http://dx.doi.org/10.1590/1678-460X2020360311

A phonetic study of Zootopia characters' voices in Brazilian Portuguese dubbing: the role of stereotypes

Um estudo fonético das vozes de personagens do filme Zootopia na dublagem em português brasileiro: o papel dos estereótipos

Alice Crochiquia[1]
Anders Eriksson[2]
Mario A. S. Fontes[3]
Sandra Madureira[4]

RESUMO

Este trabalho segue uma abordagem experimental de investigação da fala expressiva que integra procedimentos metodológicos de análises perceptiva e acústica. Como objeto deste trabalho, focalizamos a qualidade de voz e a dinâmica vocal de personagens com distintas personalidades. Amostras de fala de quatro principais personagens do filme de animação intitulado "Zootopia" dublados por atores brasileiros são analisadas.

1. Pontifícia Universidade Católica da São Paulo, São Paulo – Brasil. https://orcid.org/0000-0002-2344-5390. E-mail: liaac @pucsp.br.
2. Stockholm University, Stockholm – Sweden. https://orcid.org/0000-0002-6844-4834. Email: anders.eriksson@ling.su.se.
3. Pontifícia Universidade Católica da São Paulo, São Paulo – Brasil. https://orcid.org/0000-0001-7418-8470. E-mail: fontes@pucsp.br.
4. Pontifícia Universidade Católica da São Paulo, São Paulo – Brasil. https://orcid.org/0000-0001-8263-053X. E-mail: madusali@pucsp.br.

Devido à função expressiva da qualidade de voz, a seguinte pergunta de pesquisa foi aventada: Que tipos de ajustes de qualidade de voz e de dinâmica vocal foram usados pelos atores na dublagem de "Zootopia" para compor os perfis vocais dos personagens? A avaliação perceptiva de 54 estímulos de fala foi realizada com o Sistema de Avaliação Vocal (Laver & Mackenzie Beck, 2007). Medidas acústicas foram automaticamente extraídas, aplicando o ExpressionEvaluator script (Barbosa, 2008) no PRAAT. Os perfis de cada um dos quatro personagens foram compostos com base nas características psicológicas descritas no roteiro do filme. Os resultados da análise acústica, da análise perceptiva dos ajustes de qualidade vocal e de dinâmica vocal foram correlacionados por meio do método MFA (Análise Multifatorial) no ambiente R. Foram utilizadas 40 variáveis cuja análise resultou na distribuição dos estímulos em 6 clusters. As variáveis quantitativas que apresentaram a maior percentagem de correlação foram as relacionadas à frequência fundamental e à inclinação espectral: Standard Deviation of f0 Derivative, Standard Deviation of Spectral Tilt e f0 Median. As variáveis qualitativas com maior grau de correlação foram: Lowered Larynx, Lip Rounding, Breathy Voice e Minimised Pitch Range. A pesquisa apresentou evidências a favor do uso simbólico da matéria fônica e contribuições para o entendimento de como os estereótipos são estabelecidos.

Palavras-chave: *simbolismo sonoro; fonética acústica; qualidade de voz; estereótipos vocais; dublagem.*

ABSTRACT

This work comprises an experimental investigation approach of expressive speech that integrates methodological procedures of perceptual and acoustic analyses. As the object of this work, we have focused on voice quality and vocal dynamics. Speech samples from the four main personality-distinct characters in the animated feature film "Zootopia" dubbed by Brazilian voice actors have been analysed. Due to the expressive function of voice quality, we have posed the following question: what types of voice quality and vocal dynamics settings were used by the voice actors in the Brazilian dubbing of "Zootopia" to compose the vocal profiles of the characters? Perceptual evaluation of the 54 speech stimuli was performed using the Vocal Profile Analysis protocol (Laver & Mackenzie Beck, 2007). Acoustic measures were automatically extracted using the ExpressionEvaluator script (Barbosa, 2008) for PRAAT. The profiles for each of the four characters were composed based on the psychological traits described in the film script. The results of the acoustic analysis, the

DELTA
36.3
2020

perceptual analysis of voice quality and vocal dynamics settings were correlated using the MFA (Multiple Factor Analysis) method in the R environment based on 40 variables (quantitative and qualitative) and it turned out that the speech stimuli were distributed in 6 clusters according to the variables analysed. The quantitative variables that presented the highest correlation percentage were: Standard Deviation of f0 Derivative, Standard Deviation of Spectral Tilt, f0 Median. The qualitative variables that presented the highest correlation percentage were: Lowered Larynx, Lip Rounding, Breathy Voice and Minimised Pitch Range. The research has presented evidence in favor of the symbolic use of phonic matter and contributions to the understanding of how vocal stereotypes are established.

Keywords: *sound symbolism; acoustic phonetics; voice quality; vocal stereotypes; voice acting*

1. Introduction

The potential of speech sounds to express a variety linguistic, extralinguistic and paralinguistic meanings is what makes it such an effective means of communication. Expressivity in speech is built on the relationship between segmental elements and prosodic elements. It is the interaction of those speech structures that yields multiple meanings to the same sequence of segmental elements (Madureira, 2004, 2011, 2018).

As such, the investigation of vocal expressivity must consider speech in all its dimensions, considering the phonetic aspects which characterize the speakers´ vocal performances as animators (Goffman,1981) and impress listeners who attribute meaning effects based on vocal profiles. Therefore, vocal profiles are important to be investigated since they are powerful indices of physical and psychological states (Gobl and Chasaide, 2003).

This study aims to investigate and characterize the vocal profiles of the four main characters of the animated film "Zootopia" in its Brazilian Portuguese dubbing. The film is an animation featuring anthropomorphic animals living in a metropolis populated by mammals, called Zootopia. The film's narrative centers around the issue of

prejudice and assumptions based on stereotypes, and how harmful they can be. The "modern fable" aspect of the film and the mixture of physical and biological characteristics of animals with human traits and attitudes allows us to correlate our data with patterns described in the sound symbolism codes and examine the role of vocal stereotypes in the characters' vocal profiles.

This study takes an experimental approach to the investigation of vocal expressivity that integrates methodological procedures of perceptual and acoustic phonetic analysis, and multivariate analysis. It seeks to answer the following questions: what types of voice quality and vocal dynamics settings were used by the voice actors in the Brazilian dubbing of "Zootopia" to compose the vocal profiles of the characters? Do they reflect stereotypes understood as sound symbolic coded?

Sound symbolism emerges from the direct relations holding between the physical properties of sounds and meanings as stressed in works considering the associations between phonetic characteristics and linguistic, paralinguistic and extralinguistic meanings such as the ones by Köhler (1947), Jakobson (1978, 1979), Albano (1988), Tsur (1992), Hinton, Nichols and Ohala (1994), Abelin (1999), and Fónagy (1983, 2001), to mention a few addressing this complex issue. The fact that sounds have meaning effects can be explored in the study of vocal stereotypes, considering that both biological and cultural specific constraints might determine them, as pointed out by Kreiman and Siditis (2011) based on evidence from on the perception of personality from voice.

2. Theoretical Framework

The basic frameworks for the analysis of voice quality and vocal dynamics and findings of previous studies on the links between voice and physical, symbolic and psychological features are considered in this section.

DELTA
36.3
2020

2.1. Voice quality and vocal dynamics

The phonetic descriptive model of voice quality proposed by Laver (1980) covers both phonatory (laryngeal) and articulatory (supralaryngeal) elements of the vocal tract. The model considers the possible configurations of all those elements that define an individual's speech, and describes them in articulatory, physiological, acoustic and auditory terms.

This phonetic model for assessing voice quality considers the inherent anatomical features of the human phonatory system, as well as extrinsic factors, such as long-term articulatory and phonatory settings, polysegmental, recurrent muscle mobilizations that occur in the vocal tract during speech. These recurrent and long-lasting extrinsic factors during the production of the speech segments constitute the analytical unit of this model, the setting.

Laver presents two principles in this descriptive model of voice quality: compatibility and susceptibility. With the first principle, Laver establishes that certain settings are incompatible and cannot occur simultaneously during the production of a segment. For example, the lip spreading setting is incompatible with the lip rounding setting. This principle also rules the relationship between settings and the individual anatomy of the speaker: the individual anatomical configuration of a speaker determines the degree of ease in adopting a specific setting in their speech.

The principle of susceptibility establishes that segments are more vulnerable to certain settings, and thus, such settings are more easily perceived in those segments. In general, phonic segments are more susceptible to the influence of settings with which they do not share articulatory, acoustic or auditory features. For example, oral vowels are more vulnerable to the influence of nasal settings than nasal vowels.

From these theoretical models, the Vocal Profile Analysis protocol (VPA) was developed (Laver, 1981; Laver, 2000; Laver & Mackenzie Beck, 2007). The protocol considers phonatory (laryngeal), articulatory (supralaryngeal) and tension settings, as well as vocal dynamic features, such as prosodic and temporal elements.

The VPA model and script are based on the neutral setting, a set of settings that occurs simultaneously in the vocal tract. This setting is characterized by none disturbances at any point in the extension of the vocal tract due to the actions of the lips, jaw, tongue, pharynx or larynx (Laver, 2000).

The remaining voice quality settings are described in relation to the neutral setting. The protocol consists of two stages of analysis: 1. the identification of settings distinct from the neutral setting in the subject's voice; 2. the attribution of values to these non-neutral settings, ranging from 1 to 6; the higher the degree, the greater the difference in relation to the neutral setting. The combination of these settings and their degrees constitute the vocal profile of a subject (Laver, 1980).

The intrinsic and extrinsic settings exhibited by a speaker convey not only information about the linguistic message, but also information about the speaker, which can be divided into three categories (Laver & Trudgill, 1979): social characteristics markers, such as social class, level of education, place of origin; Markers of physical features such as sex, body type, age; attitude and affective state markers

The first category considers that an individual who speaks a certain language and belongs to a speech community adopts certain voice quality settings that relate to the sounds in their language and dialect (Honikman, 1964), and that these particular vocal tract configurations can be recognized by the listeners.

