

# ANIMATED FILM CHARACTER PROFILES: THE ROLES OF VOICE AND LEXICAL CONTENT

Alice Crochiquia<sup>1,2</sup>, Anders Eriksson<sup>1</sup>, Sandra Madureira<sup>2</sup>, Plinio A. Barbosa<sup>3</sup>

<sup>1</sup>Stockholm University, Sweden; <sup>2</sup>PUC-SP, Brazil; <sup>3</sup>UNICAMP, Brazil  
 alice.crochiquia@ling.su.se; anders.eriksson@ling.su.se;  
 sandra.madureira.liaac@gmail.com; pabarbosa.unicampbr@gmail.com

## ABSTRACT

Our study examines the attribution of physical and psychosocial features to the characters of a dubbed animated film by comparing the results of a listening test experiment with the results of a reading test experiment. The aim of this study was to investigate the potential influence of lexical content on the perceptual evaluation of speaker characteristics. In the listening test, respondents were asked to listen to audio samples of dubbed dialogues in the film produced by voice actors and rate the characters on fourteen bipolar continuous scales. The task in the reading test was the same, except for the stimuli, which were transcriptions of the audio files. Results from both tests are congruent, as the character profiles based on the perception scores from each test are nearly identical for all characters. Differences are mainly due to the degree judgement scores are amplified by the characters' voice quality profiles.

**Keywords:** dubbing, perception, animated film, voice acting, character features.

## 1. BACKGROUND

A professional film dubbing is a collaborative effort by translators, dialogue writers, dubbing directors, voice actors and sound engineers, and the result is a type of production in which features of spoken and written text are present [1]. An important step in the process of dubbing is the selection of actors to lend their voices to the characters in a new language, as audiences expect "character synchrony" [2], that is, vocal stereotyping and the association of specific vocal features to certain types of characters.

Compared to dubbing for live-action films, dubbing for animated films, especially those aimed at children, presents far fewer constraints regarding another type of synchronization, phonetic or lip synchrony, which is the matching of articulatory movements shown on screen and sounds produced by the voice actors. That is due both to the less precise articulatory movements of the characters in comparison to real actors and child audiences being less demanding [3]. Thus, in those cases, priority can be given to translations that are semantically similar to the wording in the original text but not in

articulatory detail, and in that way, keep the content of character dialogues in the target language close to the one in the source text.

This paper is about the influence of lexical content on the perception of characters from an animated film, where we compare the results from a reading test experiment to results from a previous study, which included a listening test experiment.

The previous study [4] covered the character synchrony in the Brazilian Portuguese dubbing of an animated film, in a combined analysis of the impressions of listeners regarding the attitudes, physical and social characteristics of the characters, and the acoustic parameters of each of the voice actors who performed as the characters. The results of that experiment showed that actors' voice acoustic profiles were mainly differentiated in relation to  $f_0$  mean and range, and harmonic to noise ratio (HNR). Based on the perception scores of physical, psychological, and social features of the Listening Test, five distinct character profiles were generated. From those two sets of data, correlations between acoustic, physical, social, and attitudinal characteristics were estimated. Among the correlations with the highest values were  $f_0$  maximum and Size,  $f_0$  maximum and Age, *jitter* and Honest and Kind.

A major concern in studies of the perception of physical, psychological, and social features based on speech stimuli is the influence of semantic and syntactic content of the recorded material to be judged [5].

There are several masking techniques [6] to isolate voice from linguistic information, among them reverse speech, listening to unknown languages, meaningless content recordings, randomizing splicing, content filtering, and even acoustic-driven phoneme removal transformation techniques [7] but all of them have drawbacks. Listeners perceive them as odd and perceptual judgments are affected. Neutralizing techniques, which use the same text voiced by different speakers, are also often applied to avoid lexical influence but as [5] mentions despite the same semantic and syntactic content perceptual judgments are not essentially neutral because they can be influenced by judges' social, behavioral, and individual characteristics.

