**Tuuli Nilsson**
Tampere University

**Anne-Maria Laukkanen**
Tampere University

**Tiina Syrjä**
Tampere University

**Suely Master**
Universidade Estadual Paulista (UNESP)

# Finnish and Brazilian-Portuguese listeners' evaluation of voice characteristics in Brazilian-Portuguese speaking actresses and non-actresses

## Abstract

*Auditory-perceptual evaluation is an important tool in voice training and voice therapy. International comparisons have been made for clinical evaluation tools such as GRBAS. Similar comparisons have not been made in the field of voice training. This study compares Finnish and Brazilian-Portuguese listeners' evaluation of voice characteristics in Brazilian-Portuguese speaking actresses and non-actresses. Thirty experienced actresses and 30 non-actresses, all native speakers of Brazilian-Portuguese, were recorded for a 200-word reading task in habitual conversational loudness. Six perceptual voice characteristics were rated, including overall voice quality, type of voice production (from breathy to strained), sonority, suitability of average pitch for the speaker, pitch range and nasality. The inter-rater and intra-rater reliabilities were good or moderate in both Brazilian and Finnish raters, except for nasality. Finnish and Brazilian raters' evaluations correlated strongly in voice quality, suitability of pitch and range (rho = 0.76, 0.62 and 0.82, respectively, p = 0.000). Moderate correlation was found for voice production (rho = 0.51, p = 0.000) and sonority (rho = 0.59, p = 0.000). Evaluations of nasality did not correlate. Small but significant differences were obtained in the mean evaluations, for example, Finns rated the pitch range higher (Bra =4.92 vs. Fin = 5.56, T-*

*test, $p < 0.001$). Both groups evaluated actresses' voices more positively. The results seem to suggest that voice experts share some international standards in voice quality evaluations, although ratings of some aspects may differ. This finding is in line with the cross-cultural comparisons of the evaluation of dysphonic voices.*

## Keywords

auditory-perceptual evaluation; cross-cultural; linguistic/cultural background; non-actresses; normal/supranormal voice; overall voice quality; voice training

## Introduction

The perceptual evaluation of voice and speech characteristics is an important tool in voice and speech training and therapy. Goals for training are set and the effects of training are monitored based on perceptual evaluation. In general, the basic goals of voice training seem very international, as many books about voice training published in different countries contain very similar exercises and describe the aims of exercising in the same way (Cooper 1973; Lessac 1967; Proctor 1968; Perkins 1971; Bauer 1973; Forsmark 1977; Aalto and Parviainen 1985; Rodenburg 1993; Berry 1995; DeVore and Cookman 2009; Shewell 2013; Richter 2014; Aderhold 2021). The main goal is a voice that gives the maximum acoustic output (and thus audibility) with minimal effort (Perkins 1971; DeVore and Cookman 2009; Shewell 2013), is flexible in variation to fulfil the communicative needs and endures the individual and occupational vocal challenges, thus helping to avoid the development of voice disorders due to vocal overloading.

Most research on the perceptual evaluation of voice has been conducted in a clinical context and several perceptual evaluation systems have been developed to grade the level of dysphonia and to characterize the type of voice deviation from normophonia. Grade, Roughness, Breathiness, Asthenia and Strain (GRBAS), as specified by Hirano (1989) and later modified into GIRBAS by including instability in the repertory (see, e.g., Dejonckere et al. 1998), and the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V), described by Hillman (2003) and Kempster et al. (2009), are two of the most widely recognized and studied rating protocols. As languages and cultures differ in the use of laryngeal and articulatory possibilities in speech production and vocal communication, it may also interfere with perceptual rating.

There has been a growing interest in cross-linguistic studies of pathological voice qualities (Anders et al. 1988; Doty 1998; Hartelius et al. 2003; Yamaguchi et al. 2003; Yiu et al. 2008). A cross-language study by Anders et al. (1988) studied the effects of professional backgrounds and culture (language) on the perception of hoarseness. The ratings of the control group (unprofessional listeners) mirrored those of the professionals, although with

a somewhat lower level of accuracy. No significant differences between the classes of listeners were found, yet the study indicated that US listeners rated unfamiliar German dysphonic stimuli milder than German and Finnish listeners. German listeners were most accurate in their judgements. The Finnish participants presented lower scores in identifying normal phonation and higher scores in detecting the presence and severity of hoarseness.