The second category is linked to intrinsic factors of voice quality, since they determine vocal characteristics resulting from anatomical attributes shared by groups of speakers, like shorter vocal tracts on children.

The latter category is related to voice quality settings resulting from emotional discharges and their consequences on the physiological configuration of the vocal tract, as well long-term attitudes and personality traits signaled in a speaker's voice.

DELTA
36.3
2020

2.2. Sound symbolism codes

Four kinds of sound symbolism codes have been proposed in the phonetic literature and they establish connections between the physical properties of sounds and meaning expression, the basis of speech expressivity (Madureira, Fontes and Camargo, 2019). The Frequency Code (Ohala, 1980) is linked to fundamental frequency (f0), an acoustic parameter related, in terms of speech production, to the number of vibrations per second of the vocal folds and, in perceptual terms, to the auditory sensation that goes from low to high (pitch). The Frequency Code is based on the evolution of vocalizations in animals, as a survival instinct of the species: low f0 values (few vibrations of the vocal folds per second) are linked to larger animals, and signal power, force hostility, and aggressiveness. In contrast, high f0 values (many vocal fold vibrations per second) are associated with smaller animals and signal submission, fragility.

In the animal kingdom, the use of frequency variations, combined with other forms of manipulating how one's physical size is perceived are decisive in situations of confrontation, especially in situations where the intention is exactly to avoid direct confrontation. In certain aspects of human vocalizations, the variation in f0, associated with other visual and gestural communication, attitudes and intentions of the speaker. Thus, the Frequency Code reveals the direct relationships that can exist between sound and meaning.

The Effort Code concerns articulation; the higher the degree of articulatory effort, the more precise the articulation, which carries some potential meanings. A greater degree of articulatory effort is related to tension, determination, while a lesser degree of effort is linked to disinterest, relaxation.

The Production Code, or Respiratory Code (Gussenhoven, 2004) is related to subglottic air pressure. At the beginning of a sentence, subglottic air pressure rises, and decreases at the end. Thus, it also generates potential meaning effects. Low subglottic air pressure in speech can be associated to weakness, while a high subglottic pressure can give the impression of arousal.

The Sirenic Code(Gussenhoven, 2016) is related to the voice characterized by air escaping between the vocal folds (breathy voice). This type of voice quality has a linguistic (interrogative marker) function, and refers to paralinguistic (low excitement and seduction) and extralinguistic (female sensuality and fragility) elements in speech. Its name is linked to these last two elements, a reference to the mythical tales of sirens, who attract sailors (usually to their deaths) with their voice and singing.

2.3. Vocal cues to personality and vocal stereotypes

Sapir (1927) wrote an article about the expressive value of speech and refers to speech as a personality trait. He opened the path to the formal studies on the relationship between voice and personality which were developed in the 1930s. These early studies follow the first of the three main approaches taken in studies of personality markers in the voice: accuracy studies, which were concerned with how accurate listeners were when judging a speaker's personality from their voice.

The first of such works the study by Pear (1931), conducted in England. The research consisted in the transmission of a reading excerpt from a text by Charles Dickens, produced by 9 different speakers, with different professions. After this broadcast, the listeners sent the researchers a form published in the newspaper with their opinions on the personalities and professions of the speakers.

The study had with no specific parameters adopted for the recordings and the selection of the judges, however, an important finding of the work was the existence of vocal stereotypes, shared by the judges who responded to the survey. For some of the voices, the percentage of agreement on the speaker's profession was higher than the percentage of judges who answered the question correctly (Kreiman & Sidtis, 2011).

Other formal studies with more robust experimental designs were carried out during the 1930s and 1940s, such as the study by Allport and Contril (1934), in which listeners judged characteristics such as physical appearance, professions, values, political preferences of three speakers. The results of these studies consistently replicated the

DELTA
36.3
2020

conclusions of the Pear's study. Teshigawara (2003) and Kreiman & Sidtis (2011) report that the accuracy of the findings in the research carried in the 1930s and 1940s may be affected by the inherent experimental limitation of the studies.

The studies carried out after the 1940s, for the most part, can be divided into two other approaches: studies of externalization and studies of attribution (or inference). The first approach refers to studies that aim to explore the correlation between acoustic analysis, expert analysis and coding of the voices of speakers and personality tests based on self-assessment. Studies that took this approach also found it difficult to produce solid conclusions from these correlations due to the inadequate nature of the personality assessment parameters, and the inaccuracy of the acoustic measures of the speech samples used. In addition, Scherer (1979) points out how cultural differences can make the comparison of studies in different languages difficult. The 1970s were indeed productive in terms of research works focusing the roles of voice qualities in determining psychological and social states (Scherer, 1972, 1979, 1989; Scherer, Uno and Rosenthal,1972; Scherer, Rosenthal and Koivumaki, 1972; Giles, Scherer and Taylor 1973; Scherer, London and Wolf , 1973).

The latter approach refers to the investigation of the attribution of characteristics to speakers by layman judges, without concern for the accuracy of such judgments. This approach has characterized most of the studies on voice and personality in recent decades. Within this approach are studies on the "vocal attractiveness" stereotype.

Just as physically attractive individuals are perceived as more confident, competent and sympathetic, individuals with voices assessed by judges as "attractive" are also perceived in this way (Berry, 1991, 1992; Zuckerman & Driver, 1989). According to Zuckerman, Hodgins and Miyake (1990), this stereotype, however, is prominent when the judges do not know the speaker, and apply less and less as the listener knows and becomes familiar with the speaker. This study also demonstrated that the vocal stereotype proved to be more influential than the physical stereotype in highly familiar relationships.

Some studies, such as that of Hecht and LaFrance (1995), have combined the externalization and attribution approaches. In this study,

the authors investigated whether the vocal characteristics of telephone attendants and the impressions caused by their voices correlated with the speed with which these professionals served customers. The researchers asked judges to hear statements from selected attendants and to classify personality traits and vocal characteristics. The traits were grouped into a single factor called a positive attitude; Correlations were calculated between vocal characteristics and positive attitude. The vocal characteristics that showed significant correlations with the positive attitude were described as "modulated" and "clear". This first characteristic can be understood as a high variability of pitch and intensity in relation to the duration of the speech segments. The second characteristic, in turn, may indicate a high variability of articulatory movement in speech.

The study by Yarmey (1993) combined investigations of vocal and facial gestures. In the study, judges should classify the vocal characteristics of 15 subjects in three situations: stimuli presenting only the subjects 'faces (in this situation, the judges should imagine the vocal characteristics of the subjects), stimuli presenting only the subjects' voice, and stimuli combining the visual and sound information. The results of the study suggest that the configurations of voices for non-criminal subjects are more typical and more pleasant, while the voices of criminals have unique and less pleasant characteristics, and are perceived as monotonous, rigid and unclear.

In his study of vocal expressiveness in Japanese animation (anime), Teshigawara (2013) examined the characters' voices using a modified version of Laver's voice quality model (2000), acoustic measures such as f0 and formants, and vocal types correlated with the categories of characters (heroes and villains). The results of these analyses were compared statistically with the perceptual experiment, which consisted of the assessment of layman judges regarding age, sex, physical characteristics, personality traits, emotional states and vocal characteristics of these same characters. The study showed that villains and heroes in anime show differences in voice quality, especially in relation to pharynx and larynx settings. Most villains presented pharyngeal expansion and laryngeal sphinctering, while the heroes exhibited breathy voice and neutral setting for the pharynx. The study also demonstrated that these settings have different impressive

DELTA
36.3
2020

effects on the listeners, with the judges of the perceptual experiment assigning unfavorable physical, personality and affective states to the voices that had non-neutral pharynx and larynx settings.

One issue with which discussions about vocal stereotypes are concerned is the origin of stereotypes; whether their empirical bases are biological or cultural. Montepare and Zebrowitz-McArthur (1987) tested their hypothesis that vocal stereotypes were based on ecological factors, in a study comparing the perception of "infantilized" voices by North American and Korean listeners. Listeners of both nationalities associated childish voices with weak, incompetent and affectionate personalities. The difference between the two groups was shown only in the association of female voices with affectionate personalities by North American listeners, but not by Korean listeners. These results indicate that the two cultures share these particular vocal stereotypes, and give credibility to the hypothesis of the biological origin of these stereotypes.

However, the field also has a number of studies that point to the social origins of vocal stereotypes, such as the study by Peng, Zebrowitz and Lee (1993), also focused on North American and Korean judges. Due to the differences in how both cultures see youth and old age, the authors hypothesized the personality traits perceived from the vocal characteristics associated with youth and old age would be different for North American and Korean subjects. The study showed that American judges (coming from a culture that values youth) considered voices that were louder and exhibited faster speaking rates (characteristics associated with young voices) as stronger and more dominant. For Korean judges, this was association was observed only in relation to intensity.

Thus, evidence from studies carried out in the area demonstrates both biological and cultural influences on the way in which the personality is perceived through the voice, different aspects of personality and the vocal stereotypes associated with them may have different origins.

Some studies on vocal attractiveness in children's voices suggest a possible additional influence on the relationship between voice and personality: the "self-fulfilling prophecy", as proposed by Scherer and

Scherer (1981). According to this theory, individuals may develop personality dispositions and behaviors based on inferences made by their significant others when interacting with them.

3. Material and methods

This section presents the study's material and the methodological procedures. The first subsection presents the material selected for the analysis and the criteria for the selection, following subsections give an overview of the methods applied in the study.

3.1. Material

As a first step in the methodological procedures, the Brazilian Portuguese audio track and the video file from the Blu-ray disc of "Zootopia" were extracted and the converted into .mp4 format for video and .wav for audio.

The files were then uploaded into ELAN. With the software, the different scenes of the film were annotated, given a brief description and the audio track was transcribed. Those tiers were then converted into .TextGrid files and uploaded to PRAAT along with the audio track.

The voice samples analysed in this study belong to the film's four main characters: Judy Hopps (bunny), Nick Wilde (fox), Chief Bogo (buffalo) and Assistant Mayor Bellwether (sheep). The voice acting for each of these characters in the Brazilian Portuguese dubbing is performed by a different actor.