Given the semantic, syntactic, and vocal interacting factors influencing listeners' judgments of speech stimuli in perceptual tests, investigating how lexical

content affects listeners' judgments is an important goal and the focus of the present work.

## 2. METHODS

### 2.1. Material

The material for both experiments was taken from the Brazilian dubbing of the Disney animated film Zootopia. The stimuli in the Listening Test were audio samples of dialogues produced by voice actors performing as the four main characters in the film (Assistant Mayor Bellwether, Chief Bogo, Judy Hopps, and Nick Wilde). For the experiment, five audio samples were selected for the corpus as Assistant Mayor Bellwether, who is revealed as the film's villain, has two drastically different characterizations throughout the narrative, and was thus considered as two distinct characters. The stimuli in the Reading Test were the transcriptions of those audio samples.

### 2.2. Listening Experiment

The experiment was conducted entirely online, on SurveyMonkey. The first page of the link given to the respondents presented a brief description of the experiment and a questionnaire to collect data on the respondents' background, including whether or not they had seen the film. A group of 77 Brazilian Portuguese native speakers, aged between 20 and 50, took part in the experiment.

The respondents, 46 women and 31 men, were asked to listen to the audios and rate the characters on 14 bipolar continuous scales regarding their size, age, temper, attitudes, character, social and vocal features. The respondents had to do that by using a sliding tool to indicate the point of the scale they felt best described the character, which was translated into a score from 0 to 100. Table 1 shows the scales and the pairs of descriptors used in the experiment in Brazilian Portuguese and their translation to English.

Based on the mean values of the perception scores for the five characters, a profile was established for each of them. Table 2 shows the profiles of the friendly Bellwether (B1), Bellwether's villainous persona (B2), Chief Bogo (CB), Judy Hopps (JH), and Nick Wilde (NW).

For most of the features considered for the profiles, we selected those with scores higher than 65.5 and lower than 35.5 as significant and added the qualifiers based on the scores within those ranges. 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. More details regarding the characters' profiles can be found in [4].

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

Table 1: Scales and descriptors in the Listening test.

| Characters | Features  |
|------------|---|
| <b>B1</b>  | Small, Young<br>Happy, Kind, Very Gentle,<br>Charming<br>Lively, Pleasant                       |
| <b>B2</b>  | Small, Young<br>Rather Tough, Rather Aggressive,<br>Not Charming<br>Rather Lively, Not Pleasant |
| <b>CB</b>  | Big, Not Young<br>Very Tough, Very Aggressive,<br>Not Charming<br>Not Lively, Not Pleasant      |
| <b>JH</b>  | Small, Young<br>Honest, Happy,<br>Charming<br>Rather Lively, Pleasant                           |
| <b>NW</b>  | Mid-sized<br>Not Young, Charming Not<br>Lively, Pleasant  |

Table 2: Character profiles.

### 2.3. Reading Experiment

The design for this test was nearly identical to the previous one. Among the few differences were the

exclusion of the question about hearing impairments in the questionnaire and the change for the last two scales from *Voç* to *Maneira de expressão* (Manner of expression) as the test did not include listening to the voices. The descriptors all remained the same.

The test was presented, also on SurveyMonkey, to a group of 73 Brazilian Portuguese native speakers, 43 women and 30 men, aged between 19 and 58. The stimuli in the two tests were not referred to by the names of the characters in the film, nor did any of those names occur in the audio and text samples.

#### 2.4. Interrater reliability

For a perception experiment based on ratings by listeners to be reliable, there must be a high degree of agreement between judges. As a measure of interrater agreement, we have used Cronbach’s Alpha. With this method, values can range from 0 to 1, indicating complete disagreement and perfect agreement, respectively. Values above 0.7 are considered acceptable, and scores higher than 0.9 are considered excellent.

