



Pleasantness and Wellbeing in Poem Declamation in European and Brazilian Portuguese Depends Mostly on Pausing and Voice Quality

Plinio A. Barbosa*

Department of Linguistics, University of Campinas, Campinas, Brazil

This work investigates the relationship between the sensations of pleasantness and wellbeing and the acoustics of poem declamation in two varieties of Portuguese: European and Brazilian. Ten speakers in each variety recited Alberto Caeiro's poem "Quando vier a primavera" to which dataset the declamation by two actors was added. Ten listeners in each variety participated in three perception tests for evaluating the degree of pleasantness and wellbeing of the entire reading in two different sessions by using a 5-degree Likert scale, and for choosing the more pleasant recitation from two recitations of a chunk of the poem. Both reciters and listeners were balanced in gender. A set of 22 prosodic parameters was used for predicting the listeners' evaluations in logistic regression models and linear discriminant analysis. Results showed that the parameters related to pausing and voice quality are the main explanatory variables with a minor role for the F0 median. The actors' recitations are preferred and gender differences are related to reciters, not listeners. Minor differences across the varieties were found.

Keywords: poem recitation, prosody, pleasantness, wellbeing, acoustic phonetics

OPEN ACCESS

Edited by:

Oliver Niebuhr,
University of Southern
Denmark, Denmark

Reviewed by:

Albert Rilliard,
Université Paris-Saclay, France
Pärtel Lippus,
University of Tartu, Estonia

*Correspondence:

Plinio A. Barbosa
pbarbosa@unicamp.br

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 14 January 2022

Accepted: 11 March 2022

Published: 11 April 2022

Citation:

Barbosa PA (2022) Pleasantness and Wellbeing in Poem Declamation in European and Brazilian Portuguese Depends Mostly on Pausing and Voice Quality. *Front. Commun.* 7:855177. doi: 10.3389/fcomm.2022.855177

INTRODUCTION

The vocal differences between speaking styles are related to changes in voice quality, as well as rhythm and intonation, according to studies carried out in the field for decades (see Eskénazi, 1993 for a review). Among these, professional styles such as TV and radio broadcasting (Léon, 1993), and reading in literary (book) festivals common in all Europe are remarkable for provoking effects related to pleasantness and wellbeing in the listeners (Fónagy, 1961, 2001). As regards poetry reading, although the aesthetic effect is proper to the culture in which the poetic texts were born, it also depends on prosody, with elements such as novelty and complexity being associated with prosodic aspects such as the rhythmic and intonation organization of the verses that have effects of surprise and expectation, among others (Jacobsen, 2010). A study by Tegnér et al. (2009) showed that other effects can include the increase of wellbeing in cancer patients after 6 weeks of poetry reading groups by causing an improvement of their emotional resilience and a decrease of their anxiety levels.

In a review of studies on the topic of attractiveness in speech, Rosenberg and Hirschberg (2020) showed that, although the concept of "attractiveness" is very context-dependent, common points can be found, such as the fact that male voices with a low fundamental frequency (F0) are preferred by women and are associated with dominance, whereas female voices with a

higher F0 are the preference of men. Quené (2020), Strangert and Gustafson (2008) and Hodges-Simeon et al. (2010) have reported that women also prefer men who have higher speech rates. Baumann (2017) found that women generally found the speech of other women more pleasant, but this result is controversial considering the results by Burkhardt et al. (2010), which suggested this is not true.

Work on German showed that attractiveness, regardless of gender, seems to increase with the variability of F0 (Weiss et al., 2020) as well as with the absence of disfluencies (Strangert and Gustafson, 2008; Weiss and Burkhardt, 2012). Weiss and Burkhardt (2010) also found a relationship between soft and breathy speech, and a lower spectral center of gravity with the growth of positive appreciation by listeners, a result in the same line as the one by Carlsen et al. (2018) who showed that lower harmonic-to-noise ratio (HNR) and higher jitter are related to an increase in pleasantness. As regards creakiness, Anderson et al. (2014) found that the voices of young adult females sound less attractive, less competent, and less educated when they read a phrase imitating vocal fry (creaky voice).

As regards work on Portuguese, pleasantness and aesthetical appreciation of poem recitation were lines of investigation in both European (EP) and Brazilian Portuguese (BP). The main interest of the works by Braga et al. (2007) and Pinto-Coelho et al. (2013) was the choice of human voices for EP Text-to-Speech (TTS) systems. In the first work, the authors evaluated eight subjective dimensions, including pleasantness, in a test with 2-min audio excerpts of 62 female voices reading texts of information delivery aiming at sounding talented as a speaker (these women certainly intended to sound pleasant given that instruction). They showed that pleasantness is one of the components of the choice for a particular voice, but that the general preference was for (female) voices with low fundamental frequencies. In the second work, the authors built a machine learning system for classifying and predicting the degree of the pleasantness of 3-min audio excerpts from 77 EP-speaking female speakers. The training was designed to predict a previously web-based evaluation of a 5-point pleasantness scale obtained from 112 participants from prosodic- and voice-quality acoustic parameters. Results showed that jitter, shimmer, F0 mean, derivative of intensity, and maximum of F0 derivative were the best predictors of pleasantness.

As regards the effects of poetry recitation, by working on BP, Madureira et al. evaluated the declamation of a single poem at each time by one or more reciters, both professional and non-professional. Madureira (2008) investigated the perceptual effects of the production choices of two professional speakers (male and female) reciting the same poem, “Soneto da Fidelidade” from Vinícius de Moraes. A group of 30 listeners evaluated the impressions carried out by the voices in both recitations. That of the male speaker was considered enthusiastic and splendid, whereas that of the female speaker evoked an impression of grief and sadness. The acoustic analysis of both readings revealed that whispery voice, F0 narrow range, and a great number of silent pauses were the production strategies used by the female speaker to express those negative effects, whereas the male speaker employed much varied intonation

patterns to sound more expressive to provoke a sensation of liveliness. Madureira and Fontes (2019) analyzed face movements and acoustic parameters of the recitation of the poem “A Valsa” by a Brazilian actor to show that LTAS slope, mandibular range, and F0 maximum are related to surprise and contempt inferred automatically from the Affectiva facial analysis system. It was also a single professional speaker reading a single poem that was analyzed by Madureira and Camargo (2010) to show a relation between prosodic parameters and subjective dimensions. These dimensions were evaluated by 30 college students in a semantic differential scale questionnaire with the following descriptors: activation (calm/activated), valence (pleasant/unpleasant), emotion (joy, sadness, anger, exasperation), and speech acts (advice, admonition, order, and plea). Repetitions of verses recited in different ways were used to relate prosodic variation and perceptual effect. They showed that pleasantness was related to an expanded pharynx, which causes F0 lowering (see Camargo et al., 2019). In a follow-up study of the effects of the recitation of the poem “Soneto da Fidelidade” (Menegon et al., 2021), this time recited by eight non-professional speakers, pleasantness was moderately relevant as a subjective parameter and related to voice quality and the parameter of LTAS slope. These works on BP and EP suggest that melodic, temporal, and voice quality dimensions could explain subjective positive evaluations in both poetry (BP) and ordinary prose (EP).

As can be seen from this overview, there are issues related to the link between prosodic-acoustic parameters and the sensations of attractiveness, pleasantness, and speech positive appraisal that seem to be recurrent, while others need further study such as the question of appreciation by men and women of the same-gender vs. cross-gender speech. Further investigation on acoustic parameters such as correlates of vocal effort, pausing, intensity variability could shed light on aesthetic appreciation. These measures are included here.

Besides extending the spectrum of parameters related to the acoustics of pleasant speech, additional research on Portuguese could help contribute to advance knowledge on the link of speech production and subjective appreciations such as pleasantness and wellbeing, the latter an under-studied theme.

Poetry seems to be appropriate for examining this link because declamation usually focuses on provoking a particular effect on the listener, and not only on delivering information. It is worth noting that the scientific community is developing a sense of the importance of poetry as a means of increasing quality of life and this can be exemplified with a journal specializing on the results of therapy using poem reading and declamation, the *Journal of Poetry Therapy* sponsored by the National Association for Poetry Therapy in the USA and publishing since 2003. In this vein, it can be said that poetry actually seems to increase wellbeing in a therapy situation, at least. That is why the present work investigates the link between the acoustics and the perception of pleasantness and wellbeing of poem recitation in two varieties of Portuguese: Brazilian and European. The findings of this work can also shed light on the relation between the use of voice and speech and aesthetical pleasure, which give elements for the training of professional speakers.

The work presented here concerns the evaluation of the degrees of the listener's feelings of pleasantness ("agradabilidade" in Portuguese) and wellbeing ("bem-estar" in Portuguese) caused by the listening of a Portuguese poem, as explained in detail in Section Pleasantness and Wellbeing Evaluations. Similarly, to other works reviewed here, this investigation also uses the recitation of a single poem but expands the number of speakers to twenty, which allows a larger study of prosodic variability.