A study by Yiu et al. (2008) considered cross-cultural differences in the perception of breathy and rough voice qualities in Cantonese and English synthesized voice signals that were analysed by 40 speech pathology students from Australia and Hong Kong. The results did not support the hypothesis that unfamiliar languages would be rated less severely, but rather the opposite seemed to be true: Australian listeners rated English stimuli less severely than Cantonese stimuli, and Hong Kong listeners rated Cantonese stimuli less severely than English stimuli. However, the results were not consistent across male and female voices and for both rough and breathy qualities, so some synthesis-related matters may have affected the results (Yiu et al. 2008).

A comparison between Japanese and American raters indicated that these listeners exhibit similar rating patterns across the Grade, Roughness and Breathiness scales and showed differences only in the evaluation of asthenia and strain (Yamaguchi et al. 2003). In a study by Ghio et al. (2011), French and Italian dysphonic speakers were assessed by French and Italian raters. Both groups of speakers were perceived similarly in terms of overall severity and breathiness by the listeners from each respective country. Only the roughness rating showed a noticeable difference with Italian listeners generally underestimating roughness compared to their French counterparts. An investigation by Englert et al. (2022) involved Brazilian and European listeners evaluating Brazilian-Portuguese and German voice samples. The study showed a decrease in reliability among raters when assessing samples in a non-native language. Moreover, native raters demonstrated higher diagnostic accuracy in identifying voice deviations, indicating their superior ability to evaluate voice samples in their native language, likely due to their greater familiarity and training in that linguistic context (Englert et al. 2022).

The perceptual evaluation of normophonic voices has been studied far less than the above, and no generally used rating tools exist; rather, each study has applied custom-made protocols suited for the particular aim of the study. When investigating the effects of vocal loading, the protocol often uses a modification of GRBAS (e.g. Jónsdottir et al. 2003), since vocal overloading may cause such deterioration of the voice quality that it resembles symptoms of dysphonia. Several custom-made ratings have been applied to study the effects of voice training. These ratings reflect the goals of training. Ilomaki et al. (2008) and Leppänen et al. (2009) rated overall voice quality and firmness of phonation (from breathy to pressed) as perceptual outcome parameters to evaluate the effects of voice training and manual treatment for teachers. The basic idea was that training results in optimal firmness, so-called phonation balance, which is defined as having adequate adduction between the vocal folds in relation to the subglottic air pressure used (Laukkanen 2003). Phonation balance ensures optimal vibration of the vocal folds, which in turn enhances the resonance and projection of the sound produced in the vocal tract. Resonant voice is described as projecting well, being easy to produce and involving a sensation of facial vibration (Verdolini et al. 1998; Titze 2001; Bele 2002). It is recognized by its rich harmonic content and is a central objective

in voice training for actors and singers (Verdolini et al. 1998; Titze 2001; Bele 2002).

Overall voice quality is seen as a construct of various parameters like firmness, timbre, pitch and loudness. Bele et al. (2010) rated overall quality, firmness and timbre (from dark to bright) when studying the effects of two-month voice training for Norwegian radio journalists. Master et al. (2008) rated projection (carrying power), loudness and tension when comparing the voice quality of Brazilian male actors' and non-actors' voices. When studying the voice and speech of actors or the effects of actors' long and manifold voice training, a wider parameter variety may be justified. Bele compared Norwegian male actors' and teachers' voices and aimed to find a reliable perceptual protocol for rating normophonic speakers (Bele 2002). Eleven parameters out of 16 gained satisfactory inter-rater reliability. These parameters were Overall Voice Quality, pitch range, sonority, variation in loudness, clarity of articulation, breathiness, hyperfunctional/pressed voice production, intermittent vocal fry, roughness/grating and ringing voice quality. Bele's (2002, 2007) investigation suggests that sonority, loudness variation and range are separately perceptible qualities. Furthermore, there was also strong correlation between sonority, loudness variation, voice range and articulation clarity as part of the Overall Voice Quality, and individuals with sonorous voices were using more pitch and loudness variation (Bele 2007). Additionally, roughness and breathiness are recognized as distinct characteristics, as are pressed voice and the strength of the voice in terms of acoustic energy (Bele 2007). A modified version of the protocol was applied by Nilsson (Nilsson et al. 2022). In her study, colour of voice (as in dark/bright timber), suitability of pitch for the speaker and audibility were included in the protocol. As language and culture affect perceptual rating, it is plausible to suggest that the more speech-related parameters there are in the rating tool, the more susceptible for cultural differences it may be.