### 3. RESULTS

Table 3 shows the Cronbach’s Alpha values between male and female judges, and between those who had seen the film (**S**) and those who had not (**nS**) in the Reading Test.

|            | z-norm | Male  |       | Female |
|------------|--------|-------|-------|--------|
|            |        | M/F   | S/nS  | S/nS   |
| <b>B1</b>  | 0.987  | 0.990 | 0.980 | 0.988  |
| <b>B2</b>  | 0.934  | 0.984 | 0.973 | 0.986  |
| <b>CB</b>  | 0.951  | 0.965 | 0.916 | 0.968  |
| <b>JH</b>  | 0.955  | 0.967 | 0.873 | 0.911  |
| <b>NW</b>  | 0.940  | 0.931 | 0.890 | 0.927  |
| <b>All</b> | 0.961  | 0.976 | 0.953 | 0.965  |

**Table 3:** Interrater reliability for the Reading Test.

Table 4 presents the mean score values. Low values are closer to the descriptor on the left end of the scale, and higher values to the right end of it. To compile more detailed impressions of the characters, we’ve chosen to highlight scores higher than 80 and lower than 20 in this experiment as strong, as opposed to the 65.5 and 35.5 values in the first experiment. These cutoff points are not definitive and may still be modified in planned future analyses and experiments.

Table 5 shows the comparison between the mean values of the perception scores from listeners and those from readers who had not seen the film, who did not have any information on the characters beyond the reading of the transcriptions during the test. The correlations show that even in the complete absence of speech information and without previous

knowledge of the characters in the film, the reading profiles are similar to listening profiles.

### 4. DISCUSSION

Results of the Listening Test and the Reading Test were found to be coherent, that is, the same characteristics were attributed to the characters in the two tests. However, some of the characteristics were evaluated with stronger scores in one of them.

Strong scores in both tests were attributed to Bellwether 1 on the following characteristics: Docile, Energetic, Happy, Lively; to Bellwether 2, Energetic; and to Chief Bogo, Dominant and Tough.

The similarities between the results for the two tests extended beyond the very high or very low scores. Judy and Nick received overall weaker scores in both tests. The strongest score attributed to Judy was Energetic (83 in the Listening Test and 74 in the Reading Test) and to Nick was Dominant (34 in the Listening Test and 21 in the Reading Test) and Bold (67). Nick did not have scores below 31 or above 60 on any of the other scales. In the Listening Test, his strongest scores were for Charming and for a Pleasant voice.

Coherence in relation to both the Listening Test and the Reading Test can be found in the characteristic attributions to Bellwether 1 as Docile, Gentle and Honest and to Bellwether 2 as Aggressive, Tough, and Dishonest; to Chief Bogo as Big, Old, and Tough; to Judy as Energetic and Honest. Nick had a few close scores in Listening and Reading, however, most of them were around the mid-mark of the scales, unlike the strong scores of both Bellwethers and Chief Bogo.

Comparison inter-characters shows that Chief Bogo got the strongest scores for Old, Aggressive and Tough; Bellwether 1 for Young, Kind, Docile, Pleasant, Happy, Lively, and Charming; Bellwether 2 for Bold.

For many of the acoustic parameters considered in the previous study, a similar pattern to the perception scores emerged, especially those related to *f0*. The values for Bellwether 1, Bellwether 2 and Chief Bogo were placed at the ends of the spectra, and Judy and Nick fell somewhere in the middle.

The slightly stronger scores on the Listening Test than in the Reading Test convey the impact of pitch and voice quality settings on characteristic attributions. Chief Bogo’s low-pitched voice was paired to Bigger, Older, more Dominant, Tougher judgments; Bellwether 1’s, Bellwether 2’s and Judy’s high-pitched voices were paired with Younger, and more Lively judgments. Nick’s Whispery and Harsh voice quality with slightly higher scores for Charming and Age (Older).