The characters were profiled based on information taken from interviews with the film's creators, writers and voice actors, and promotional materials released by the studio.

The speech samples were taken from scenes in from different points of film's narrative, in which these characters interacted with at least one of the other three selected characters, and their demeanor was consistent with their psychological profiles. The samples did not have a musical score or sound effects that would interfere with the acoustic

DELTA
36.3
2020

and perceptual analyses of the voices. In total, 16 speech samples were selected, four for each character. The samples were between 7 and 14 seconds long.

For the VPA analysis and the extraction of acoustic parameters in PRAAT, the 16 speech samples were edited into smaller sections (stimuli) to eliminate longs pauses, sound and noises that were not part of the lines of the character being analysed. This generated 54 stimuli.

Chart 1 below lists the numbers of stimuli analysed for each character.

Number of Stimuli	Characters (Gender)
12	Assistant Mayor Bellwether (F)
12	Chief Bogo (M)
16	Judy Hopps (F)
14	Nick Wilde (M)

Figure 1 – Stimuli selected for analysis

The full transcription of the stimuli is included at the end of this paper.

3.2. Analysis procedures

3.2.1. Perceptual analysis

The stimuli were rated according to the VPA protocol (Laver and Mackenzie Beck). This rating was performed by a phonetician with twenty years of experience with the VPA.

The protocol consists of 55 settings, defined in relation to the neutral setting. Those settings are divided into six categories: Vocal Tract Features, Overall Muscular Tension, Phonation Features, Prosodic Features, Temporal Organization and Other Features.

The rating of voices with the protocol is a two-stage evaluation. In the First Pass, raters note if a subject's voice presents any of those 55 non-neutral settings. In the Second Pass, raters assess the degree of the non-neutral settings marked in the First Pass in a six-point grading scale; 1 to 3 represent moderate degrees, 4 to 6 represent extreme degrees.

The VPA protocol is included at the end of this paper. The ratings were done in PRAAT and annotated in .TextGrid files.

3.2.2. Acoustic analysis

For the acoustic analysis of the characters' voice, the script ExpressionEvaluator (Barbosa, 2008) for PRAAT to the 54 stimuli. The script was used to extract five classes of acoustic parameters: fundamental frequency (F0), the first derivative of the fundamental frequency (dF0), intensity, spectral tilt (SpTt) and Long-Term Average Spectrum (LTAS). One to four descriptors were used for each class producing the following parameters: F0 median; F0 inter-quartile semi-amplitude; F0 0.995 quantile; F0 skewness; F0 first derivative mean; F0 first derivative standard deviation; F0 first derivative skewness; Global intensity skewness; Spectral tilt; Spectral tilt standard deviation; Spectral tilt skewness and Long-Term Average Spectrum (LTAS) standard-deviation. About these parameters, some remarks are worth mentioning since they are not obvious choices. Spectral tilt, for example, is a known correlate to vocal effort. The first derivatives are sensitive to changes in the respective parameters. The first derivative of F0 can thus be used to detect abrupt changes in the F0 contour. Skewness finally is a way of quantifying deviation from a normal distribution.

The script operates with the option of parameter reference values for male and female speakers. Thus, the reference values of the characters' genre were set for the each of measurements. With the extracted data, the script automatically generates a .txt file and PRAAT graphics.

The values for each of the parameters extracted to the .txt file were then exported to an Excel spreadsheet along with the results of the VPA rating.

DELTA
36.3
2020

3.2.3. Multivariate analysis

Multivariate analysis allows the projection of data on different planes and the weights of measures in different dimensions. By combining main component methods, hierarchical grouping and partitioning, this method of analysis makes it possible to highlight the similarities and differences between stimuli (Husson et al., 2013).

For multivariate analysis, we used R, a programming language and free software for statistical calculations and data visualization, and within this environment, the R Commander and FactoMineR interfaces.

From the package of FactoMineR programs, specific for Exploratory Analysis of Multiple Variables, we chose the MFA (Multiple Factor Analysis) method, an extension of the PCA (Principal Component Analysis) method, used for analyses in which a set of elements is characterized by variables structured as groups, which can come from different sources of information. The method calculates the main component of each of the groups of variables, to map the similarities between the stimuli in relation to all the variables and to group them based on these similarities (clustering).

The basic methodology of this analysis consists in looking at the data from different angles to understand its complexity in a data "cloud". The variables are represented by points in a dimensional space that can be defined, provided that each of these variables has a value per stimulus. The multivariate analysis presented in this work was performed in 7 dimensions.

In this work, the variables are the VPA protocol settings identified in the 54 stimuli and acoustic parameters extracted by ExpressionEvaluator script. All values of the considered variables were normalized by z-score.

4. Results

This section presents the results of the analyses performed in the study, which are divided into three subsections. The first of these

subsections, Perceptual analysis, is divided into four subsections, each with a brief overview of the results of the perceptual analysis for the four characters.

4.1. Perceptual analysis

We begin this section presenting the results for perceptual analysis using the VPA protocol. In the table 3 below, we have listed the settings exhibited by each of the characters in their stimuli. The settings are color coded according to type of setting.

Table 1 – Voice quality and vocal dynamics settings exhibited in the characters' speech samples

Character	**VPA Settings**
Judy Hopps	Lip spreading
	Minimised labial range
	Extensive mandibular range
	Raised larynx
	Tense vocal tract
	High mean pitch
	High pitch variability
	High mean loudness
	Fast speaking rate
	Interrupted continuity
Nick Wilde	Backed tongue body
	Raised larynx
	Lowered larynx
	Lax vocal tract
	Tense larynx
	Breathy Voice
	High mean pitch
	Low mean pitch
	Minimised pitch range
	High pitch variability
	Low pitch variability
	Low mean loudness
	Fast speaking rate
	Slow speaking rate

DELTA
36.3
2020

Chief Bogo	Lip rounding Open jaw Lowered tongue body Lowered larynx Tense vocal tract Tense larynx Creaky voice Low mean pitch Minimised pitch range High pitch variability High mean loudness
Assistant Mayor Bellwether	Lowered tongue body Pharyngeal constriction Raised larynx Tense vocal tract Tense larynx Breathy voice Harsh voice High mean pitch Extensive pitch range High pitch variability High mean loudness Fast speaking rate Slow speaking rate
Vocal tract features	
Overall muscular tension	
Phonation features	
Prosodic features	
Temporal organization	

4.1.1. Judy Hopps

The character was the only one that did not exhibited any non-neutral phonatory setting in any of her stimuli.

High pitch variability was the setting that was most repeated recurrent setting in the stimuli from the character, appearing in 13 of the 16 stimuli analysed, always in a moderate degree 2.

4.1.2. Nick Wilde

Nick Wilde was the character who exhibited the highest number of non-neutral voice quality and vocal dynamics settings; 14, in total.

The breathy voice setting was the main setting identified in the character's stimuli, appearing in 12 of the 14 stimuli, always in moderate degrees.

4.1.3. Chief Bogo

The character was the only one among the four who exhibited non-neutral settings to extreme degrees.

The Lip rounding and Lowered larynx settings appear in 10 of the 12 stimuli from the character, in similar degrees. The main differences between the stimuli were the muscle tension and vocal dynamics settings.

4.1.4. Assistant Mayor Bellwether

In her first six stimuli, the character maintained a stable set of voice quality and vocal dynamics settings. In the final six stimuli, when the character shows her true manipulative personality, there was greater variation between the non-neutral settings present in the segments.

Among the settings exhibited in the first six stimuli, the breathy voice and high mean loudness were the only ones that did not appear in any of the final stimuli. On the other hand, the tense vocal tract setting appeared in most of the final stimuli and in none of the six first stimuli.

4.2. Acoustic Analysis

This section presents the results of the analysis of the stimuli's acoustic and statistical parameters, related to fundamental frequency and intensity. Those parameters are: F0 median (MED), F0 inter-quartile semi-amplitude (SPQ), F0 0.995 quantile (QAF), F0 skewness (ASF), F0 first derivative mean (MEF), F0 first derivative standard deviation (DPF), F0 first derivative skewness (ADF), Global intensity skewness (AMT), Spectral tilt (MÊS), Spectral tilt standard deviation (DES), Spectral tilt skewness (AES), and Long-Term Average Spectrum (LTAS) standard-deviation (SLT).

DELTA
36.3
2020

Figure 2 and 3, below, represent the mean values of all ExpressionEvaluator measurements for each character.