Languages have different phonemes and phonology. Also, languages may differ from each other in terms of vocal characteristics, such as fundamental frequency and voice quality (Keating et al. 2023). Pitch (fundamental frequency) and its use can be language and culture dependent (see studies including Pegoraro Krook 1988; Ohara 1992, 1999; Lewis 2002; Mennen Schaeffler and Docherty 2007). Andreeva et al. (2014) found that German and English speakers use lower pitch maxima, narrower pitch span and generally less variable pitch than Bulgarian and Polish speakers. A study by Mennen et al. (2007) reported a broader pitch range among female speakers of Southern Standard British English in comparison to female speakers of Northern Standard German. It is also common to describe Russians as having a broader pitch range and speaking with higher tones compared to Finns, whose speech is often perceived as lacking variation (Ullakonoja 2007). Pitch plays a crucial role in conveying lexical tone, intonation and stress (Gordon and Applebaum 2010; Best 2019). For instance, tonal languages like Mandarin Chinese, Thai and Yoruba use it to differentiate word meanings. In non-tonal languages, pitch changes may indicate functions like grammar, emotion and emphasis, shaping speech rhythm and melody (Chan and Chang 2019). However, pitch changes and their importance in signalling affect can vary between languages. For example, according to Yanushevskaya et al. (2018), in Japanese and Spanish the F0 contour may play a more important role, and a tense voice plays a lesser role in affect signalling than in Irish English and Russian.

Previous studies have shown some mutual tendencies between languages, for example, high or rising pitch in questions, and low or descending pitch in

nonquestions and statements (Ohala 1983). In Finnish, however, intonation is typically descending, and questions are not marked with rising intonation. Politeness is in many languages signalled with higher pitch while assertiveness with lower. It is even more common both in language and culturally that vocalizations indicating diminutiveness and submissiveness are expressed with a higher pitch, and largeness and threats are expressed with lower pitch; this can even be found between species (Ohala 1983).

Variation in voice quality may have a significant grammatical and phonetical role in languages. Voice quality also provides information about the speaker, participates in the construction of stance in interaction and serves to identify the individuality of the speaker (Podesva and Callier 2015).

In some languages, phonation types such as creaky, breathy and harsh – the same that are evaluated in pathological voices – are used for phonemic or allophonic distinctions, thus affecting linguistic information (Gordon and Ladefoged 2001; Keating et al. 2023). For instance, Jalapa Mazatec utilizes breathy and creaky phonation to alter word meanings (Garellek and Keating 2011). Many languages, like English, use non-modal phonation types for prosodic or sociolinguistic purposes such as emphasis or mood (Ladefoged and Maddieson 1998; Keating et al. 2023). In certain prosodic contexts, nonmodal phonation types are employed to achieve communication goal or grammatical nuances.

Wagner and Braun (2003) have reported intercultural and interlanguage differences in voice quality. In their study, F0, HNR (harmonics-to-noise ratio) and shimmer (period-to-period variation in voice amplitude, one correlate of hoarseness) varied between speakers of German, Italian and Polish, which distinguished these speaker groups from each other. They suggest that the higher F0 and harmonics-to-noise ratio contribute to the impression of brightness in Polish speakers' voices, and high shimmer values contribute to the impression of a rougher voice quality in Italian speakers.

Studies have shown that voice quality can act, for example, as a marker of ethnicity and regionality (Szakay 2012; Loakes and Gregory 2022). Szakay (2012) has found that Maori English speakers are creakier than European New Zealanders analysed by H1–H2 spectral tilt in vowels. Loakes and Gregory (2022) reported that for male speakers, pitch and voice quality, measured in H1–H2, differed significantly according to dialect, and for female speakers to location. Also, Szakay (2012) found that listeners are sensitive to phonation differences and are able to rely on phonation cues in the identification of speaker ethnicity.