| Scales                          | B1    |    | B2    |    | CB    |    | JH    |    | NW     |    |
|---------------------------------|-------|----|-------|----|-------|----|-------|----|--------|----|
|                                 | L     | R  | L     | R  | L     | R  | L     | R  | L      | R  |
| <i>Small - Big</i>              | 30    | 19 | 24    | 29 | 87    | 63 | 34    | 35 | 53     | 48 |
| <i>Young - Old</i>              | 23    | 27 | 23    | 36 | 65    | 59 | 28    | 36 | 60     | 34 |
| <i>Aggressive - Docile</i>      | 87    | 83 | 31    | 28 | 14    | 25 | 51    | 49 | 57     | 32 |
| <i>Dominant - Submissive</i>    | 56    | 50 | 12    | 22 | 9     | 16 | 31    | 32 | 34     | 21 |
| <i>Lazy - Energetic</i>         | 88    | 83 | 84    | 83 | 78    | 68 | 83    | 74 | 44     | 60 |
| <i>Tough - Gentle</i>           | 87    | 77 | 31    | 24 | 9     | 17 | 42    | 45 | 55     | 31 |
| <i>Charming - Uninteresting</i> | 22    | 67 | 37    | 41 | 49    | 61 | 32    | 39 | 34     | 46 |
| <i>Kind - Mean</i>              | 18    | 22 | 63    | 55 | 28    | 60 | 38    | 38 | 48     | 57 |
| <i>Cautious - Bold</i>          | 50    | 36 | 74    | 72 | 72    | 64 | 68    | 52 | 46     | 67 |
| <i>Happy - Sad</i>              | 11    | 16 | 35    | 37 | 55    | 48 | 29    | 39 | 49     | 41 |
| <i>Dishonest - Honest</i>       | 73    | 71 | 39    | 55 | 41    | 49 | 73    | 70 | 46     | 43 |
| <i>Sophisticated - Simple</i>   | 55    | 53 | 36    | 45 | 50    | 55 | 41    | 44 | 45     | 59 |
| <i>Lively - Monotonous</i>      | 7     | 15 | 26    | 31 | 52    | 48 | 22    | 40 | 42     | 39 |
| <i>Pleasant - Unpleasant</i>    | 26    | 18 | 44    | 49 | 49    | 65 | 29    | 40 | 33     | 57 |
| <b>Correlation</b>              | 0.875 |    | 0.937 |    | 0.841 |    | 0.928 |    | -0.198 |    |

**Table 4:** Mean values for perception scores for the characters' features in the two experiments.

| Scales                          | B1    |       | B2    |       | CB    |       | JH    |       | NW     |       |
|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|--------|-------|
|                                 | L     | R(nS) | L     | R(nS) | L     | R(nS) | L     | R(nS) | L      | R(nS) |
| <i>Small - Big</i>              | 30    | 23    | 24    | 27    | 87    | 56    | 34    | 38    | 53     | 47    |
| <i>Young - Old</i>              | 23    | 33    | 23    | 36    | 65    | 54    | 28    | 38    | 60     | 31    |
| <i>Aggressive - Docile</i>      | 87    | 81    | 31    | 32    | 14    | 26    | 51    | 45    | 57     | 32    |
| <i>Dominant - Submissive</i>    | 56    | 46    | 12    | 22    | 9     | 17    | 31    | 32    | 34     | 20    |
| <i>Lazy - Energetic</i>         | 88    | 81    | 84    | 84    | 78    | 66    | 83    | 73    | 44     | 64    |
| <i>Tough - Gentle</i>           | 87    | 72    | 31    | 24    | 9     | 20    | 42    | 47    | 55     | 29    |
| <i>Charming - Uninteresting</i> | 22    | 35    | 37    | 40    | 49    | 64    | 32    | 38    | 34     | 45    |
| <i>Kind - Mean</i>              | 18    | 23    | 63    | 52    | 28    | 60    | 38    | 38    | 48     | 59    |
| <i>Cautious - Bold</i>          | 50    | 38    | 74    | 71    | 72    | 62    | 68    | 50    | 46     | 67    |
| <i>Happy - Sad</i>              | 11    | 18    | 35    | 36    | 55    | 46    | 29    | 41    | 49     | 38    |
| <i>Dishonest - Honest</i>       | 73    | 72    | 39    | 58    | 41    | 46    | 73    | 70    | 46     | 44    |
| <i>Sophisticated - Simple</i>   | 55    | 53    | 36    | 46    | 50    | 55    | 41    | 40    | 45     | 55    |
| <i>Lively - Monotonous</i>      | 7     | 17    | 26    | 30    | 52    | 44    | 22    | 42    | 42     | 35    |
| <i>Pleasant - Unpleasant</i>    | 26    | 21    | 44    | 45    | 49    | 67    | 29    | 40    | 33     | 58    |
| <b>Correlation</b>              | 0.966 |       | 0.921 |       | 0.767 |       | 0.893 |       | -0.215 |       |

