

Chaume, F., Ranzato, I. & Zanotti, S. (2018). The challenges and opportunities of audiovisual translation: An interview with Frederic Chaume. In *CULTUS,* 2018(11)*,* 10-17.

Crochiquia, A. (2020). *A voz na construção de personagens em um desenho animado.* Master's thesis. Pontifical Catholic University of São Paulo.

Crochiquia, A., Eriksson, A., Fontes, M.A.S. & Madureira, S. (2020). A phonetic study of Zootopia characters' voices in Brazilian Portuguese dubbing: the role of stereotypes. In *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, 36(3), 1-46. https://doi.org/10.1590/1678-460X2020360311

Crochiquia, A., Eriksson, A., Barbosa, P.A. & Madureira, S. (2022). A perceptual and acoustic study of dubbed voices in an animated film. In Frota, S., Cruz, M. & Vigário, M. (Eds.), *Proceedings of the 11th International Conference on Speech Prosody*, Lisbon, Portugal, 23-26 May 2022, 565-569. https://doi.org/10.21437/SpeechProsody.2022-115

Crochiquia, A., Eriksson, A., Madureira, S. & Barbosa, P.A. (2023). Animated film character profile: the roles of voice and lexical content. In Skarnitzl, R., Volín, J. (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*, Prague, Czech Republic, 7-11 August 2023, 466-47. https://guarant.cz/icphs2023/862.pdf

Danan, M. (1991). Dubbing as an Expression of Nationalism. In *Meta,* 36(4)*,* 606–614. https://doi.org/10.7202/002446ar

Díaz-Cintas, J. (1990). Dubbing or subtitling: The eternal dilemma (film, translation techniques). In *Perspectives,* 7*,* 31–40. https://doi.org/10.1080/0907676X.1999.9961346

Díaz-Cintas, J. (2015). Preface. In Ranzato, I. (Ed.), *Translating Culture Specific References on Television: The Case of Dubbing*. New York: Routledge.

Díaz-Cintas, J. (2019). Film censorship in Franco's Spain: the transforming power of dubbing. In *Perspectives,* 27*,* 182-200. https://doi.org/10.1080/0907676X.2017.1420669

Emond, C., Rilliard, A. & Trouvain, J. (2016). Perception of smiling in different modalities by native vs. non-native speakers. In *Proceedings of the 8th International Conference on*

*Speech Prosody*, Boston, USA, 31 May / 1-3 June 2016, 639-643. https://doi.org/10.21437/SpeechProsody.2016-131

Fasoli, F., Mazzurega, M. & Sulpizio, S. (2017). When Characters Impact on Dubbing: The Role of Sexual Stereotypes on Voice Actor/Actress' Preferences. In *Media Psychology*, 20, 450-476. https://doi.org/10.1080/15213269.2016.1202840

Gussenhoven, C. (2016). Foundations of Intonational Meaning: Anatomical and Physiological Factors. In *Topics in Cognitive Science,* 8*,* 425-434. https://doi.org/10.1111/tops.12197

Hinton, L., Nichols, J. & Ohala, J. (1995). Introduction: Sound-symbolic processes. In Hinton, L., Nichols, J. & Ohala, J. (Eds.), *Sound Symbolism*. Cambridge: Cambridge University Press, 1–12. https://doi.org/10.1017/CBO9780511751806.001

Imhof, M. (2010). Listening to Voices and Judging People. In *International Journal of Listening*, 24, 19–33. https://doi.org/10.1080/10904010903466295

Kreiman, J., Sidtis, D. (2011). Voice, Emotion and Personality. In Kreiman, J., Sidtis, D., *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Oxford: Wiley-Blackwell, 302–360.

Laver, J. (1980). *The phonetic description of voice quality.* Cambridge: Cambridge University Press.

Liu, W., Zhang, X. & Liang, C. (2022). An acoustic study on character voices of dominators and subordinates: A case study on male characters in Empresses in the Palace. In *Frontiers in Communication,* 7. https://doi.org/10.3389/fcomm.2022.1088170

Mahrholz, G, Belin, P. & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. In *PLoS One*, 13, e0204991. https://doi.org/10.1371/journal.pone.0204991

Nybakk, L.M. (2014). *Representations of linguistic variation in audiovisual translation: A study of American animated films and their Norwegian dubbed translations.* Master's thesis. Norwegian University of Technology and Science.

Ohala, J.J. (1984). An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice. In *Phonetica,* 41(1)*,* 1–16. https://doi.org/10.1159/000261706

Pedersen, J. (2010). Audiovisual translation – in general and in Scandinavia. In *Perspectives,* 18(1)*,* 1–22. https://doi.org/10.1080/09076760903442423

Sánchez-Mompeán, S. (2016). "It's not what they said; it's how they said it": A corpus-based study on the translation of intonation for dubbing. In Rojo-López, A.M., Sánchez-Mompeán, S. (2020). Introduction: Unhiding the Art. In Sánchez-Mompeán, S. *The Prosody of Dubbed Speech: Beyond the Character's Words*. New York: Palgrave Macmillan, 1-10. https://doi.org/10.1007/978-3-030-35521-0_1

San Segundo, E., Mompean, J.A. (2017). A Simplified Vocal Profile Analysis Protocol for the Assessment of Voice Quality and Speaker Similarity. In *Journal of Voice,* 31(5)*,* 644. e11–644.e27. https://doi.org/10.1016/j.jvoice.2017.01.005

Tveit, J.E. (2009). Dubbing versus Subtitling: Old Battleground Revisited. In Díaz-Cintas, J., Anderman, G. (Eds.), *Audiovisual Translation: Language Transfer on Screen*. New York: Palgrave Macmillan, 85–96. https://doi.org/10.1057/9780230234581_7

Weirich, M. (2008). Vocal Stereotypes. In *Proceedings of 2nd Tutorial and Research Workshop on Experimental Linguistics*. Athens, Greece, 25-27 August 2008, 229–232. https://www.isca-archive.org/exling_2008/weirich08_exling.pdf

Xu, A., Lee, A. (2018). Perception of vocal attractiveness by Mandarin native listeners. In Klessa, K., Bachan, J., Wagner, A., Karpiński, M. & Śledziński, D. (Eds.), *Proceedings of the 9th International Conference on Speech Prosody*. Poznan, Poland, 13-16 June 2018, 344–348. https://doi.org/10.21437/SpeechProsody.2018-70

Whitman-Linsen, C. (1992). *Through the Dubbing Glass: The Synchronization of American Motion Pictures into German, French and Spanish*. Bern: Peter Lang.