Esposito (2010) found that Gujarati speakers, whose language contains voice quality-related phonemic contrasts, are better at distinguishing between breathy and modal phonation compared to English or Spanish speakers, whose languages lack distinct phonemic voice quality contrasts. In another study of Kreiman and Gerratt (2010), native speakers of Gujarati, Thai and English were observed to differ in their sensitivity to phonation changes, with Gujarati speakers showing higher sensitivity. The sensitivity of Thai and English speakers was lower and context-dependent due to post-lexical phonation variation (e.g. prosodical or sociolinguistic variation) (Kreiman and Gerratt 2010). In a study, Ding et al. (2017) compared Chinese and German listeners' preferences in German speech synthesis samples and discovered that both groups had a strong correlation and similar preference of voice quality (breathy voice). However, German listeners primarily focused on voice quality, while Chinese listeners were more influenced by pitch variations (Ding et al. 2017).

Additionally, cultures differ in what kind of voice quality is regarded as attractive.

Japanese and Chinese listeners favour higher fundamental frequency and less breathy voices, while Brazilians prefer more breathy voices (Erickson et al. 2020). The results by Waaramaa et al. (2021) suggest that Arabic listeners associated breathy voice quality with attractiveness, while Finnish-speaking listeners linked breathy voice with many negative personality traits (unpleasant, unbalanced, untrustworthy).

In summary, voice is a cultural construct, and, for example, ethnicity, gender and spoken language can affect the fundamental frequency and other acoustic parameters of a voice, but the degree of their influence can be difficult to access (Behlau and Murry 2011). As languages and cultures display differences in the use of voice characteristics and voice-related preferences, the inter-rater reliability in evaluating various voice features is supposed to decrease when assessing non-native samples. Furthermore, cultural background may also affect what kind of voice qualities are preferred. Voice quality serves as a significant marker of social identity and affiliation, with individuals adjusting it to fit social norms or express cultural identity within their linguistic communities (Olwage 2004).

The voice training world is increasingly globalized, with various speaking and actors' voice training methods as well as singing methods spread internationally. Yet, no generally accepted perceptual rating protocol exists, and as far as we know, no cross-cultural comparisons have been made in the field of aesthetic voice use. The question arises as to what extent the training objectives and thus the quality criteria would be the same. Therefore, there is strong incentive to discover how international voice experts' quality assessments of normal and supranormal voices are. Finding evidence of whether or not a listener's and speaker's linguistic and cultural background affects a voice professional's perceptual judgement of normal and supranormal (e.g. actors' and public speakers') voices would simplify the process of determining how best to use research data from geographically disparate subjects.

To the best of our knowledge, the present study is the first where inter-cultural comparison is made on normophonic voices including supranormal voices (trained actors' voices). The material is based on Master et al. (2012). Evaluations by Brazilian and Finnish speech experts are compared. This material was chosen since (1) it is large enough and systematically collected (i.e. it includes well-matched groups of speakers) and (2) Brazilian and Finnish languages and cultures show a lot of differences. Finnish belongs to the Uralic language group (Suomi et al. 2008), and Brazilian-Portuguese to the Romance languages (Barbosa and Albano 2004). Differences between Finnish and Brazilian-Portuguese are manifested both at segmental and suprasegmental levels, for example, in intonation and phonemes. In Finnish, neutral complete statements have a smoothly descending pitch contour; the first syllable is slightly above (or at) the middle of the speaker's habitual speaking voice range, followed by declination and the final fall (Suomi et al. 2008; Lennes 2010). Also, word stress in Finnish is mostly regular – emphasis always on the first syllable – and while some exceptions can be found, none affect grammatical meaning (Blankenship 2002). Differences in phonemes are vast (Barbosa and Albano 2004; Suomi et al. 2008). One major difference in respect to vowels, in addition to formant frequencies between the two languages, is that Brazilian-Portuguese uses nasalized vowels while Finnish does not (Barbosa and Albano 2004; Suomi et al. 2008). The Finnish language also has a quantity opposition,