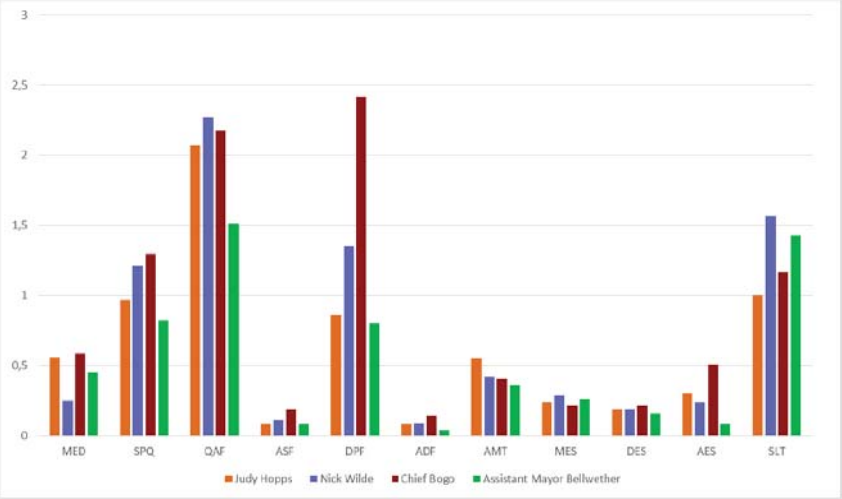

Figure 2 – Mean values for ExpressionEvaluator measurements

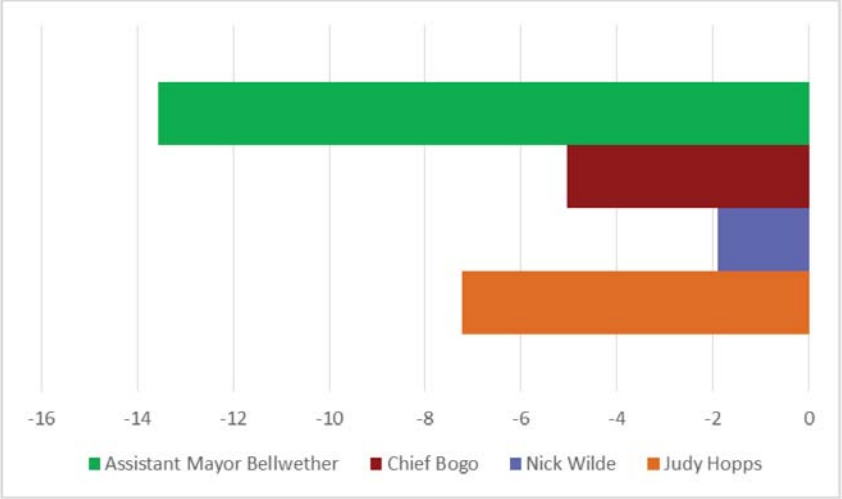

Figure 3 – Mean values for MEF measurement

Chief Bogo exhibited the highest mean values for 7 of the 12 measurements. Among them were the SPQ (F0 inter-quartile semi-amplitude), and DPF (F0 first derivative standard deviation) measurements. These measurements are variance indexes (SPQ and DPF).

The stimuli from Nick Wilde and Assistant Mayor Bellwether presented higher values for the SLT (LTAS standard-deviation) measurement than those of Chief Bogo and Judy Hopps. Higher SLT values indicate breathy voice. Nick Wilde's stimuli also presented higher values for the QAF (F0 0.995 quantile) measurement.

Regarding the AMT (Global intensity skewness), the values for this measurement were higher in the stimuli from Judy Hopps than those in the stimuli from other characters. Higher values of AMT may be linked to high pitch and laryngeal tension.

The average values for F0 were calculated in PRAAT, correcting extraction errors. The F0 averages of the characters' speech stimuli were: 239 Hz for Assistant Mayor Bellwether in the scenes where she appears to be friendly and affable, and 282 Hz in the scenes where she acts as the villain of the film; 126 Hz for Chief Bogo; 326 Hz for Judy Hopps; 145 Hz for Nick Wilde.

4.3. Multivariate Analysis

For this analysis, 40 variables were considered. Those variables are divided into two groups: the ExpressionEvaluator script measurements (quantitative data), and voice quality and vocal dynamics settings (qualitative data) identified as non-neutral in the selected stimuli. Charts 4 and 3, below, show the variables that make up each of the groups.

DELTA
36.3
2020

Chart 3 – Gc1 group descriptors

Group	Method	Type	Variables
Gc1	ExpressionEvaluator (12 parameters)	Quantitative	F0 median (MED) F0 inter-quartile semi-amplitude (SPQ) F0 0.995 quantile (QAF) F0 skewness (ASF) F0 first derivative mean (MEF) F0 first derivative standard deviation (DPF) F0 first derivative skewness (ADF) Global intensity skewness (AMT) Spectral tilt (MÊS) Spectral tilt standard deviation (DES) Spectral tilt skewness (AES) Long-Term Average Spectrum (LTAS) standard-deviation (SLT)

Chart 4 – Gq1 group descriptors

Group	Method	Type	Variables
Gq1	VPA protocol (28 settings)	Qualitative	Lip rounding Lip spreading Minimised labial range Open jaw Extensive mandibular range Backed tongue body Lowered tongue body Pharyngeal constriction Raised larynx Lowered larynx Tense vocal tract Lax vocal tract Tense larynx Lax larynx Creaky voice Breathy voice Harsh voice High mean pitch Low mean pitch Extensive pitch range Minimised pitch range High pitch variability Low pitch variability High mean loudness Low mean loudness Interrupted continuity Fast speaking rate Slow speaking rate

The analysis showed that the group of quantitative variables (Gc1) was farther projected than the group of qualitative variables (Gq1) in Dimension 1 of the vector space. In Dimension 2, the two groups of variables were equally projected. This means that these two dimensions are relevant to our analysis. The two dimensions have 21.68% (11.75% in Dim1, and 9.33% in Dim2) of explanatory power for the data. Figure 4, below, shows the projection of each of the groups of variables in the two dimensions.

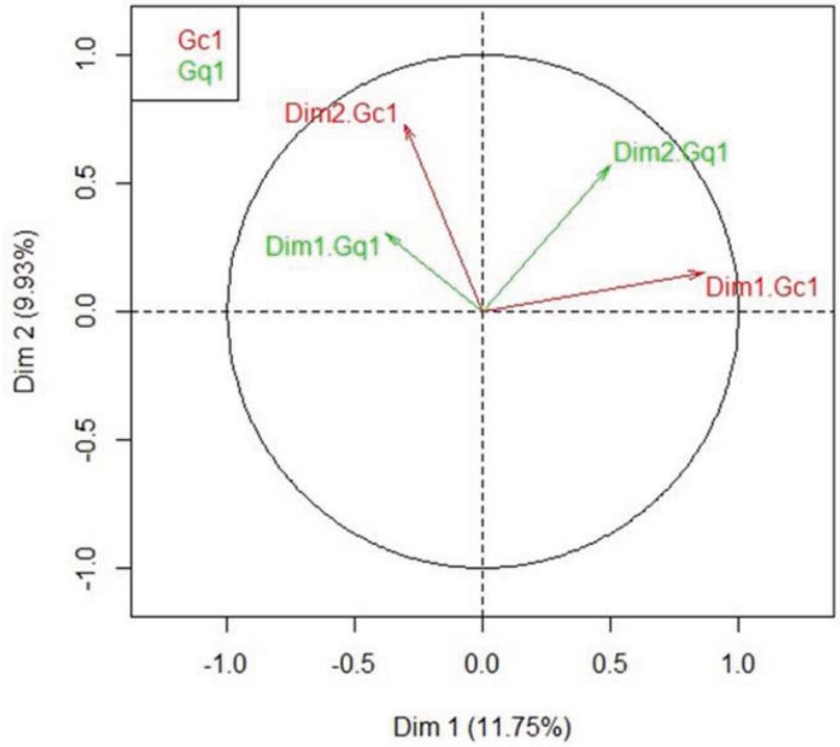

Figure 4 – Projection of Gc1 and Gq1 in Dimensions 1 and 2

The two groups of variables were relevant to differentiate the stimuli. The relevance of each of the groups can be verified from the distance between them and the zero and the two axes: the more distant, the more relevant. This distance from the groups is shown in Figure 5, below.

DELTA
36.3
2020

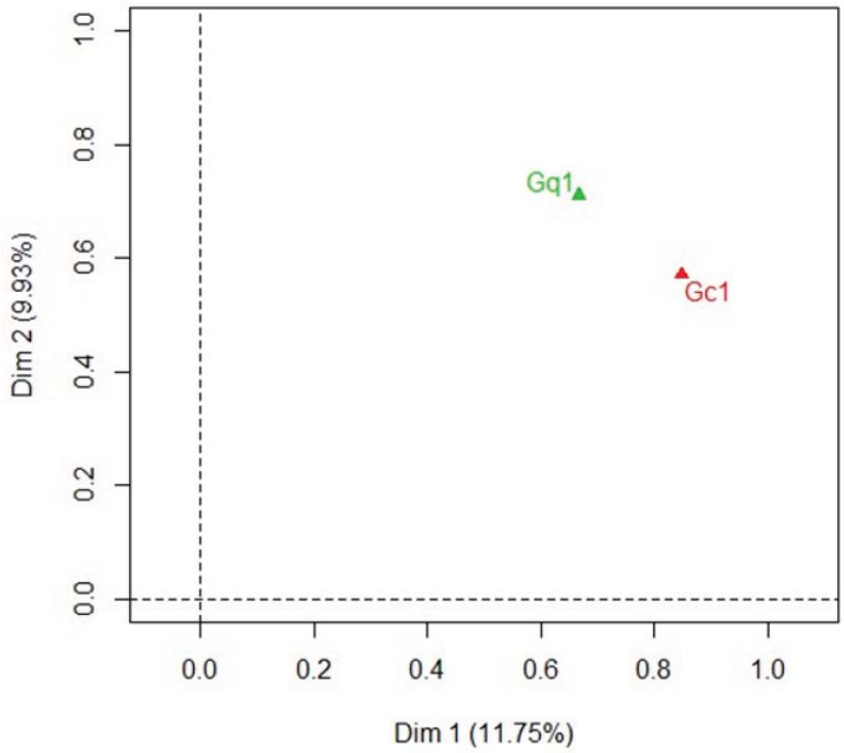

Figure 5 – Projection of the Gc1 and Gq1 in Dimension 1 and Dimension 2 axes

The Lg coefficient, which explains the degree of projection of the variables in the vector space, indicated that the group of variables with the highest projection (highest value of Lg) in this study was Gq1 (VPA settings). Table 5, below, lists all reported values for Lg.

Table 2 – Lg values

	Gc1	**Gq1**	**MFA**
Gc1	2,75	1,25	2,65
Gq1	1,26	4,71	3,95
MFA	2,65	3,95	4,36

The Rv coefficient, which explains the degree of similarity between the groups, indicated that the group of variables with the highest similarity index (highest Rv value) in this study was also the Gq1 group. In Table 3, below, all reported values for Rv are listed.

Table 3 – Rv values

	Gc1	**Gq1**	**MFA**
Gc1	1	0,35	0,76
Gq1	0,35	1	0,87
MFA	0,76	0,87	1

Table 7, below, lists the value the contribution of the two groups in the 7 dimensions; the higher the value, the more relevant the contribution of the group to that dimension. In this study, we have considered the contribution of the two groups of variables only in Dimensions 1 and 2, which have sufficient explanatory power for the data analysed.