Based on the examined literature and general expectations based on cultural aspects or common sense, the main hypotheses for poem recitation in this study are:

- H1. Pleasantness and wellbeing appraisals are equivalent, but the second has a weaker relation to acoustics because it is more subjective than the first;
- H2. Lower mean F0 increases both the sensations of pleasantness and wellbeing for women listening to the recitation of men;
- H3. Higher mean F0 increases both the sensations of pleasantness and wellbeing for men listening to the recitation of women;
- H4. Diverging from what is said in the literature on attractiveness, due to the appraisal of care with the content in poem recitation, slowing down and more frequent pauses increase both pleasantness and wellbeing;
- H5. Breathier speech increases pleasantness and wellbeing;
- H6. More variable F0 increases pleasantness and wellbeing;
- H7. The lesser the vocal effort, the higher the sensations of pleasantness and wellbeing;
- H8. Portuguese listeners appreciate the declamation by Brazilians more than Brazilians appreciate the declamation by Portuguese speakers, due to the greater exposure in Portugal to the speech of Brazilians in Portuguese TV, especially Brazilian soap operas;
- H9. The acoustic parameters related to pleasantness and wellbeing are the same and in the same degree of relative importance in the two varieties of Portuguese, in the absence of evidence to the contrary;
- H10. Men and women use the same prosodic-acoustic parameters to make the speech better appreciated by the listener;
- H11. Actors' voices are more pleasant and cause more wellbeing than those of lay reciters.

In the next section, the methodology for testing the eleven hypotheses above is presented. Section Results sets forth the results, which are discussed in detail in Section Discussion. Section Conclusion summarizes the main findings.

METHODOLOGY

The PROS-POIESIS corpus is formed by the declamation of the poem "Quando vier a primavera" (When Spring comes) by Alberto Caiero, one of Fernando Pessoa's heteronyms, by ten Brazilian speakers and ten Portuguese speakers in balanced gender. All speakers do not have voice professional training and have between 25 and 50 years of age to avoid an age group with

effects of vocal aging (Stathopoulos et al., 2011) that usually affect the values of parameters such as F0 median, jitter, shimmer, and speech rate.

Due to the Covid-19 pandemic, the participants themselves used the Easy Voice app on their cell phones to make all recordings. Because this app allows choosing among different codifications, instructions were given to record all audio files in PCM format (WAV) at a sampling rate of 48 kHz. The author, who is a trained phonetician, further evaluated all audio files. The recordings were resampled at 16 kHz and leveled to the same maximum intensity level at 65 dB.

As control recordings, the declamations of the same poem by Brazilian actor Ivan Lima (72 years old at the time of recording) and by Portuguese actor Pedro Lamares (33 years old at the time of recording) were included in the dataset. The reason for choosing these two actors was simply the public availability of these recordings for the same poem. It is expected that they get the highest scores of pleasantness and wellbeing due to their professional voices. The recitation by Pedro Lamares is available here since 2011: <<https://www.youtube.com/watch?v=ZNWEmKLLFWA>> and that by Ivan Lima, here since 2020: <https://www.youtube.com/watch?v=7QP_cEEQ8RA>. Unfortunately, no actresses were found reading the same poem in both varieties of Portuguese.

Acoustic Parameters

The Prosody Descriptor Extractor script for Praat (Boersma and Weenink, 2021) implemented by the author (Barbosa, 2020) was used to extract 22 prosodic-acoustic parameters from 14 chunks from one to two verses of the spoken poem including the silent pause at the end, when applicable. This segmentation is shown in the Appendix in the original version followed by a translation for the sake of making the meaning transparent for the general reader. The criteria used for choosing this segmentation are two-fold: (1) to ensure an acceptable unit of meaning when each chunk is heard in isolation; (2) to ensure a minimum duration of about 3 s for the computation of the acoustic parameters in each chunk as explained in the following paragraphs.

The segmentation of the chunks was the same across the 20 reciters and the reading of the poem lasted from 60 to 95 s. A set of 22 parameters was computed for each chunk, a single value per chunk. This set can be divided into the following classes: (1) Class melody: twelve descriptors of fundamental frequency (F0): median, semi-amplitude between quartiles (F0SAQ), F0 minimum and maximum, standard-deviations of values and time positions of F0 local peaks, F0 peak rate, and F0 mean peak bandwidth within each chunk, mean and standard-deviations of rates of F0 rises and falls; (2) Class intensity: two intensity descriptors: spectral emphasis and the coefficient of variation of global intensity (the ratio between global intensity standard-deviation and mean); (3) Class VQ: four voice quality descriptors: long-term averaged spectrum (LTAS) slope, Harmonic-to-Noise Ratio (HNR), jitter and shimmer; finally, (4) Class timing: four temporal descriptors: silent pause average duration, pausing rate, speech, and articulation rates.

Melodic measures such as the statistical descriptors of centrality and dispersion (median and semi-amplitude between

quartiles) and measures of extreme values (F0 minimum and maximum) are more easily found in the literature of prosodic variation across styles. To them, we added parameters aiming at describing other aspects of melodic variability. Those are the descriptors related to the peaks of F0 within a chunk. These peaks are automatically obtained by the Prosody Descriptor Extractor script after a three-step technique for obtaining a smoothed F0 contour in Praat: first, a 5-Hz low-pass filter applied to the extracted F0 contour; second, a quadratic interpolation and then a Praat-built function to select the extreme values. From these values, we computed four peak-related measures: the standard-deviation of the F0 peak values, the standard deviation of the F0 peak time positions, the mean of F0 peak rate, and the mean of F0 peak bandwidth. The first three of these F0 peak-related measures are connected to liveliness because it concerns how F0 peaks are distributed across a sound excerpt. The majority of these F0 peaks are pitch accents because of the care to use a 5-Hz smoothing function to highlight the main F0 peaks. Since the mean of F0 peak rate is about 1 s in our data, a minimum duration of 3 s for a chunk allows the computation of mean and standard deviations of F0 peak intervals, which is useful for further investigating F0 peak rate variability across the chunks.

As for F0 bandwidths, by studying the possible correlates of charismatic speech in the speeches of Steve Jobs (Apple) and Mark Zuckerberg (Facebook, recently renamed Meta), Niebuhr et al. (2018) have shown, among other aspects, that Jobs has a broader mean pitch accent bandwidth than Zuckerberg, which possibly contributes to the former sound more charismatically. Because charisma is a positive evaluation that can be attributed to listening to an audio file, we decided to include mean F0 peak bandwidth as a possible parameter for explaining the positive dimensions of pleasantness and wellbeing.

Spectral emphasis was computed according to Traunmüller and Eriksson (2000) as the difference between the spectral energy of the whole band (E) and that of the low band with a cut-off frequency of 400 Hz (E_0) for computing the energy of F0. This cut-off frequency is an approximation of the original proposal of 1.5 times the mean fundamental frequency, which has the same goal of computing the energy of F0 in the lower band. According to the authors, the higher the value of spectral emphasis, the higher the vocal effort. In previous works, we computed spectral emphasis in intervals as short as vowel intervals and showed that this measure was one of the three relevant ones (with duration and F0 standard deviations) to signal levels of lexical stress in Brazilian Portuguese and Swedish (Barbosa et al., 2013; Eriksson et al., 2013). That is why its computation within a 3-s minimum duration chunk is robust enough for the analysis presented here.

LTAS slope, computed here by the difference in mean energy between the bands 0–1 and 1–4 kHz, varies according to differences in voice quality for a series of reasons including breathiness (the flatter the slope, the breathier the voice, see Mendoza et al., 1996), differences in voice projection for singing and acting (Sundberg, 1994), and other aspects not explored here that can be found in the review by Master et al. (2006). This parameter was included to evaluate its potential in explaining the listeners' preferences. A word of caution is necessary though because authors such as Löfqvist and Mandersson (1987) have

shown that LTAS measures stabilize after 10 s of continuously voiced frames of speech, due to variability in articulation. Nevertheless, we found in our data that its variability across the chunks of the same reciter was lesser than its variability across reciters, which can be used as a criterion of discriminability for this parameter across the recitations and a potential predictor of the listeners' preferences for a particular recitation.

Harmonic-to-Noise Ratio (HNR) is a correlate of roughness and breathiness (de Krom, 1994; Verdonck-de Leeuw et al., 2001) as well as creakiness (Khan et al., 2015): the lower its value, the rougher, the breathier and the creakier the speech is. Although often computed from sustained vowels, it has been used in connected speech with success for predicting roughness (Severin et al., 2005). Because low values of HNR can be associated with distinct qualities such as creaky and breathy voices, this measure should be combined with others, such as LTAS, jitter, and shimmer, to discriminate between the specific voice quality.

Jitter and shimmer are measures of irregularity across glottal cycles, the former of glottal period irregularity and the latter of glottal amplitude irregularity. Both are usually computed for sustained vowels but their computation in connected speech could reveal differences in voice qualities because, according to Laver (1980), voice quality is “the characteristic auditory coloring of an individual speaker's voice [...] a cumulative abstraction over a period of time of a speaker-characterizing quality.”