a possibility for long and short variations in speech segments for practically all sounds (Järvikivi et al. 2010), while in Brazilian-Portuguese there is a quality opposition (Barbosa and Albano 2004). Brazilian-Portuguese vowels are characterized by four qualities referred to as 'open', 'closed', 'reduced' and 'nasal'. As in the case of the word *almoço* (/aɬˈmosu/), which can mean either 'meal' or 'I lunch', a word's definition may depend upon the vowel's pronunciation. In the case of *almoço*, the qualifying factor resides in the initial 'o', which can be pronounced 'open' or 'closed'. With respect to consonants, the major differences between Finnish and Brazilian-Portuguese are that the Finnish /h/ occurs in many positions in words, and its distribution is wide: it can be, for example, a glottal continuant or oral fricative, depending on the allophone. The glottal /h/ is not present in Brazilian-Portuguese, and Finnish lacks the dental affricates present in Brazilian-Portuguese.

## Methods

### *Participants*

Text reading samples were obtained from an earlier study (Master et al. 2012). Thirty experienced actresses, and 30 non-actresses were recruited. Participants with an age range of 18–50 years were native speakers of Brazilian-Portuguese. The actors had a minimum of five years of work experience in theatre, at least one year of formal training and no current or past record of voice disorders or complaints of voice problems. The non-actresses had neither used their voice professionally nor formally participated in any voice training and had no current voice disorders or voice complaints.

### *Reading test*

The participants read a text of 200 words in Brazilian-Portuguese. They were asked to simply 'read the text' in a habitual voice level. Continuous speech, in comparison with sustained vowels, which are often used in voice research, involves continuous modification and control of glottal and supraglottal mechanisms; it contains rapid voice onsets and offsets, shifts between consonants and vowels and variation in pitch, loudness and other prosodic features (Parsa and Jamieson 2001; Eadie and Baylor 2006). As this study investigates cross-linguistic affects, it was reasonable to use connected speech as test material.

The recordings were made in an acoustically treated booth, with an ambient noise level below 30 dB. For the recording, a Marantz PMD-671 Solid State Recorder (Marantz, NY) and a microphone headset AKH C420L (AKG, AUT) were used, positioned at a distance of 8 cm from the lips' contralateral commissure. Samples were recorded digitally at a sample rate of 44.1 kHz with 16 bit/sample quantization. All the subjects were aware and gave their consent that the recordings would be used for research purposes.

### *Raters*

The assessments were made by ten experienced voice experts (age range 35–55, all females) without any known impairment to hearing. Their professional experience ranged from a minimum of 3–5 years to over twenty years of working with and evaluating normal and supranormal voices. Five of the experts were Brazilian speech language pathologists and voice trainers, and five were Finnish vocologists and voice trainers (one of them was also a speech and language pathologist). Table 1 shows the background of experience in years.

*Table 1: Finnish and Brazilian voice experts' experience (in years) working with and evaluating voice disorders and working with and evaluating normal and supranormal voices.*

| Voice disorders (years) | Normal and supranormal voices (years) |
|---|---|
| No | 10–15 |
| 3–5 | Over 20 |
| Over 20 | Over 20 |
| No | 3–5 |
| No | 20 |
| 3–5 | 5–10 |
| 3–5 | 15–20 |
| 15–20 | 10–15 |
| 5–10 | 3–5 |
| 5–10 | 3–5 |

This difference between formal educational labels is attributed to the fact that in Brazil there is no specific education for speech trainers, while in Finland speech and language pathologists typically work only with voice patients and do not serve as coaches for voice professionals (e.g. actors.). All Brazilian listeners were native Brazilian-Portuguese speakers, and the Finnish listeners were native Finnish speakers. All had proficiency in at least one foreign language. None of the Finnish listeners spoke Brazilian-Portuguese.

## Rating scale and procedure

Voice samples were rated on six vocal parameters using a Likert scale from 0 to 10. The following characteristics were assessed: overall voice quality (0 = very poor, 10 = excellent), voice production (0 = hypofunctional/breathy, 10 = hyperfunctional/pressed, strained), sonority (0 = not sonorous, 10 = very sonorous), pitch (0 = too low, 10 = too high for the speaker), pitch range (0 = very narrow, 10 = very wide) and nasality (0 = very nasal, 10 = very de-nasal) (Table 2). These characteristics, except for nasality, were chosen from an earlier study assessing the evaluation criteria for normal and supranormal voices (Bele 2002, 2007).