Table 4 – Gc1 and Gq1 contributions to Dimensions 1 and 2

	Dim,1	**Dim,2**	**Dim,3**	**Dim,4**	**Dim,5**	**Dim,6**	**Dim,7**
Gc1	56,02	44,52	49,73	12,65	51,99	39,47	26,58
Gq1	43,98	55,48	50,26	87,34	48,00	60,53	73,41

The multivariate analysis has shown that quantitative variables with the highest percentage of correlation in the two dimensions were: F0 median (MED), F0 first derivative standard deviation (DPF), and Spectral tilt standard deviation (DES). DPF may be interpreted as the degree of abrupt fundamental frequency changes whereas DES may be seen as the degree of vocal effort variation.

DELTA
36.3
2020

In Table 5, below, we have listed all the Gc1 variables that showed significance in Dimensions 1 and 2. In the table, the percentage of correlation and the statistical significance of each variable (p, value) are reported. The variable with the correlation value closest to 1 is the variable that best describes a dimension.

Table 5 – Correlation percentage and p value of Gc1 variables

Dim,1 quanti		
	correlation	**p,value**
Z,MED	0,67	0
Z,AES	0,48	0,0002
Z,DPF	0,47	0,0003
Z,QAF	0,45	0,0006
Z,SPQ	0,36	0,0069
Z,ASF	-0,40	0,0026
Z,DES	-0,41	0,0019
Z,MÊS	-0,42	0,0013
Z,SLT	-0,61	0
Dim,2 quanti		
	correlation	**p,value**
Z,DPF	0,65	0
Z,DES	0,57	0
Z,MEF	0,52	0,0001
Z,SPQ	0,39	0,0033

Figure 6, below, shows the projection of all Gc1 variables in the vector space; the closer to the edge of the circumference, the more significant is the variable.

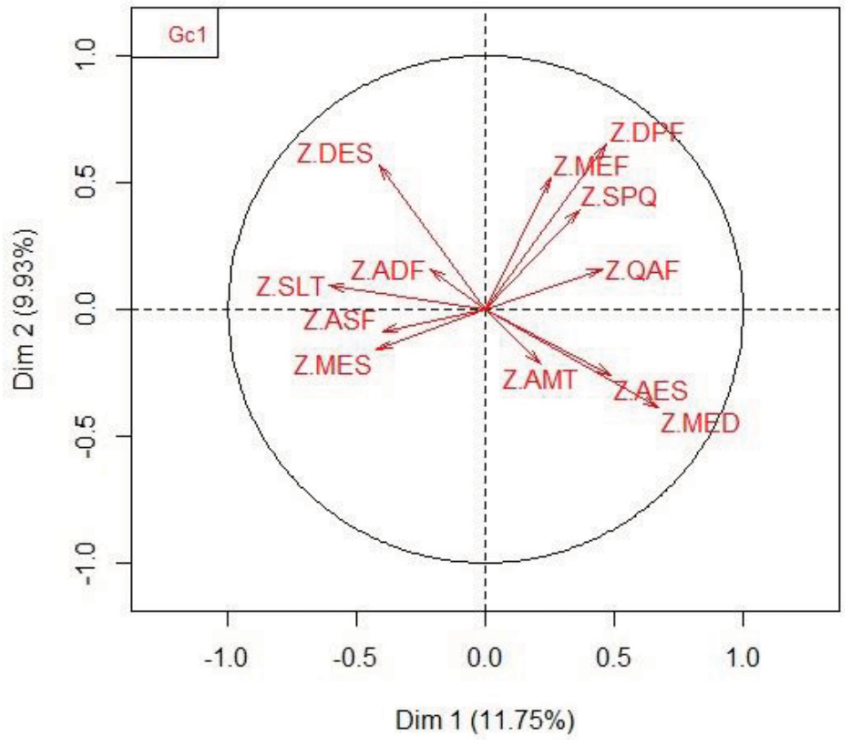

Figure 6 – Projection of all Gc1 variables in the vector space

Table 6, below, presents the values for the three most significant variables in all stimuli: F0 median (MED), F0 first derivative mean (DPF), and Spectral tilt standard deviation (DES).

DELTA
36.3
2020

Table 6 – DPF, DES and MED values

Stimuli	DPF	DES	MED	Stimuli	DPF	DES	MED	Stimuli	DPF	DES	MED
1	0.41,	0.19,	-0.04,	**19**	2.49,	0.26,	-0.80,	**37**	0.97,	0.23,	-0.39,
2	0.52,	0.27,	-0.009,	**20**	4.79,	0.33,	-0.72,	**38**	2.43,	0.20,	0.14,
3	0.72,	0.19,	-0.04,	**21**	0.20,	0.24,	-0.68,	**39**	1.25,	0.12,	1.84,
4	0.64,	0.18,	0.43,	**22**	1.59,	0.21,	-0.27,	**40**	1.34,	0.18,	0.79,
5	0.81,	0.14,	0.74,	**23**	1.40,	0.24,	-0.23,	**41**	2.03,	0.17,	0.94,
6	1.29,	0.14,	1.09,	**24**	5.61,	0.14,	2.26,	**42**	0.64,	0.14,	-0.08,
7	2.23,	0.10,	2.37,	**25**	0.91,	0.15,	-0.57,	**43**	0.87,	0.18,	1.04,
8	0.64,	0.13,	0.24,	**26**	0.54,	0.25,	0.02,	**44**	1.45,	0.22,	-0.20,
9	0.48,	0.15,	-0.12,	**27**	0.72,	0.24,	1.23,	**45**	1.07,	0.22,	0.47,
10	0.44,	0.11,	0.05,	**28**	1.89,	0.19,	0.19,	**46**	0.59,	0.20,	-0.13,
11	0.77,	0.11,	0.61,	**29**	1.32,	0.19,	0.95,	**47**	1.03,	0.22,	0.26,
12	0.60,	0.15,	0.13,	**30**	1.69,	0.18,	0.18,	**48**	0.50,	0.17,	-0.10,
13	1.23,	0.24,	-0.04,	**31**	1.76,	0.18,	2.17,	**49**	0.61,	0.17,	0.48,
14	2.56,	0.20,	0.17,	**32**	1.63,	0.12,	0.84,	**50**	0.40,	0.18,	1.63,
15	2.12,	0.20,	0.09,	**33**	1.04,	0.17,	-0.71,	**51**	0.49,	0.19,	0.75,
16	1.12,	0.09,	0.84,	**34**	1.04,	0.17,	-0.38,	**52**	0.48,	0.20,	1.14,
17	1.51,	0.21,	-0.56,	**35**	1.29,	0.19,	-0.47,	**53**	0.43,	0.23,	-0.19,
18	1.20,	0.17,	0.38,	**36**	1.65,	0.16,	0.25,	**54**	0.51,	0.18,	0.32,

As for the Gq1 group, the quantitative variables that presented the highest percentages were: Lowered larynx, Lip rounding and Breathy voice in Dimension 1; Lowered larynx, Lip rounding and Minimised pitch range in Dimension 2. In table 7, below, the correlation (R2) and statistical significance of each setting (p, value) are reported. The variable with the closest value to 1 in R2 is the one that best describes a dimension.

Table 7 – Correlation percentage and p value of Gq1 variables

Dim,1 quali		
	R2	**p,value**
Lowered larynx	0,39	0,0001
Lip spreading	0,28	0,0003
Breathy Voice	0,28	0,001
Dim,2 quali		
	R2	**p,value**
Lowered larynx	0,66	0,0001
Lip spreading	0,59	0,0003
Minimised pitch range	0,44	0,001

The acoustic and vocal variables clustered the 54 stimuli in 6 groups, under the influence of the acoustic measurements (Gc1) and the voice quality and vocal dynamics settings (Gq1). The grouping of stimuli in the factor map is shown in Figure 7 below. The factor map shows the distribution of the stimuli in Dimensions 1 and 2. Each cluster in the factor map is shown in a different color. The small squares within the graph are the centroids of each cluster, which represent the mean values for the stimuli that make up each cluster.

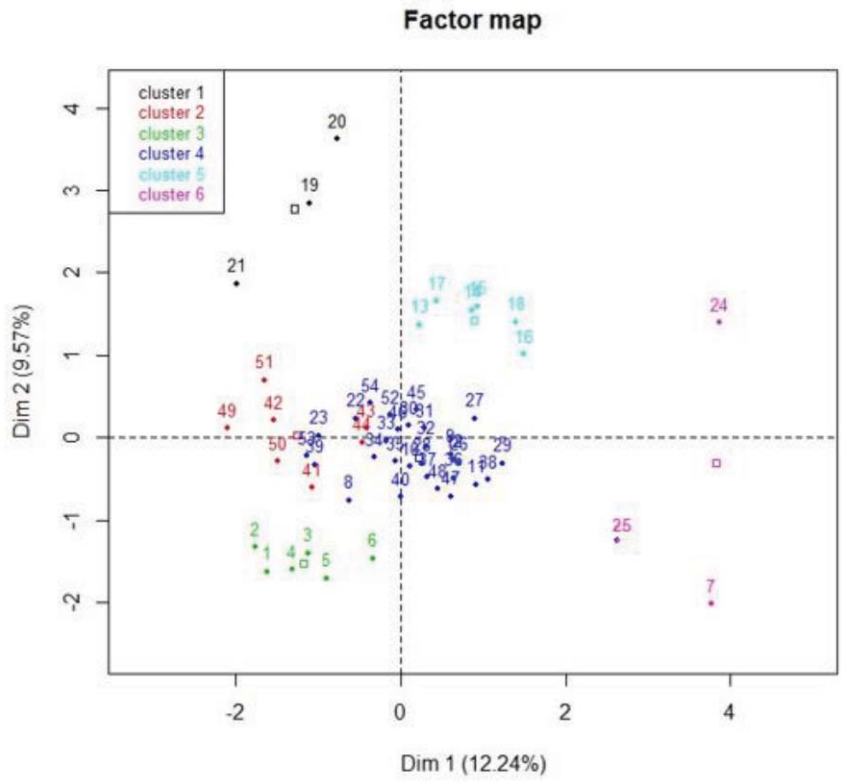

Figure 7 – *Factor map of the stimuli*

Cluster 1, in black, contains three stimuli from Chief Bogo´s speech productions.