Pleasantness and Wellbeing Evaluations

Five perception tests were carried out, where the first four consisted in the attribution of the degree of pleasantness or wellbeing felt by the listeners, according to the instruction given to them: (1) one Likert-scale test for evaluating pleasantness applied to 10 Brazilian listeners and containing recitations by Brazilian speakers; (2) one Likert-scale test for evaluating wellbeing applied to the same 10 Brazilian listeners and containing recitations by Brazilian speakers; (3) one Likert-scale test for evaluating pleasantness applied to 10 Portuguese listeners and containing recitations by Portuguese speakers; (4) one Likert-scale test for evaluating wellbeing applied to the same 10 Portuguese listeners and containing recitations by Portuguese speakers; and (5) a discrimination test applied to all Brazilian and Portuguese listeners for evaluating paired chunks from both European (EP) and Brazilian Portuguese (BP) recitations. All instructions were given in Portuguese.

As part of the four Likert-scale tests carried out with Brazilian and Portuguese listeners, we included the declamation of the actors to the ones of the ten lay speakers in the data set of each variety of Portuguese. This allowed comparing the evaluations of the actors with those of the lay reciters to answer Hypothesis 11.

The scale for pleasantness evaluation varied in five degrees from “very unpleasant” (“muito desagradável” in Portuguese, degree 1) to “very pleasant” (“muito agradável” in Portuguese, degree 5) with the neutral response (“neutro” in Portuguese) having degree 3. Similarly, the scale for wellbeing evaluation varied in five degrees from “very low wellbeing” (“grande mal-estar” in Portuguese, degree 1) to “very high wellbeing” (“grande bem-estar” in Portuguese, degree 5) with the neutral response also having degree 3. Because the labels have general use in

Portuguese for evaluating external events as being pleasant (“agradável”) or causing wellbeing (“bem-estar”), no particular instructions were given to the participants on how to interpret them. Although the term “mal-estar” could also be interpreted as related to illness, because it was presented in a scaling factor containing the term “bem-estar,” the possibility of interpreting “mal-estar” as causing a sensation of illness to the listener must be discarded. The similarity of the responses for both pleasantness and wellbeing strongly suggests that this illness-related interpretation was not considered by the listeners in the context of the effect of a poem recitation.

As for the discrimination test, a set of 44 pairs using the 14 chunks of the poem recitations separated by a short musical tone and combining different genders and speakers was prepared. The chunks being compared in each pair have the same content and are from the same language variety for avoiding a choice biased toward differences in content or variety. The pool of paired audio files contains all chunks of the poem presented from 2 to 6 times from different reciters to combine readings from the same gender (either males or females) and across genders, taking at least a chunk from all reciters. Apart from that, the choice of a particular pair of reciters in an audio file pair was random. To evaluate the behavior of listeners according to variety, the pooled group of 20 listeners from Brazil and Portugal evaluated the same set. Their task was to say which of the readings is the one that sounds more pleasant to him/her, the first or the second.

For the five tests, the listeners were selected in the same age range as the reciters, that is, between 25 and 50 years of age, to avoid the effects of differences in evaluation from younger or older age groups. The discrimination test also allowed checking the possible effect of the variety of the reciter and the listener when evaluating the pairs of chunks in both BP and EP, as well as whether the mean degrees attributed in the Likert scale tests could predict the performance of the listeners in the discrimination test. Each listener had online participation in the three tests, which were built in the Survey Monkey platform with the help of a research assistant. The discrimination test was performed after the Likert scale tests and the three tests were completed in 30–45 min.

Statistical Models

To test our hypotheses, three kinds of statistical models were run in the R (R Development Core Team, 2008) software. The first kind were correlation tests applied to verify if Brazilian and Portuguese listeners answered in a similar way to the stimuli of the discrimination test and if these responses are correlated with the differences in degree from the Likert-scale test. The correlation for evaluating the degree of similarity of the responses for pleasantness and wellbeing by the listeners of the two varieties was computed to answer Hypothesis 1.

The second kind of statistical model was logistic regression. After a process of progressive model simplification, four final logistic regression models, corresponding, respectively, to the four Likert-scale tests presented in the previous section, were retained that predicted the degree of pleasantness or wellbeing (one of the two predicted variables in each model) from the corresponding prosodic-acoustic parameters (the predictor

variables). For doing so, the degrees from 1 to 5 were, respectively, transformed to 0 to 100% in intervals of 25%, allowing the use of a logistic model. The quasi-binomial family in the R *glm* function was used for prediction when there were important differences between degrees of freedom and amount of deviance, according to statistical theory (Roback and Legler, 2021, chapter. 6). When this was not the case, the binomial family was used. Pseudo-correlation measures were used to evaluate the degree of explained variance of these models, which are measures of effect size. Model simplification always started by using non-correlated acoustic parameters as predictors and successively eliminating those that were not significant (see below for the process of eliminating correlated parameters). The final models contain only significant predictors.

The third kind of statistical model was a model based on linear discriminant analysis (LDA). The final model predicts one of the two possible responses of the corresponding discrimination test (first or second chunk) from the differences of the mean values of each parameter from the two recitations of the respective chunk. In the LDA analysis, there were no predictors related to pausing (pause duration and rate of pause) because a single silent pause ends the majority of the chunks, and this final pause coincides with the end of an utterance, then, pause duration cannot be apprehended. The process of simplification also considered the maximum number of non-correlated predictors with a progressive elimination of non-significant predictors for retaining only significant predictors in the final model.

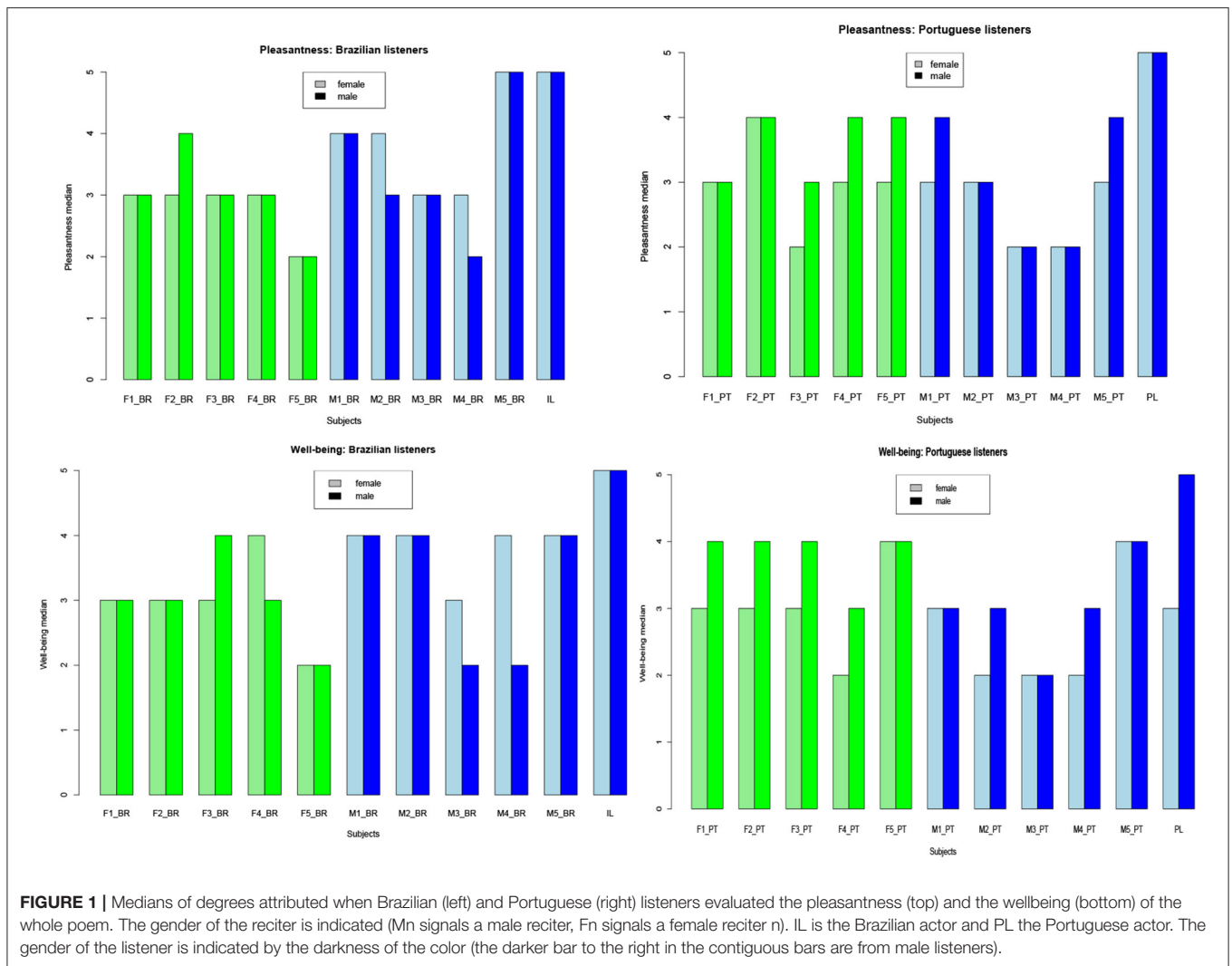
For all models, when two or more predictors of the same class correlated by more than 50%, one of them was taken as a predictor and the other one(s) was not used in the initial model. That procedure allowed us to discard the following parameters: F0 minimum and maximum (both correlated to more than 60% with F0 median), standard-deviations of rates of F0 rises and falls (both correlated to more than 60% with F0SAQ), and speech rate (correlated with more than 70% with articulation rate). In all the statistical models built, the 5% level of significance was chosen.