The judges of the present study were instructed to use high-quality headphones while evaluating the voices. The order of samples was randomized. Ten samples were repeated in random order to assess intra-rater reliability.

## Statistical analyses

Statistical analyses were carried out using PASW Statistics 26.0 software for Windows/MacOS (SPSS, Inc., Chicago, IL). Intra-rater reliability was calculated

*Table 2: Short definitions of vocal characteristics used in this study (Bele 2002, 2007).*

| | |
|---|---|
| Sonority | A voice that is perceived by its resonant quality, indicating richness in overtones. |
| Hyperfunctionality | Perceived vocal strain, as if the vocal folds are excessively compressed during phonation, often requiring significant laryngeal effort. |
| Hypofunctionality | Perceived insufficient vocal fold adduction, produced with minimal laryngeal effort, resulting in a weak, slack and non-resonant voice. |
| Pitch | The auditory correlate by fundamental frequency (determined by the rate of vibration of the vocal folds in the larynx); perceived highness or lowness of a sound. |
| Pitch range | The auditory correlate of the variation in fundamental frequency during running speech. |
| Nasality | The quality of voice when airflow is directed through the nasal cavity during speech (Sweeney et al. 1996). |
| Overall voice quality | The holistic assessment of various attributes of a person's voice. |

using ICC, a two-way random model, and inter-rater reliability was computed using ICC, a two-way mixed model (Koo and Li 2016). ICC of 0.75 or above presents good reliability (Koo and Li 2016). The Spearman's rank order correlation coefficient (Spearman's rho) was used to measure the strength of the association between Brazilian and Finnish listeners' evaluations based on the mean values of each variable. The significance of differences in the rating of the six vocal parameters between Finnish and Brazilian raters was calculated using a paired samples T-test for parameters with normal distribution (result from the Kolmogorov-Smirnov test non-significant), and using the Wilcoxon signed rank test for the parameters that did not show a normal distribution. Additionally, differences between the actresses' and non-actresses' ratings in both groups of listeners were studied using the independent samples T-test and Mann-Whitney U-test. Significant difference was set at $p \leq 0.05$ in all analyses. Bonferroni correction was obtained by the calculation $0.05 / 6 \leq 0.008$.

## Results

### *Intra-rater reliability*

For intra-rater reliability or test-retest consistency, ten voices were rated twice by the listeners. Results for both groups of listeners are presented in Table 3. Finnish listeners' mean intra-rater reliability was good in voice quality (0.87), sonority (0.80) and range (0.80). The Finnish raters' evaluations of voice production (0.72), nasality (0.72) and pitch (0.74) received moderate ICC values. Similarly, the Brazilian-Portuguese listeners' overall values of ICC were good in voice quality (0.91), sonority (0.84) and range (0.86), and moderate in voice production (0.72), nasality (0.77) and pitch (0.76).

### *Inter-rater reliability*

The inter-rater reliability (ICC) was calculated for the five Finnish listeners' and five Brazilian-Portuguese listeners' ratings. The results are presented in Table 4. The Finnish listeners' overall inter-rater reliability ranged from moderate to good for five of the six parameters. The four parameters that obtained

good ICC values were voice quality (0.84), voice production (0.86), pitch (0.80) and range (0.81), whereas sonority obtained a moderate value (0.75). Nasality (0.35) received low reliability. When looking at the actresses and non-actresses separately, ICC values were mostly good (Table 5). Non-actresses' pitch (0.72), range (0.72) and sonority (0.56) obtained only moderate reliability, whereas evaluations of nasality in both actresses (0.33) and non-actresses (0.44) obtained poor (<0.50) inter-rater reliability.

The Brazilian-Portuguese listeners had good inter-rater reliability (between 0.75 and 0.9) in voice quality, voice production and sonority (Table 4). Pitch, range and nasality obtained moderate reliability (between 0.5 and 0.75).