Cluster 2, in red, contains the stimuli from Nick Wilde´s speech productions

DELTA
36.3
2020

Cluster 3, in green, contains the stimuli from Assistant Mayor Bellwether´s speech productions in earlier scenes, when she still presented herself as helpful and friendly.

Cluster 4, in dark blue, presents the largest number of stimuli. It contains the stimuli from Judy Hopps' speech productions, as well as half of the stimuli from Nick Wilde´speech productions. Furthermore, this cluster contains five of the stimuli from Assistant Mayor Bellwether, taken from the final scenes of the film. The cluster also contains two stimuli from Chief Bogo´s speech productions.

Cluster 5, in light blue, is composed solely of speech samples from Chief Bogo´s speech productions.

The sixth and final cluster, in magenta, is composed of three stimuli made up of fixed expressions: one from Judy Hopps', one from Chief Bogo's and one from Assistant Mayor Bellwether's. As fixed expressions they tend to have determined pitch contour and are used produced with a higher rate of articulation. These two factors, pitch and rate, belong to the prosodic part of VPA.

We interpret the fact that some characters´ speech samples were distributed into a greater number of clusters in relation to the complexity of the roles they play in the animation. Chief Bogo´s speech samples are distributed into four clusters, Assistant Mayor Bellwether into three, Nick Wilde into two and Judy Hopps into one. Chief Bogo alternates between a very severe attitude when demanding the policeman to work to a flexible one when welcoming a bunny (Juddy Hopps) and a fox (Nick Wilde) into his team of brave policemen. He also shows empathy towards Judy acting as a counsellor when she mentions that she has decided to leave the Police Department because she feels guilty. Assistant Mayor Bellwether alternates from pretending to be docile to revealing herself as a villain and Nick Wilde changes from a deceptive to a trustworthy attitude. Judy remains diligent and optimistic all the time.

The cluster analysis was derived by means of the MFA (Multivariate Factor Analysis) method, because there are two types of variables (qualitative and quantitative). The number of clusters is based on the inertia gain in relation to the number of variables, the number of tokens

and the precision index. According to these settings, the cutting in the dendrograms of the hierarchical cluster analysis is defined (Husson and Pagès, 2011). Figure 8 shows the hierarchical clustering of the 54 stimuli in 6 groups, under the influence of the acoustic measurements (Gc1) and the voice quality and vocal dynamics settings (Gq1).

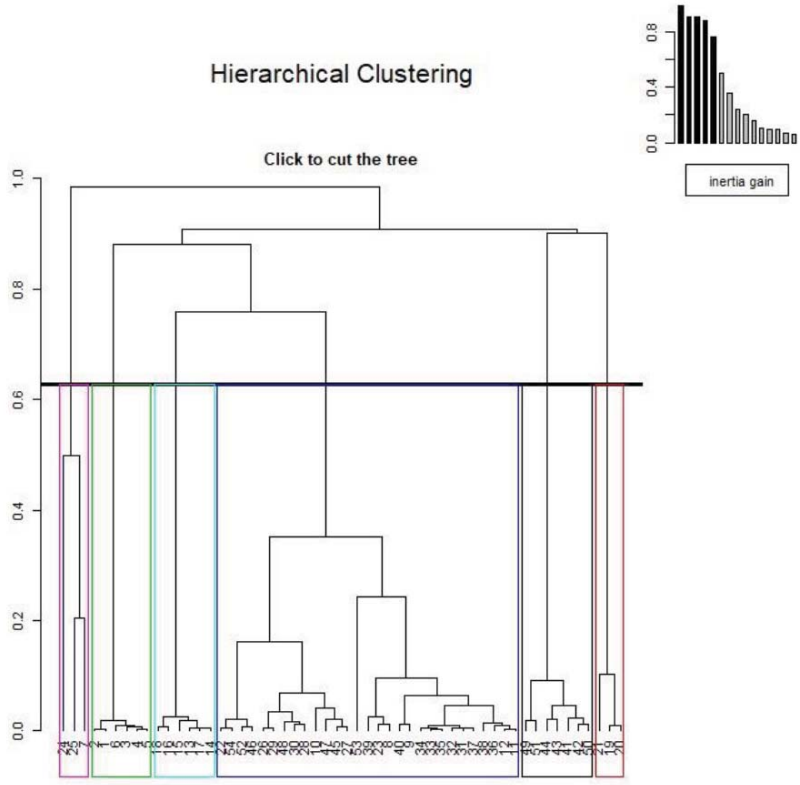

Figure 8 – Hierarchical clustering of the stimuli

The stimuli were not grouped according to the gender of the characters, but according to the vocal features and the resulting acoustics that reflected personal traits of the characters and their attitudes in specific contexts of the film.

DELTA
36.3
2020

5. Discussion

The results of the acoustic analysis showed variation in the measurements extracted by the ExpressionEvaluator script, in relation to speech situations, that is, the same character in more relaxed or tense situation would change their voice quality settings, which would affect their acoustic outputs and the values for the acoustic parameters. Thus, the clusters generated by the MFA showed the stimuli were dispersedly distributed across the bidimensional space and mostly grouped in multiple clusters for each character.

Judy Hopps, the bunny, was the only one whose stimuli were grouped in a single cluster. The character, described as determined and optimistic, and also perceived as fragile and silly by the other characters, showed no behavioral changes throughout the narrative. The main setting identified in the character's stimuli was High pitch variability, which was exhibited in 13 of the 16 stimuli. According to the Frequency Code, higher F0 values correlate to smaller animals and amiable attitudes.

The stimuli from Nick Wilde, described as charming and persuasive, were located in two distinct clusters, 2 and 4. These clusters, however, are shown to be close together in the bidimensional space created by the MFA method, as seen in the factor map, Figure 7. The first of these clusters contains all the stimuli of the character that present Minimised pitch range, Low mean loudness, and Tense larynx and vocal tract. The stimuli in the second cluster present none of these settings. The most recurring vocal quality setting used by Nick is Breathy voice.

The latter setting also appears in the first six stimuli of Assistant Mayor Bellwether, when she presents herself as friendly and helpful as a front to earn the trust of the other characters. According to the Sirenic Code, Breathy voice is related to seductiveness and bashfulness.

The settings of laryngeal and vocal tract tension presented by the character of Assistant Mayor Bellwether in her final scenes characterize her speech at a time when she demonstrates more assertive facets of her personality, which contrasts with the docile attitude she presented in the earlier parts of the narrative.

Chief Bogo, the buffalo, was the character that showed the most variation in the attitudes shown in the scenes, and the distribution of his speech stimuli in the clusters reflected this variation, as his stimuli were distributed across 4 clusters. Unlike Nick Wilde, Chief Bogo's stimuli not only form different clusters, these clusters are also shown to be dispersed in two-dimensional space, as demonstrated in the factor map of Figure 7.

The Lowered larynx and Lip rounding, which were present in most of his stimuli, result in lower F0 values, which correlate to a lower pitch. These acoustic features of Chief Bogo's voice also align with patterns described in the Frequency Code, in which lower pitched vocalizations are linked to larger animals, and in human voices relate to aggressive and dominant attitudes.

In his first three stimuli, taken from a scene in which he berates Judy, the authoritarian character of the character's personality is reflected in his speech, in an aggressive and hostile way of speaking. These stimuli are all grouped into a single cluster (1). Cluster 5 consists of stimuli taken from a scene in the film in which the character seeks to console Judy Hopps, still as an authority figure, but with a compassionate disposition.

In stimuli 22 and 23, which are grouped with the stimuli of the other characters in cluster 4, the character presents an increased pitch variability setting. At this point in the film, the character adopts a friendlier demeanor.

The final stimulus, 24, on the other hand, exhibits Lowered larynx and Lip rounding setting, and represents the return to his hostile and dismissive attitude (albeit as a joke).

6. Conclusion

As presented in our Introduction, this study sought answer the following questions: what types of voice quality and vocal dynamics settings were used by voice actors in the Brazilian dubbing of "Zootopia" to compose the vocal profile of the characters? Do they reflect stereotypes understood as sound symbolic coded?

DELTA
36.3
2020

We found that all four of the characters used non-neutral settings for Vocal Tract features, Overall Muscular Tension and Prosodic Features. Three of the characters used at least one non-neutral Phonation setting; Judy was the only one not to exhibit this type of setting in any of her stimuli. Temporal organization settings also appeared in the stimuli of three characters; Chief Bogo was the only exception.

We found that the overall configuration of the vocal profiles of each character were distinct from one another, showing symbolic uses of phonic matter for the expression of meanings, and according to the context of interaction between the characters.

The voice actor for Judy Hopps, the bunny, used non-neutral settings that shortened the vocal tract (Lip spreading, Decreased lip extension, Raised larynx, Extensive jaw range and tense vocal tract) and vocal dynamics settings like high mean pitch, high mean loudness and fast speaking rate. These settings and the resulting acoustic features fit with the idea of a small, agile animal.

The voice actor for Chief Bogo, the buffalo, used non-neutral settings that lengthen the vocal tract (Lip rounding, Lowered larynx, Opened jaw and Lowered tongue body), and vocal dynamics settings like low mean pitch and high mean loudness. This vocal characterization is consistent with the strong and authoritarian image of the character.

For the character Nick Wilde, the fox, settings that lengthen the vocal tract (such as Backed tongue body and Lowered larynx) and settings that shorten the vocal tract (Raised larynx) were exhibited. Unlike Judy Hopps and Chief Bogo, the voice actor exhibited lax settings for both the vocal tract and the larynx. Breathy voice was the most recurring setting in the character's stimuli. In terms of vocal dynamics, the settings for mean pitch (high and low) and speaking rate (fast and slow) changed between segments, but the degrees of pitch extension and loudness variability were maintained through all the stimuli.