All stimuli are publicly available for listening at the following URL: <https://figshare.com/authors/Plinio_Barbosa/11320902>.

RESULTS

Figure 1 presents the medians of the degrees attributed to the scales of pleasantness and wellbeing according to the variety and gender of the groups of listeners. No differences between the evaluations of male and female listeners were observed when checking the closeness of median responses either for male or female reciters. The correlation between the evaluation of the two scales was 82% for Brazilian listeners and 89% for Portuguese listeners. This confirms the first part of Hypothesis 1, the close relationship between the two scales as well as the fact that the left part (“mal-estar”) of the wellbeing scale was not interpreted differently from the right part (“bem-estar”), otherwise, the correspondence between the two scales would be low.

In the logistic models, the gender of the reciter in each spoken variety happened to produced different sets of predictors,



that is why each one of the four original models was split in two according to the reciter's gender and these results are then presented gender-wise for each language variety. For the datasets of the two varieties of Portuguese, if a parameter in a specific chunk could not be computed for mathematical reasons, the measures of all the other parameters were not included in the models, for maintaining the whole set of predictors in each model. An example of this case is the computation of peak rate, which requires at least two F0 peaks in the chunk. That is why some boxes in **Figures 2–5** do not present any variability but a median value only. For pleasantness, the actors in each community got the 5 median degrees, whereas for wellbeing the Portuguese actor (PL) got the 4 median degrees and the Brazilian actor (IL) got the 5 median degrees (see **Figure 1**).

The degree of pleasantness attributed to Brazilian female reciters by the Brazilian listeners is explained by three parameters in order of importance: rate of pause (the lower, the more pleasant), the slope of LTAS (the faster the slope, the more pleasant), and pause duration (the longer, the more pleasant). Pseudo-correlation explains 42% of the variance. The degree of

pleasantness attributed to Brazilian male reciters by the same listeners is explained by four parameters: the same as for female reciters in the same order of importance plus F0 median in the last position (the higher, the more pleasant). The parameters explained 23% of the variance. The distribution of the four parameters' values of the BP dataset that are the predictors of the two models can be checked graphically with the boxplots in **Figure 2**. In the case of the rate of pause (top, right) the tendency is not linear and higher means are found for degrees 3 to 5 (recall that the bold horizontal segment in boxplots stands for the median, not the mean).

As for the degree of pleasantness attributed to Portuguese female reciters by the Portuguese listeners, only pause duration was significant (the longer, the more pleasant), explaining 44% of the variance. As for the effect of male recitations on the same listeners, three parameters in order of importance explain the listeners' choices: spectral emphasis, F0 median (the lower, the more pleasant in both cases), and pause duration (the longer, the more pleasant), which explains 38% of the variance. The distribution of the parameters' values that are the predictors of

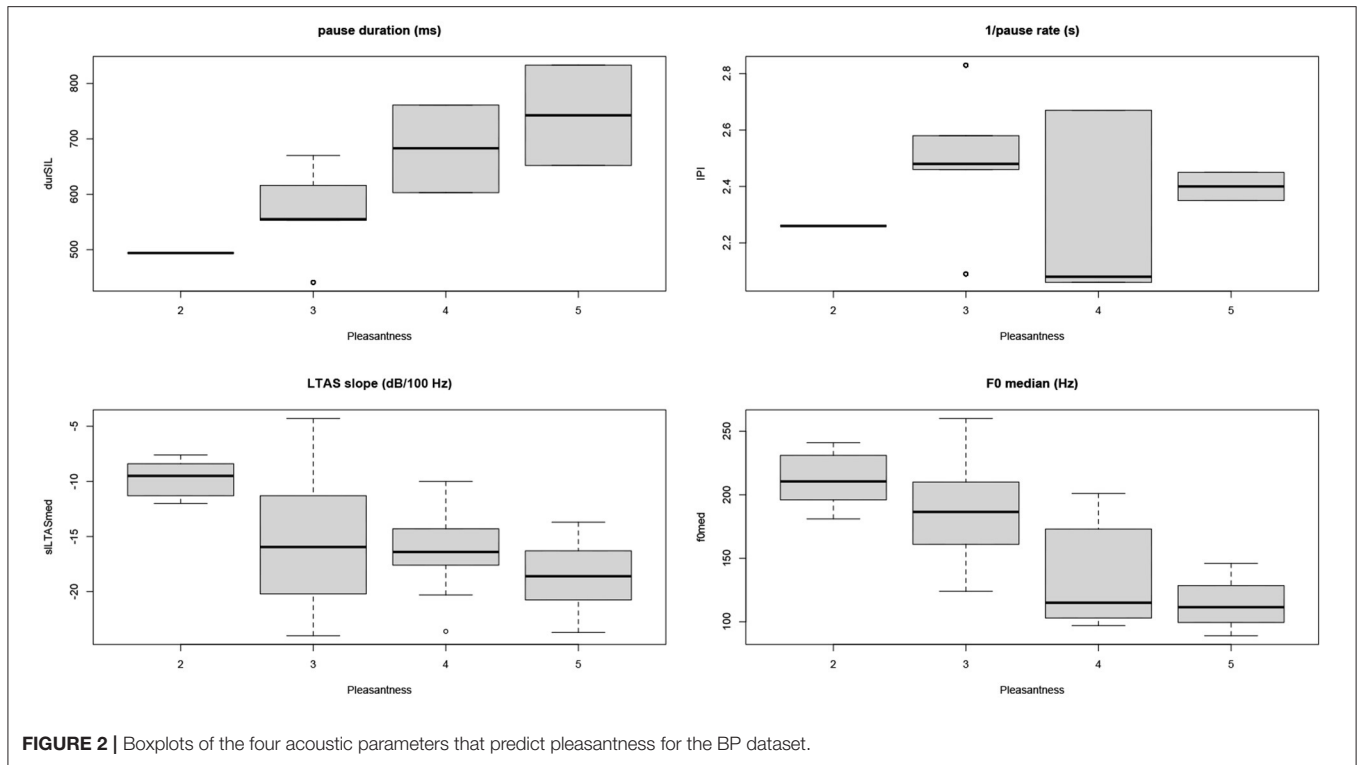


FIGURE 2 | Boxplots of the four acoustic parameters that predict pleasantness for the BP dataset.