**Table 5:** Perception scores of listeners and readers who had not seen the film.

## 5. RESULTS

Overall, results indicate that the lexical content in the dialogues was enough to lead readers to an accurate assessment of the characters especially in the case of Bellwether and her two personas, and we must point out that the majority of the dialogue lines did not contain qualifiers that would immediately make the features of the characters explicit to the listeners and readers.

However, the results in this study have not been enough to lay out in detail how the lexical content may have been a deciding factor in the perception of these characters. We hope to gain more insight into the influence of the lexical content with additional analyses of the dialogue lines and with similar listening and reading experiments with Zootopia in different languages.

With the comparison of the results of both experiments, evidence of the influence of parameters like pitch means and phonation voice quality settings

can be inferred from the higher intensity of characteristics attributed to size, age, and some attitudinal features. The voice performances did match the characters' features displayed in the text and in most cases served to amplify them.

Even in Nick's case, where none of his perception ratings met the new threshold for strong scores, the similar patterns in the results for the perception experiment and the acoustics may still provide some insight into the use of voice as characterization as he is portrayed in the film as a hard-to-read character.

Some of these correlations highlighted as significant in the previous study are already well-documented in the literature on vocal expressivity, like high values for parameters related to  $f_0$  and small size, and the opposite for low values, as well as low values for Harmonic-to-Noise Ratio as a vocal aging index. However, correlations like low H1-H2 values and higher scores for Dominance, Aggressiveness are not as efficiently explained by findings in previous studies and require further inspection.

## 6. ACKNOWLEDGEMENTS

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)* - Finance Code 001, grant #88881.689915/2022-01. Participation at ICPHS 2023 by the first author has been supported by a grant from *The Swedish Foundation for International Cooperation in Research and Higher Education (STINT, BR2027-7253)*. The third author acknowledges a grant from PIPEq #21672, PUC-SP. The fourth author acknowledges a grant from CNPq #302194/2019-3.

## 7. REFERENCES

- [1] C. Bosseaux, “Investigating dubbing: Learning from the past, looking to the future,” in *The Routledge Handbook of Audiovisual Translation*, L. PérezGonzález, Ed. Routledge, 2018, pp. 48–63.
- [2] C. Whitman-Linsen, *Through the dubbing glass: the synchronization of American motion pictures into German, French, and Spanish*. Peter Lang, 1992.
- [3] F. Chaume, “Synchronization in dubbing: a translational approach,” in *Topics in Audiovisual Translation*, P. Orero, Ed. John Benjamins Publishing Company, 2004, pp. 35–52.
- [4] A. Crochiquia, A. Eriksson, P. Barbosa, S. Madureira, “A perceptual and acoustic study of dubbed voices in an animated film”, *Proc. Speech Prosody 2022*, pp. 565–569, 2022.
- [5] J. Pittam, *Voice in Social Interaction: An Interdisciplinary Approach*. SAGE, 1994.
- [6] K. R. Scherer, S. Feldstein, R. N. Bond, and R. Rosenthal, “Vocal cues to deception: a comparative channel approach,” *Journal of Psycholinguistic Research*, vol. 14, pp. 409–425, 1985.
- [7] C. Noufi, J. Berger, M. Frank, K. Parker, and D. L. Bowling, “Acoustically-driven phoneme removal that preserves vocal affect cues,” *arXiv Preprint*, 2022.