The voice actor for Assistant Mayor Bellwether also used settings that are opposite either in their configuration or their impressive effects: Fast and slow speaking rate; Lowered tongue body and Raised larynx;

Breathy voice, which is pleasant, and Harsh voice and Pharyngeal constriction, that cause unpleasant impressive effects.

The presence of opposite settings in Assistant Mayor Bellwether's speech samples signals the different roles played by the character in different moments of the narrative; the affable little sheep who is friendly and helpful, and the bitter, conniving villain detailing her plan.

The first six stimuli, which were taken from earlier scenes in the film when she presented herself as affable, all exhibited the Breathy voice setting. That setting appeared in none of the final six stimuli, where she reveals herself as the villain; in those stimuli, the most recurring setting was Tense vocal tract.

As for the acoustic parameters, we highlight F0 values to separate the voices of large and small animals, and the spectral tilt descriptor to characterize breathiness in the expression of persuasion and tension in contexts of conflict between the characters.

The comparison between F0 values demonstrate the relationship between the vocalizations and the size of the animals according to the Frequency Code and the expressions of attitude (low values indicating dominance, and high values, fragility) and emotion in speech (low values for affable feelings and high values for irascible feelings).

As such, the results of the analyses confirmed our hypotheses that the vocal profiles for each of the characters would be distinct from one another, and that the vocal profiles and their resulting acoustic features would reflect the patterns described in the Frequency Code (Laver, 1980) and the Sirenic Code (Gussenhoven, 2016).

The findings of this study regarding the tension settings and Breathy voice can be compared to the results of Teshigawara (2003). In his study, Teshigawara noted that the Breathy voice setting appears in the voices of the heroes in Japanese animations. In the Brazilian Portuguese dubbing of "Zootopia", the setting appears in Nick's voice, who is described as charming and persuasive, and in Assistant Mayor Bellwether's voice, when she presents herself as friendly and docile. In contrast, the laryngeal tension setting appears the latter character's stimuli, when she assumes the role of villain. In the Teshigawara (2003),

DELTA
36.3
2020

the voices of the villains were characterized by the larynx sphinctering, which correlates to the laryngeal tension setting.

This last setting also appears in the voice of Chief Bogo in the scenes in which he antagonizes Judy and shows a hostile demeanor.

When compared, these results point towards biological and universal basis for the link between those settings and the characters who exhibit them, since they show that characters meant to be sympathetic or attractive to audience and "villains" have similar voice quality settings in two different cultures.

Choices like the bunny for an energetic character, who is seen as naive by the other characters, and the fox for a con man show how the biological foundation and the folklore associated with the behavior of these animals have shaped the set up for the narrative. The presence of vocal stereotypes in the character's vocal profiles, especially those related to F0, can be seen as an extension of the stereotypes that guide the choice of each animal to incorporate the characters regarding their aspects of personality, attitudes and social roles in the allegorical narrative of the film. The fox, for instance is in a lot of cultures (oriental and occidental) is associated with "cunningness" as manifested in sayings such as "as sly as a fox" in English, "Listig som en räv" (smart as a fox) in Swedish, "esperto como uma raposa" in Portuguese and "rusé comme un renard" in French, "astuto commo un zorro" in Spanish as well as in other languages such as Japanese, Chinese, Turkish, Chinese and Russian. The same for the association between the rabbit and activeness, bull and strength, sheep and meekness and black sheep and oddness.

Phonetic knowledge of production and perception of meaning effects linked to the voice is of importance for the characterization of attitudes and emotions in speech. The choice of settings of voice quality and vocal dynamics by the speaker affects the attribution of meanings by the listeners due to the symbolic value that the acoustic features convey about the speaker (Ohala, 1983; 1984; Chuenwattanapranithi, 2008; Gussenhoven, 2002; 2004, 2016).

This study brings contributions to the field of investigation of expressive vocal prosody in the sense that it provides evidence in favor

of the symbolic use of sound, that is, on the motivated relationship between sound and meaning in a genre that is little explored in phonetic literature.

Concerning voice acting, we believe this study will contribute to the improvement of training for voice actors and provide them with theoretical and practical resources that will help them with their performances when playing animated characters.

As "Zootopia" has been dubbed in over 25 languages, we intend to extend this research by investigating the characters' vocal profiles in different languages in order to map linguistic and cultural factors that may interfere in the codification of sound symbolism, and to investigate how the interactions between biological, psychological, social and linguistic factors shape vocal stereotypes.

Acknowledgements

The authors thank Plínio A. Barbosa for providing the ExpressionEvaluator script used in the study. The first author acknowledges a grant from the São Paulo Research Foundation (FAPESP grant #2017/10725-3).

References

ABELIN, Asa. 1999. *Studies in Sound Symbolism.* Doctoral dissertation. Gothenburg: Göteborg University.

ALBANO, Eleonora. 1988. Fazendo sentido do som. *Ilha do Desterro* - A Journal of English Language, Literatures in English and Cultural Studies 19, 11–26.

BARBOSA, Plínio A. 2009. Detecting changes in speech expressiveness in participants of a radio program. In: *INTERSPEECH.* p. 2155-2158.

BOERSMA, Paul; WEENINK, David. 2018. *Praat: doing phonetics by computer* [software]. Versão 6.0.42.

CHUENWATTANAPRANITHI, Suthathip; XU, Yi; THIPAKORN, Bundit. 2008. Encoding emotions in speech with the size code — A perceptual investigation. *Phonetica,* v. 65, n. 4, p. 210-230.

ELAN [software]. 2018. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Versão 5.4.

FÓNAGY, Iván. 2001. *Languages within Language: an Evolutive Approach.* Amsterdam: John Benjamins.

FÓNAGY, Iván. 1983. La vive voix: *Essais de psycho-phonétique*. Paris: Payot.

FONTES, Mário Augusto S. 2014. *Gestualidade vocal e visual, expressão de emoções e comunicação falada.* 193f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem). Programa de Estudos Pós-Graduados em Linguística Aplica e Estudos da Lingugem – Pontifícia Universidade Católica, São Paulo.

FOX, John; BOUCHET-VALAT, Milan. 2019. *Rcmdr: R Commander.* R package, versão 2.6-1.

GILES, Howard; SCHERER, Klaus R.; TAYLOR, Donald M. 1979. Speech markers in social interaction. In: Scherer, K.R & Giles, H. (Ed.). *Social markers in speech.* Cambridge, UK : Cambridge University Press, p. 343-381.

GOBL, Christer; CHASAIDE, Ailhbe N. 2003. The role of voice quality in communicating emotion, mood and attitude, *Speech Communication,* 40(1–2), 189–212.

GOFFMAN, Erving. 1981. *Forms of Talk.* Filadélfia: University of Pennsylvania Press.

GUSSENHOVEN, Carlos. 2002. Intonation and interpretation: Phonetics and Phonology. In: International Conference on Speech Prosody. 1., 2002. *Proceedings of the First International Conference on Speech Prosody.* Aix-en Provence: Laboratoire Parole et Language, p. 47-57. Disponível em: <http://gep.ruhosting.nl/carlos/aix2002guss.pdf>. Acesso em 5 de setembro de 2016.

_____. 2004. *The Phonology of Tone and Intonation.* Cambridge: Cambridge University Press.

_____. 2016. Foundations of Intonational Meaning: Anatomical and Physiological Factors. *Topics in Cognitive Science*, v. 8, n. 2, p. 425-434, March 2016.

JAKOBSON, Roman. 1978. *Six Lectures on Sound and Meaning.* Cambridge, MA: MIT Press.

JAKOBSON, Roman; WAUGH, Linda R. 1979. *The Sound Shape of Language*. Indiana University Press and Harvester Press.

HECHT, Marvin A.; LAFRANCE, Marianne. 1995. How (fast) can I help you? Tone of voice and telephone operator efficiency in interactions. *Journal of Applied Social Psychology,* v. 25, n. 23, p. 2086-2098, December 1995.

HINTON, Leanne; NICHOLS, Johanna; OHALA, John J. (Ed.). 1994. *Sound Symbolism.* Cambridge: Cambridge University Press.

HONIKMAN, Beatrice. 1964. Articulatory settings. In: ABERCROMBIE, David et al. (Eds.). In: *Honour of Daniel Jones.* Londres: Longman. p. 73-84.

HUSSON, Fançois; LÊ, Sebastien; PAGÈS, Jérôme. 2011. *Exploratory Multivariate Analysis by Example Using R.* Boca Raton: CRC Pres, Taylor and Francis Group.

HUSSON, François et al. 2013. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R. R package version 1.25. Disponível em: http://CRAN.R-project.org/package=FactoMineR.

KREIMAN, Jody; SIDTIS, Diana. 2011. *Foundations of voice studies:* An interdisciplinary approach to voice production and perception. Boston: Wiley-Blackwell.

KÖHLER, W. 1947. *Gestalt Psychology.* New York: Liveright.

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

_____. 2000. Phonetic evaluation of voice quality. In: KENT, Raymond D.; BALL, Martin John (Ed.). *Voice quality measurement.* San Diego: Singular Publishing Group. p. 37-48.

LAVER, John et al. 1981. *A perceptual protocol for the analysis of vocal profiles. Edinburg University Department of Linguistics Work in Progress.* Edinburgh, 1981; 14, p. 139-55.

LAVER, John; MACKENZIE BECK, Janet. 2007. *Vocal Profile Analysis Scheme – VPAS.* Edimburgo: Speech Science Research Centre - Queen Margareth University College (QMUC).

LAVER, John; TRUDGILL, Peter. 1979. Phonetic and linguistic markers in speech. In: SCHERER, Klaus R.; GILES, Howard (Ed.). *Social markers in speech.* Cambridge: Cambridge University Press. p. 1-32.

MACKENZIE BECK, Janet. 2005. Perceptual analysis of voice quality: the place of the Vocal Profile Analysis. In: HARDCASTLE, William J.; MACKENZIE BECK, Janet (Ed.). *A Figure of Speech:* a Festschrift for John Laver. Mahwah: Lawrence Erlbaum Associates, Publishers. p. 285-322.