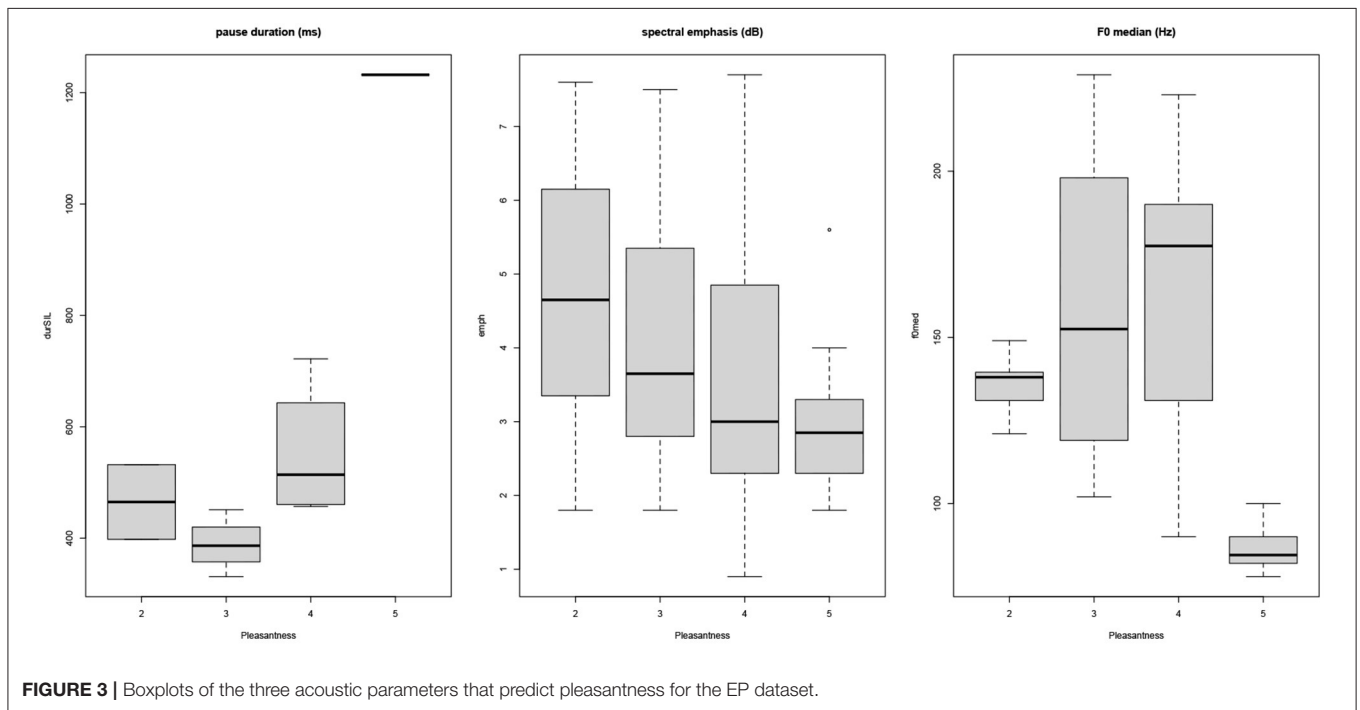
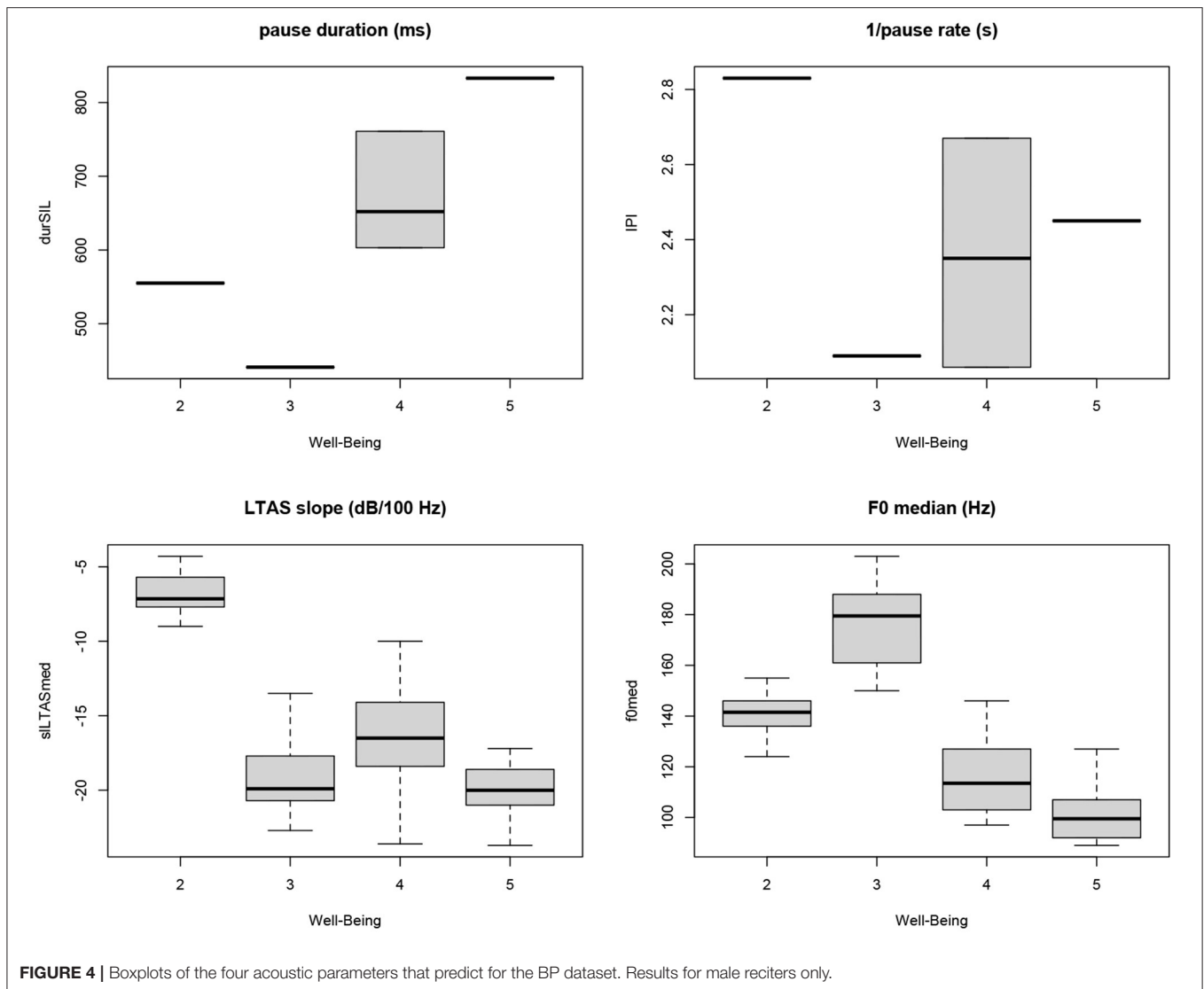


FIGURE 3 | Boxplots of the three acoustic parameters that predict pleasantness for the EP dataset.

the two models can be checked graphically with the boxplots for the three predictors for the EP dataset in **Figure 3**.

The degree of wellbeing attributed to Brazilian female reciters by the Brazilian listeners has a very low explained variance, only 5%. Only two parameters explain the responses, in order

of importance: HNR (the higher, the greater the wellbeing) and slope of LTAS (the faster the slope, the greater the wellbeing). By contrast, the degree of wellbeing attributed to Brazilian male reciters by the same listeners has 27% of explained variance, higher than the one from the model that evaluated pleasantness,



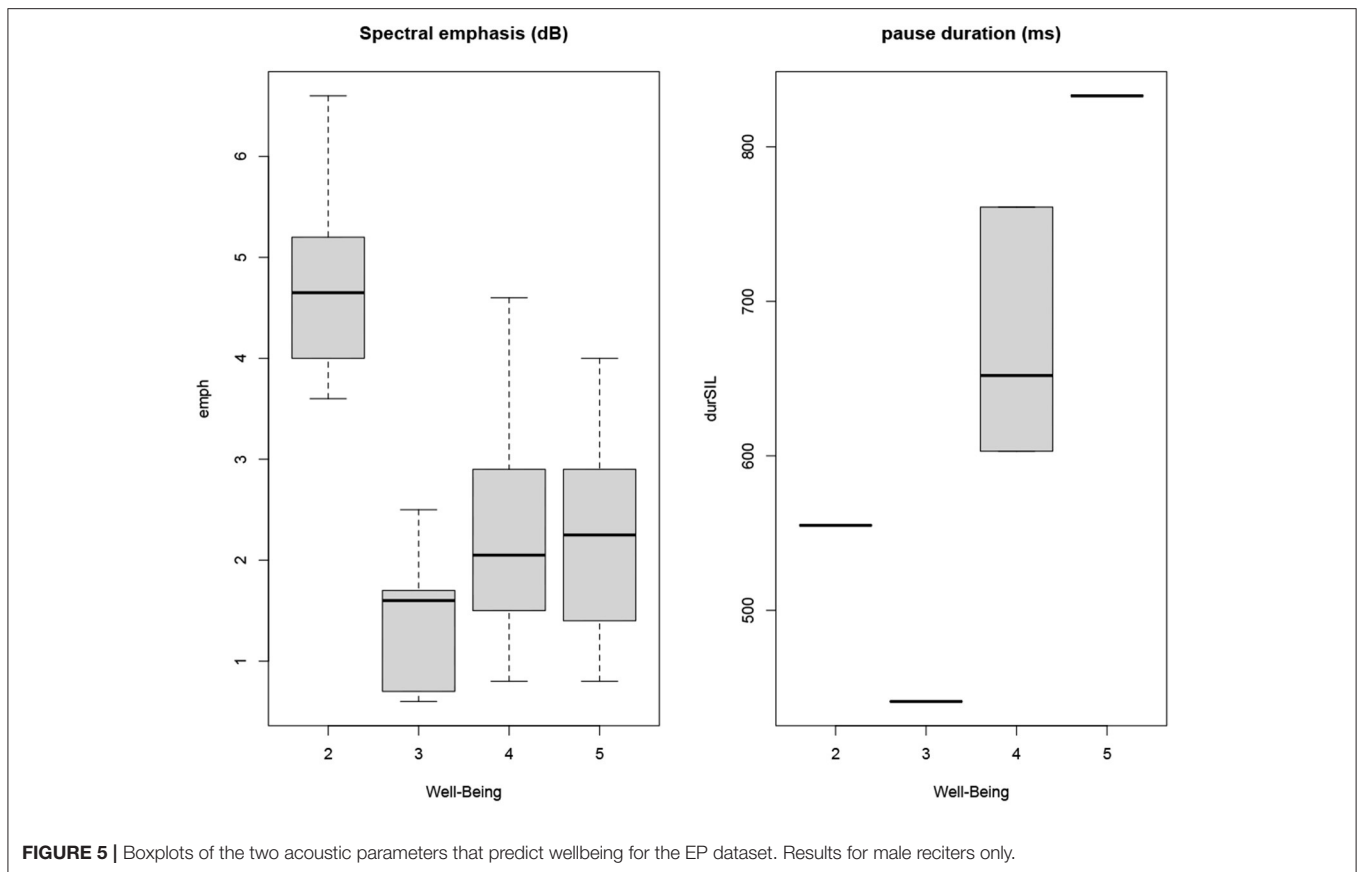
and explained by the same four parameters: rate of pause (the lower, the higher the wellbeing), pause duration (the longer, the higher the wellbeing), the slope of LTAS (the faster the slope, the higher the wellbeing) and F0 median (the higher, the higher the wellbeing). **Figure 4** shows the boxplots for the parameters explaining wellbeing evaluation of the male recitations due to its higher explained variance. The same tendencies found for pleasantness prediction can be seen, although less gradual. Degree 3 is clearly the one that deviates from a linear trend for degrees 2, 4, and 5 in all four plots, and certainly contributed to the positive contribution of the F0 median to explain wellbeing (the higher, the higher the wellbeing).