MADUREIRA, Sandra. 2004. A expressão de atitudes e emoções na fala. In: KIRILLOS, Leny (Org.). *Expressividade.* São Paulo: Revinter. p. 15-25.

_____. 2011. "The Investigation of Speech Expressivity". In: H. Mello, A. Panunzi, T. Raso (Eds.). *Illocution, modality, attitude, information patterning and speech annotation.* Firenze: Firenze University Press, 1, p. 101-118.

_____. 2018. Brazilian Portuguese rhotics in poem reciting: perceptual, acoustic and meaning-related issues. In: Mark Gibson; Juana Gil. (Org.). *Romance Phonetics and Phonology.* 1 ed. Oxford: Oxford University Press, v. 1, p. 76-96.

MORTON, Eugene S. 1977. On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds. *The American Naturalist*, v. 111, n. 981, p. 855-869, setembro/outubro 1977.

OHALA, John J. 1984. An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice. *Phonetica,* v. 41, n. 1, p.1-16.

R CORE TEAM. 2013. *R: A language and environment for statistical computing.* Viena: R Foundation for Statistical Computing.

SAPIR, Edward. 1927. Speech as a Personality Trait. *American Journal of Sociology*, v.32, p. 892-905.

SCHERER, Klaus R. 1979. Personality markers in speech. In: SCHERER, Klaus R.; GILES, Howard (Ed.). *Social markers in speech.* Cambridge: Cambridge University Press, p. 147–209.

_____. 1972. Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. In: *Journal of Personality,* vol. 40, n° 2, p. 191-210.

_____. 1989. Vocal measurement of emotion. In: PLUTCHIK, Robert; KELLERMAN, Henry. (Org.). *Emotion:* Theory, research, and experience. Vol. 4: The measurement of emotion. New York: Academic Press.

SCHERER, Klaus R.; ROSENTHAL, Robert; KOIVUMAKI, Judy. 1972. Mediating interpersonal expectancies via vocal cues: Differential speech intensity as a means of social influence. In: *European Journal of Social Psychology,* vol. 2, n° 2, p. 163-175.

SCHERER, Klaus R.; LONDON, Harvey; WOLF, Jared J. 1973. The voice of confidence: Paralinguistic cues and audience evaluation. In: *Journal of Research in Personality,* vol. 7, n° 1, p. 31-44.

SCHERER, Klaus; SCHERER, Ursula. 1981. Speech Behavior and Personality. In: DARBY, John K. (Ed.). *Speech Evaluation in Psychiatry.* Nova York: Grune & Stratton. p. 115-135.

TESHIGAWARA, Mihoko. 2003. *Voices in Japanese Animation:* A Phonetic Study of Vocal Stereotypes of Heroes and Villains in Japanese Culture. 224f. Dissertation (Doctor of Philosophy). University of Victoria, Victoria.

YARMEY, A. Daniel. Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Apllied Cognitive Psychology*, v. 7, n. 5, p. 419-431, October 1993.

TSUR, Reuvent. 1992. *What Makes Sound Patterns Expressive?* The Poetic Mode of Speech Perception. Durham, NC: Duke University Press.

ZUCKERMAN, Miron; HODGINS, Holley; MIYAKE, Kunitate. 1990. The vocal attractiveness stereotype: Replication and elaboration. *Journal of Nonverbal Behavior,* v. 14, n. 2. p. 97-112, June 1990.

ZUCKERMAN, Miron; MIYAKE, Kunitate. 1993. The attractive voice: What makes it so? *Journal of Nonverbal Behavior,* v. 17, n. 2, p. 119–135, June 1993.

Recebido em: 27/07/2020
Aprovado em: 03/09/2020

DELTA
36.3
2020

Annex A

VPA protocol (Laver & Mackenzie Beck, 2007)

	FIRST PASS		SECOND PASS						
	Neutral	Non-neutral	SETTING	Moderate			Extreme		
				1	2	3	4	5	6
A. VOCAL TRACTFEATURES									
Labial			Lip Rounding/p rotrusion						
			Lip spreading						
			Labiodenta Liz ation						
			Minimised range						
			Extensive range						
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 bocy						
			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						
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								

D. PROSODIC FEATURES									
		Neutral	SETTING	Moderate			Extreme		
				1	2	3	4	5	6
13. Pitch	Mean		High						
			Low						
	Range		Minimised range						
			Extensive 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. Dyplophonia			Absent						

DELTA
36.3
2020

Appendix A

Brazilian Portuguese stimuli selected for analysis and corresponding lines in the original English dub

Stimulus	Character	Transcription
1	Assistant Mayor Bellwether	*A maioria da população é de presas, Judy, e agora, estão todos com medo.* (Our city is ninety percent prey, Judy, and right now, they're just really scared.)
2		*Você é a heroína deles, confiam em você. E, é por isso, que o Chefe Bogo e eu –* (You're a hero to them. They trust you. And so that's why Chief Bogo and I -)
3		*- queremos que seja garota propaganda do departamento.* (- want you to be the public face of the ZPD.)
4		*Judy, se dedicou tanto pra chegar até aqui.* (Judy, you've worked so hard to get here.)
5		*Queria isso desde que era um filhote.* (It's what you wanted since you were a kid.)
6		*Não pode desistir.* (You can't quit.)
7		*Pensa nisso, -* (Think of it, -)
8		*- noventa por cento da população unida contra um inimigo em comum.* (- ninety percent of the population united against a common enemy.)
9		*Nós seremos imbatíveis.* (We'll be unstoppable!)
10		*Armei pro Leãonardo, -* (I framed Lionheart, -)
11		*- posso armar pra vocês dois.* (- I can frame you too!)
12		*É a minha palavra contra a de vocês.* (It's my word against yours.)

13		*Abandonou o seu posto, incitou o pânico, -* (Abandoning your post, inciting a scurry, -)
14		*- expôs os roedores ao perigo. Mas, -* (reckless endangerment of rodents but, -)
15		*pra ser justo, impediu um gênio do crime de roubar duas dúzias de cebolas!* (-to be fair, you did stop a master criminal from stealing two-dozen moldy onions!)
16		*Insubordinação.* (Insubordination.)
17		*Agora, eu vou abrir essa porta e você vai dizer a ela que é uma ex-agente de trânsito com mania de grandeza, -* (Now I'm going to open this door and you're going to tell that otter you're a former meter maid with delusions of grandeur -)
18	Chief Bogo	*- e que não vai pegar esse caso!* (-who will not be taking the case!)
19		*Não queira levar todo o crédito por isso, Hopps.* (Don't give yourself so much credit, Hopps.)
20		*A divisão do mundo sempre existiu. Pra isso, temos bons policiais.* (The world has always been broken, that's why we need good cops.)
21		*Como você.* (Like you.)
22		*Nós temos alguns novos recrutas hoje aqui, -* (We have some new recruits with us this morning, -)
23		*- incluindo a nossa primeira raposa.* (- including our first fox.)
24		*Tô nem aí.* (Who cares.)

25		*Peraí, peraí.* (Hey, hey!)
26		*Ninguém diz pra mim o que eu posso ou não posso ser.* (No one tells me what I can or can't be.)
27		*Principalmente uma raposa –* (Especially not some jerk -)
28		*- que leva a vida enganando os outros vendendo –* *(*- who never had the guts to try and be anything more than -)
29		- picolés falsos. (- a pawpsicle hustler.)
30		*Umm, eu não queria discordar do senhor, mas não são cebolas.* (Mmm, hate to disagree with you, sir, but those aren't onions.)
31		*É uma variedade de crocus, chamada Midnicampum holicithias. Classe C em botânica, senhor.* (Those are a crocus variety called Midnicampum holicithius. They're a Class C botanical, sir.)
32	Judy Hopps	*Eu cresci numa família muito ligada à agronomia...* (Well, I grew up in a family where plant husbandry was kind of a thing...)
33		*O que foi que você disse?* (What was it you said?)
34		*É muito fácil verificar uma placa?* (Any moron can run a plate?)
35		*Poxa, se tivesse algum malandro por aqui que não fosse tão covarde.* (Gosh, if only there were a moron around who were up to the task.)
36		*Ah, Nick!* (Oh, Nick!)
37		*Os uivantes não são lobos, -* (Night howlers aren't wolves, -)
38		*- são plantas tóxicas.* (-they're toxic flowers.)
39		*Alguém tá fazendo isso com os predadores –* (I think someone is targeting predators on purpose -)
40		*- pra que se tornem selvagens.* (- and making them go savage.)

41		*Tá bom, olha, -* (all right, look, -)
42		*- todos que vêm para Zootopia acham que podem ser o que quiserem, só que não.* (- everyone comes to Zootopia thinking they can be anything they want. Well, you can't.)
43		*Só pode ser o que já é.* (You can only be what you are.)
44		*Raposa esperta, coelha tosca.* (Sly fox, dumb bunny.)
45		*Olha, vocês dão a ela uma roupa de palhaço, um triciclo ridículo e dois dias pra resolver um caso que vocês trabalham a duas semanas.* (Look, you gave her a clown vest, a three-wheeled joke-mobile, and two days to solve a case you guys haven't cracked in two weeks.)
46		*É, por isso ela buscou ajuda de uma raposa.* (Yeah, it's no wonder she needed to get help from a fox.)
47	Nick Wilde	*Porque vocês não ajudaram, -* None of you guys were gonna help her, -)
48		*- não foi?* (were you?)
49		*Se o mundo sempre vai ver uma raposa como traiçoeira –* If the world's only gonna see a fox as shifty -)
50		*- e não confiável, -* (- and untrustworthy, -)
51		*- não faz sentido querer ser outra coisa.* (- there's no point in trying to be anything else.)
52		*Se liga, se eu quisesse evitar as câmeras por tá fazendo alguma coisa ilegal, –* (You know, if I wanted to avoid surveillance because I was doing something illegal, -)
53		*- o que eu nunca fiz –* *(*- which I never have, -)
54		*- usaria o túnel de manutenção 6B.* *(*- I would use the maintenance tunnel 6B.)