As for the degree of wellbeing attributed to Portuguese female reciters by the Portuguese listeners, the same 5% of explained variance of the BP dataset evaluation by female listeners was found, but associated to different parameters: F0med (the higher, the higher the wellbeing) and pause duration (the longer, the higher the wellbeing). As for results for male reciters given by the

same listeners, two parameters in order of importance explain the data: pause duration (the longer, the higher the wellbeing) and spectral emphasis (the lower, the higher the wellbeing), which explains 32% of the variance. **Figure 5** shows the boxplots for the parameters explaining wellbeing attributions for the male recitations only because it is the model with a higher explained variance where lower values for spectral emphasis and higher values for pause duration are associated with higher degrees of wellbeing.

The results presented above signal, as far as female reciters are concerned, that the relation of wellbeing to acoustic parameters is weaker than that for pleasantness. For male reciters, the relation of both scales to acoustic parameters is very similar, though. This partially confirms the second part of Hypothesis 1.

As for the discrimination results, the overall correlation between the responses of Brazilian and Portuguese listeners was 78%. When the variety of the reciter was BP, their correlation was 82 and 76% when the variety of the reciter was EP. When



both recitations of a chunk were from male speakers, the two groups of listeners correlated 78%, in comparison with 83% when they evaluated female reciters only and 76% for pairs containing mixed-gender reciters. Prediction for first or second preferred chunk in terms of pleasantness got the same proportion of hits in the LDA model: 75%.

The overall prediction of the listeners' choices is explained by four parameters in order of importance in the final model: articulation rate (faster rates are preferred), F0 falls (smoother slopes are preferred), spectral emphasis (lower values, meaning softer readings are preferred), and having a lower effect, the slope of LTAS (sharper slopes are preferred, which is related to less effort). If the responses are separated according to the variety of the reciter, only the first three parameters of the overall model are significant in different orders of importance: the same of the overall model when Brazilian reciters were evaluated, with 88% of hits in both directions, and the following order when Portuguese reciters were evaluated: spectral emphasis (lower values are preferred), F0 falls (smoother slopes are preferred), and articulation rate (faster rates are preferred). The preference for the first reading was predicted with 78% of hits and 75% of hits for the second reading choice, signaling a balance of preference between the two positions.

The median response by both Portuguese and Brazilian listeners for the chunk read by the Brazilian actor was for his reading in 6 out of 6 comparisons with lay reciters, whereas the

median response for the chunk read by the Portuguese actor was for his reading in 5 out of 6 comparisons with lay reciters. A Portuguese female, the lay reciter was preferred over the Portuguese actor in this single exception. This confirms an overall preference for the actors in the discrimination test as well. All statistical results are summarized in **Table 1**.

For the sake of examining the relationship between the acoustic parameters and the preferences for particular reciters, **Table 2** gives the mean values for the significant predictors considering into account the models involving Portuguese reciters. It can be seen that male reciters M5 and PL (the actor), who were more pleasant, have lower values for spectral emphasis and F0 median and higher pause duration means, which is in accordance with the logistic regression model. Female reciter F2 has the highest pause duration mean, and was also one of the females with high scores for pleasantness.

Although the LDA model evaluates a particular chunk, it can be seen that higher articulation rates and lower spectral emphasis are associated with higher scores of pleasantness.

Table 3, on the other hand, gives the mean values for the significant predictors considering into account the models involving Brazilian reciters. It can be seen that for male reciters M5 and IL (the actor), the ones who were more pleasant, have low rates of pause, the longest pauses among males, and fast LTAS slopes, which is in accordance with the prediction of the corresponding logistic regression model. Female reciter F2 has

TABLE 1 | Statistical models' summary. "Log. reg." stands for "logistic regression," and "pleasant." stands for "pleasantness."

Model type	Dataset	Language	Reciters' gender	Var. predicted	Model fitting	Predictor(s)
Log. reg.	Poem	BP	Female	Pleasant.	42%	Pause rate > LTAS slope > pause duration
Log. reg.	Poem	BP	Male	Pleasant.	23%	Pause rate > LTAS slope > pause duration > F0 median
Log. reg.	Poem	BP	Female	Wellbeing	5%	HNR > LTAS slope
Log. reg.	Poem	BP	Male	Wellbeing	27%	rate of pause > pause duration > LTAS slope > F0 median
Log. reg.	Poem	EP	Female	Pleasant.	44%	Pause duration
Log. reg.	Poem	EP	Male	Pleasant.	38%	Spectral emphasis > F0 median > pause duration
Log. reg.	Poem	EP	Female	Wellbeing	5%	F0 median > pause duration
Log. reg.	Poem	EP	Male	Wellbeing	32%	Pause duration > spectral emphasis
LDA	Paired chunks	EP & BP	Male & female	Pleasant.	88%	Articulation rate > F0 fall > spectral emphasis > LTAS slope

Refer to Section Acoustic parameters for an explanation of acoustic parameters.

TABLE 2 | Mean values for significant predictors for Portuguese reciters (gender indicated by the first letter: F, female; M, male).

	F0 median (Hz)	F0 fall (Hz/frame)	Pause duration (ms)	1/pause rate (s)	Articulation rate (syll/s)	HNR (dB)	SE (dB)	LTAS slope (dB/100 Hz)	Pleasant
F1	210	-4.6	384	1.7	6.7	13.0	5.5	-13.5	3
F2	185	-3.2	564	2.2	6.1	9.3	5.8	-9.3	4
F3	185	-3.8	389	2.5	6.3	11.9	4.9	-13.1	3
F4	176	-4.2	457	1.9	7.3	9.5	3.8	-14.0	4
F5	191	-3.5	464	2.2	6.3	14.0	2.4	-17.4	4
M1	118	-3.1	331	1.6	7.5	7.4	2.9	-14.2	3
M2	120	-1.8	451	1.7	7.4	7.4	2.8	-16.2	3
M3	133	-1.7	532	2.3	6.0	7.6	6.2	-9.5	2
M4	139	-1.8	398	1.8	6.5	10.4	3.2	-11.8	2
M5	100	-2.2	722	2.2	6.7	5.7	2.3	-16.0	4
PL	86	-0.8	1,232	2.9	8.1	1.8	3.0	-11.9	5

PL is the Portuguese actor.

TABLE 3 | Mean values for significant predictors for Brazilian reciters (gender indicated by the first letter: F, female; M, male).

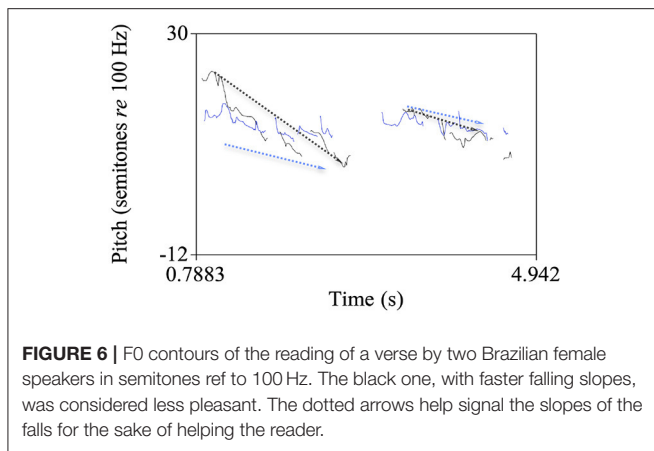
	F0 median (Hz)	F0 fall (Hz/frame)	Pause duration (ms)	1/pause rate (s)	Articulation rate (syll/s)	HNR (dB)	SE (dB)	LTAS slope (dB/100 Hz)	Pleasant
F1	237	-4.5	616	2.5	6.1	12.2	1.6	-20.6	3
F2	182	-2.0	683	2.1	7.2	11.0	1.9	-16.0	4
F3	212	-3.0	553	2.6	5.1	13.4	4.9	-14.7	3
F4	177	-2.7	670	2.5	6.5	12.7	1.6	-15.0	3
F5	212	-4.5	494	2.3	5.5	8.8	4.4	-9.8	2
M1	115	-2.4	761	2.7	6.1	7.5	1.6	-18.6	4
M2	107	-1.5	603	2.1	6.1	7.4	2.6	-14.1	4
M3	141	-2.8	555	2.8	5.7	6.7	4.8	-6.8	3
M4	175	-3.6	441	2.1	5.5	12.5	1.4	-19.0	3
M5	127	-3.3	652	2.4	5.5	7.5	2.4	-16.9	5
IL	101	-1.9	833	2.5	5.8	6.5	2.3	-19.9	5

IL is the Brazilian actor.

the highest pause duration mean, and one of the fastest LTAS slopes.

According to the LDA model, it can be seen that higher articulation rates and lower spectral emphasis tend to have higher scores of chunk-size pleasantness (final column).

For the sake of examining the consequences for the perception of less-known prosodic parameters, **Figure 6** presents an illustration of the F0 contours for the reading of a one-verse chunk by two Brazilian female speakers. The reading corresponding to the contour with sharper decreasing slopes



from the F0 peaks (black) was considered less pleasant. The direction of this choice happened 61% out of all choices.

If the direction of the differences in the degree of pleasantness attributed to the whole poem's reciter by the respective linguistic community is used as a prediction of the response in the discrimination test, the significant correlations between predicted-from-Likert-test and actual discrimination tests are 54% for Brazilian listeners and 63% for Portuguese listeners. For that correlation, one of the vectors was formed by the difference of the median degrees from the Likert-scale test considering the recitation from each chunk was extracted for building the pair and, the other vector, the responses of the discrimination test: 2, if the second reading is more pleasant than the first and 1, in the opposite case.

DISCUSSION

The most relevant parameters for predicting pleasantness in the evaluation of both the whole poem and the comparison of the respective chunk readings point to the classes of pausing and tempo (pause duration, pausing rate, and articulation rate), voice quality (spectral emphasis and LTAS slope) and melody (F0 falls), with a minor prediction power for F0 median in Brazilian and Portuguese male reciters, according to the models summarized in **Table 1**. On the other hand, the degree of pleasantness attributed to the recitation of the whole poem by Portuguese female reciters depends only on pause duration. It is also important to signal a word of caution as regards the models involving the evaluation of wellbeing caused by female recitations in both varieties of Portuguese due to the extremely low explained variance, only 5%, despite the significance of the predictors in the final models. Because the correlation of that evaluation with the one of pleasantness is superior to 75%, the reasons for the weak correspondence with acoustic parameters should be looked at something non-acoustic in the female recitations, such as the relation of some parameters with the content, for instance, a pause realized in a particular word boundary. This can only be disentangled by coding these aspects and including other poems in future research.

The prediction of wellbeing degree in the Likert-scale test is mainly explained by parameters related to pausing (duration and rate), voice quality (LTAS slope and HNR), and F0 median as far as Brazilian reciters are concerned. The evaluation of the wellbeing caused by Portuguese recitations concerns the same domains but is restricted to the parameters of pause duration, spectral emphasis, and F0 median. For this variety, the results of the evaluation of male recitations are more relevant, due to higher explained variance.

Actors' recitations were preferred, both in direct comparisons and in the whole poem recitation, where pause rate and duration are highly relevant, parameters for which the actors exhibit the longest median pauses. The Portuguese actor also has the lowest rate of pause, the highest articulation rate, and the smoothest F0 mean falls.

It is important to point out, as far as pleasantness and wellbeing are concerned, that the following parameters were not significant predictors in any case: time position of F0 local peaks, F0 peak rate, F0 mean peak bandwidth, rates of F0 rise mean, the coefficient of variation of global intensity (CVInt), jitter, and shimmer. As we have seen in the introduction, jitter is associated with pleasantness in German, and its role in Portuguese is found in EP as a parameter for predicting pleasantness, as seen in the work by Pinto-Coelho et al. (2013) where low values of this parameter and for shimmer are preferred. As for the other parameters, their low variation across the chunks, especially CVInt (with an intra-speaker/inter-speaker ratio varying below 1%), indicates that a higher variation would be more related to liveliness than to the scales measured here. The theme of the poem, the insignificance of the poet in comparison with the passage of the seasons, especially the return of Spring, does not favor a lively recitation.

As for the hypotheses raised in the Introduction, we have already seen that Hypothesis 1 (pleasantness and wellbeing appraisals are equivalent, but the second has a weaker relation to acoustics because it is more subjective than the first) and Hypothesis 11 (Actors' voices are more pleasant and cause more wellbeing than those of lay reciters) were confirmed. In the following, the other hypotheses will be examined.

Hypotheses 2 (lower mean F0 increases both the sensations of pleasantness and wellbeing for women listening to the recitation of men) and 3 (higher mean F0 increases both the sensations of pleasantness and wellbeing for men listening to the recitation of women) were not entirely confirmed because F0 median was significant only in the evaluation of male recitations for which higher-pitched men were preferred by the two Brazilian genders of listeners, whereas lower-pitched men were preferred by the two Portuguese genders of listeners. Because articulation rate was significant only in the case of paired chunks and in the direction of higher rates being preferred and the rate of pauses predicted less frequent pauses as being more pleasant, Hypothesis 4 (slowing down and more frequent pauses increase both pleasantness and wellbeing) was not confirmed. Hypothesis 5 (breathier speech increases pleasantness and wellbeing) was not confirmed because higher values HNR increases the sensation of wellbeing and not lower values, which are related to more roughness and breathiness. This is in relation to the work by

Madureira (2008) for BP, where the speaker with a whispery voice sounded sad. Hypothesis 6 (more variable F0 increases pleasantness and wellbeing) was not confirmed because F0SAQ was not a significant predictor. Hypothesis 7 (the lesser vocal effort, the higher the sensations of pleasantness and wellbeing is) is confirmed because the spectral emphasis was significant for Portuguese male reciters for both pleasantness and wellbeing, for which the lower values increase both sensations. The high correlations of responses between Brazilian and Portuguese listeners broke down according to the variety of the reciter disconfirms Hypothesis 8 (Portuguese listeners appreciate the declamation by Brazilians more than Brazilians appreciate the declamation by Portuguese speakers).

Hypotheses 9 and 10 were not confirmed due to the differences in the models. As for Hypothesis 9 (the acoustic parameters related to pleasantness and wellbeing are the same and in the same degree of relative importance in the two varieties of Portuguese), although there is a close correspondence between the acoustics of the reciters' varieties of Portuguese, meaning that rhythmic parameters such as pause duration and articulation rate are significant predictors, with F0 median added in the case of male reciters in both varieties and measures of vocal effort, namely spectral emphasis and LTAS slope, the order of importance and specific combination of predictors is not the same in the two varieties of Portuguese. And because the predictors change depending on the gender of the reciter, Hypothesis 10 (men and women use the same prosodic-acoustic parameters to make the speech better appreciated by the listener) is not confirmed.

One key component of the acoustics of pleasantness, in particular in the models with higher amounts of variance explained, seems to be pausing because recitations with longer pause durations and lower pausing rates are more pleasant when our participants listen to the whole poem declamation. In pair comparisons, higher articulation rates, another temporal parameter, got higher degrees of pleasantness, which is in accordance with the higher attractiveness of faster male speech judged by women in the literature reviewed in the Introduction. Our results show that, at least for pleasantness, higher articulation rates are preferred by both male and female listeners irrespective of the gender of the reciter.

As far as the training of professional speakers is concerned, our results suggest the following relation between the use of voice/speech and the increase of the aesthetical pleasure in poetry declamation as related to sound pleasant and causing an increase

of wellbeing in the listener: long pauses, more frequent pauses, higher articulation rates, lower vocal effort (related to sharper LTAS slopes and lower values of spectral emphasis).

CONCLUSIONS

The acoustics of pleasantness has some points in common with the acoustics of attractiveness when our results are compared with the findings reviewed in the Introduction mainly for English and German. One of these points is the role of tempo and vocal effort, as highlighted in the previous section. Despite minor differences in the evaluation of pleasantness depending on the variety of the reciter, the overall picture is quite the same, with the main role for tempo and voice quality through vocal effort acoustic correlates. The role of F0 falls in pair comparisons must be further investigated, but could also be related to slowing down, because smoother (slower) falls were considered more pleasant.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://figshare.com/authors/Plinio_Barbosa/11320902.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comitê de Ética em Pesquisa from the University of Campinas. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The author thank our reciters and listeners from the two sides of the Atlantic Ocean and the CNPq grant # 302194/2019-3. The author also in debt with research assistant Alice Crochiquia for her help with the online perception tests. This work was registered in the Ethics Committee under number 41315420.8.0000.8142.

REFERENCES

- Anderson, R. C., Klofstad, C. A., Mayew, W. J., and Venkatachalam, M. (2014). Vocal fry may undermine the success of young women in the labor market. *PLoS ONE* 9, e97506. doi: 10.1371/journal.pone.0097506
- Barbosa, P. A. (2020). *Prosody Descriptor Extractor*. Available online at: <https://github.com/pabarbosa/prosody-scripts/tree/master/ProsodyDescriptorExtractor>.
- Barbosa, P. A., and Eriksson, A., Åkesson, J. (2013). "Cross-linguistic similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese," in *Nordic Prosody XI, 2012 Tartu Nordic Prosody; Proceedings From the XIth Conference*. Frankfurt am Main: Peter Lang.
- Baumann, T. (2017). "Large-scale speaker ranking from crowdsourced pairwise listener ratings," in *Proceedings of INTERSPEECH 2017*. (Stockholm), 2262–2266. doi: 10.21437/Interspeech.2017-1697
- Boersma, P., and Weenink, D. (2021). *Praat: Doing Phonetics by Computer*. Available online at: <http://www.praat.org/>
- Braga, D., Coelho, L., Resende, F. G. V. Jr., and Dias, M. S. (2007). "Subjective and objective evaluation of Brazilian Portuguese TTS voice font quality,"

- in *Advances in Speech Technology, 14th International Workshop, June 27–29 2007* (Maribor).
- Burkhardt, F., Eckert, M., Johannsen, W., and Stegmann, J. (2010). “A database of age and gender annotated telephone speech,” in *Proceedings of LREC 2010* (Valletta), 1562–1565.
- Camargo, Z., Madureira, S., Reis, N., and Rilliard, A. (2019). “The phonetic approach of voice qualities: challenges in corresponding perceptual to acoustic descriptions,” in *Subsidia: Tools and Resources for Speech Sciences*, eds J. M. Lahoz-Bengoechea, and R. P. Ram-Åsn (Malaga: University of Malaga), 11–17.
- Carlsen, J. M., Franke, M., Huttner, L. M., and Radtke, A. (2018). “How to measure a pleasant voice,” in *Proceedings of the Conference Phonetics and Phonology in the German Language Area (P&P14)* (Vienna), 22–26.
- de Krom, G. (1994). “Spectral correlates of breathiness and roughness for different types of vowel fragments,” in *Proceeding 3rd International Conference on Spoken Language Processing (ICSLP 1994)*. (Yokohama), 1471–1474.
- Eriksson, A., Barbosa, P. A., and Åkesson, J. (2013). “The acoustics of word stress in Swedish: a function of stress level, speaking style and word accent,” in *Proceedings of the 14th Annual Conference of the International Speech Communication Association, Lyon (Interspeech 2013)* (London: Causal Productions), 778–781. doi: 10.21437/Interspeech.2013-226
- Eskénazi, M. (1993). “Trends in speaking styles research,” in *Proceedings of the Third Eurospeech* (Berlin), 501–509.
- Fónagy, I. (1961). Communication in poetry. *Word* 17, 194–218. doi: 10.1080/00437956.1961.11659754
- Fónagy, I. (2001). *Languages Within Language: An Evolutive Approach*. Philadelphia, PA: John Benjamins Publishing. doi: 10.1075/fof.13
- Hodges-Simeon, C. R., Gaulin, S. J., and Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Hum. Nat.* 21, 406–427. doi: 10.1007/s12110-010-9101-5
- Jacobsen, T. (2010). Beauty and the brain: culture, history and individual differences in aesthetic appreciation. *J. Anat.* 216, 184–191. doi: 10.1111/j.1469-7580.2009.01164.x
- Khan, S. U. D., Becker, K., and Zimman, L. (2015). Acoustic correlates of creaky voice in English. *J. Acoust. Soc. Am.* 137, 2267–2267. doi: 10.1121/1.4920276
- Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.
- Léon, P. (1993). *Précis de Phonostylistique*. Paris: Nathan, p. 157–184.
- Löfqvist, A., and Mandersson, B. (1987). Long-Time average spectrum of speech and voice analysis. *Folia Phoniatr. Logopaed.* 39, 221–229. doi: 10.1159/000265863
- Madureira, S. (2008). “Reciting a sonnet: production strategies and perceptual effects,” in *Proceeding Speech Prosody 2008*. (Campinas), 697–700.
- Madureira, S., and Camargo, Z. (2010). “Exploring sound symbolism in the investigation of speech expressivity,” in *Proceedings of the 3rd ISCA Tutorial and Research Workshop on Experimental Linguistics (ExLing 2010)*, ed A. Botinis (Athens: ISCA and University of Athens), 105–108.
- Madureira, S., and Fontes, M. A. S. (2019). “The analysis of facial and speech expressivity: tools and methods,” in *Subsidia: Tools and Resources for Speech Sciences*, eds J. M. Lahoz-Bengoechea, R. P. Ram-Åsn (Malaga: University of Malaga), 19–26.
- Master, S., Biase, N. D., Pedrosa, V., and Chiari, B. M. (2006). The long-term average spectrum in research and in the clinical practice of speech therapists. *Pró Fono Rev. Atualiz. Cientf.* 18, 111–120. doi: 10.1590/S0104-56872006000100013
- Mendoza, E., Valencia, N., Muñoz, J., and Trujillo, H. (1996). Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). *J. Voice* 10, 59–66. doi: 10.1016/S0892-1997(96)80019-1
- Menegon, P. S., Madureira, S., and Fontes, M. (2021). Análise de interpretações orais de um poema quanto aos aspectos expressivos. *Intercâmbio* 49, 113–137. doi: 10.23925/2237.759X.2021V49.56747
- Niebuhr, O., Thumm, J., and Michalsky, J. (2018). “Shapes and timing in charismatic speech—Evidence from sounds and melodies,” in *Proceeding 9th International Conference of Speech Prosody* (Poznań). doi: 10.21437/SpeechProsody.2018-118
- Pinto-Coelho, L., Braga, D., Sales-Dias, M., and Garcia-Mateo, C. (2013). On the development of an automatic voice pleasantness classification and intensity estimation system. *Comp. Speech Lang.* 27, 75–88. doi: 10.1016/j.csl.2012.01.006
- Quené, H. (2020). “Attractiveness of male speakers: effects of pitch and tempo,” in *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speaker*, eds B. B. Weiss, J. Trouvain, M. Barkat-Defradas, and J. J. Ohala (Singapore: Springer), 153–164. doi: 10.1007/978-981-15-6627-1_9
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <http://www.r-project.org>
- Roback, P., and Legler, J. (2021). *Beyond Multiple Linear Regression: Applied Generalized Linear Models and Multilevel Models in R*. London: Chapman and Hall/CRC. doi: 10.1201/9780429066665
- Rosenberg, A., and Hirschberg, J. (2020). “Prosodic aspects of the attractive voice,” in *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speaker*, eds B. B. Weiss, J. Trouvain, M. Barkat-Defradas, and J. J. Ohala (Singapore: Springer), 17–40. doi: 10.1007/978-981-15-6627-1_2
- Severin, F., Bozkurt, B., and Dutoit, T. (2005). “HNR extraction in voiced speech, oriented towards voice quality analysis,” in *2005 13th European Signal Processing Conference* (Antalya: IEEE).
- Stathopoulos, E. T., Huber, J. E., and Sussman, J. E. (2011). Changes in acoustic characteristics of the voice across the life span: measures from individuals 4–93 years of age. *J. Speech Lang. Hear. Res.* 54, 1011–1021. doi: 10.1044/1092-4388(2010)10-0036
- Strangert, E., and Gustafson, J. (2008). “What makes a good speaker? Subject ratings, acoustic measurements and perceptual evaluations,” in *Proceedings of INTERSPEECH 2008*, 1688–1691. doi: 10.21437/Interspeech.2008-368
- Sundberg, J. (1994). Perceptual aspects of singing. *J. Voice* 8, 106–122. doi: 10.1016/S0892-1997(05)80303-0
- Tegnér, I., Fox, J., Philipp, R., and Thorne, P. (2009). Evaluating the use of poetry to improve well-being and emotional resilience in cancer patients. *J. Poetry Ther.* 22, 121–131. doi: 10.1080/08893670903198383
- Traunmüller, H., and Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *J. Acoust. Soc. Am.* 107, 3438–3451. doi: 10.1121/1.429414
- Verdonck-de Leeuw, I. M., Festen, J. M., and Mahieu, H. F. (2001). Deviant vocal fold vibration as observed during videokymography: the effect on voice quality. *J. Voice* 15, 313–322. doi: 10.1016/S0892-1997(01)00033-9
- Weiss, B., and Burkhardt, F. (2010). “Voice attributes affecting likability perception,” in *Proceedings of INTERSPEECH 2010* (Chiba), 2014–2017. doi: 10.21437/Interspeech.2010-570
- Weiss, B., and Burkhardt, F. (2012). “Is ‘not bad’ good enough? Aspects of unknown voices likability,” in *Proceedings of INTERSPEECH 2012*, 510–513. doi: 10.21437/Interspeech.2012-97
- Weiss, B., Trouvain, J., and Burkhardt, F. (2020). “Acoustic correlates of likable speakers in the NSC database,” in *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speaker*, eds B. B. Weiss, J. Trouvain, M. Barkat-Defradas, and J. J. Ohala (Singapore: Springer), 245–262. doi: 10.1007/978-981-15-6627-1_13

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Barbosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.