*Table 3: Intraclass correlation coefficient for auditory-perceptual evaluation by Brazilian-Portuguese speaking and Finnish listeners based on ten repeated voice samples.*

| | **Finnish** | **Brazilian** |
|---|---|---|
| Voice quality | 0.87 | 0.91 |
| Voice production | 0.72 | 0.72 |
| Sonority | 0.80 | 0.84 |
| Pitch | 0.74 | 0.76 |
| Range | 0.80 | 0.86 |
| Nasality | 0.72 | 0.77 |

*Table 4: Inter-rater reliability coefficients (ICC) of six perceptual variables rated by five Finnish and five Brazilian-Portuguese listeners for all 60 samples.*

| | **Finnish** | **Brazilian** |
|---|---|---|
| Voice quality | 0.84 | 0.82 |
| Voice production | 0.86 | 0.80 |
| Sonority | 0.75 | 0.79 |
| Pitch | 0.80 | 0.60 |
| Range | 0.81 | 0.78 |
| Nasality | 0.35 | 0.67 |

*Table 5: Inter-rater reliability coefficients (ICC) of six perceptual variables rated by five Finnish and five Brazilian-Portuguese listeners separately for actresses (n = 30) and non-actresses (n = 30).*

| | **Finnish** | | **Brazilian** | |
|---|---|---|---|---|
| | **Actresses** | **Non-actresses** | **Actresses** | **Non-actresses** |
| Voice quality | 0.84 | 0.77 | 0.77 | 0.76 |
| Voice production | 0.83 | 0.87 | 0.81 | 0.78 |
| Sonority | 0.79 | 0.56 | 0.74 | 0.53 |
| Pitch | 0.83 | 0.72 | 0.65 | 0.54 |
| Pitch range | 0.77 | 0.72 | 0.78 | 0.50 |
| Nasality | 0.33 | 0.44 | 0.60 | 0.71 |

Actresses' voice quality, voice production and pitch range showed good reliability; other parameters reached moderate ICC values (between 0.5 and 0.75; Table 5). Non-actresses' ICC values were good for voice quality, voice production and nasality, while sonority, pitch and pitch range received moderate values (Table 5).

## Comparison between brazilian and finnish listeners' ratings

Table 6 shows Spearman's rho correlations between Brazilian and Finnish listeners' evaluations based on the mean values of each variable. The non-parametric correlation was from moderate to very strong in all characteristics other than nasality, where no correlation was found ($r = 0.16$, $p = 0.225$).

Tables 7–9 compare the Brazilian and Finnish raters' mean evaluations for the six voice characteristics. When all 60 samples are considered (Table 7), Finnish listeners perceived voices as significantly more sonorous (Bra = 4.82, Fin = 5.53) and having a larger pitch range (Bra = 4.92, Fin = 5.56). Figure 1 illustrates the distribution of ratings by Finnish and Brazilian listeners.

Table 8 presents a comparison of actresses' voices and Table 9 of non-actresses' voices. A significant difference in the evaluation was found for the actresses' voice quality, which was evaluated better by Brazilian-Portuguese listeners (Bra = 6.39, Fin = 5.88). Finnish listeners evaluated non-actresses' samples significantly more sonorous (Bra = 4.07, Fin = 5.15) and having firmer voice production (Bra = 4.62, Fin = 5.30). Furthermore, Finnish listeners perceived the pitch range to be larger in both actresses and non-actresses

*Table 6: Correlations between Finnish and Brazilian listeners' ratings. Spearman's rank order correlation coefficient.*

| | Spearman's rho | Sig. (two-tailed) |
|---|---|---|
| Voice quality | 0.76 | <.001 |
| Voice production | 0.51 | <.001 |
| Sonority | 0.59 | <.001 |
| Pitch | 0.62 | <.001 |
| Range | 0.82 | <.001 |
| Nasality | 0.16 | 0.225 |

*Table 7: Differences between the evaluations of all 60 Brazilian-Portuguese samples by Brazilian and Finnish listeners. Paired samples T-test for others, related samples Wilcoxon signed rank test for voice production and nasality.*

| | Voice quality | | Voice production | | Sonority | | Pitch | | Pitch range | | Nasality | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** |
| Mean | 5.64 | 5.43 | 4.86 | 5,03 | 4.82 | 5.53 | 5.28 | 5.37 | 4.92 | 5.56 | 5.01 | 4.83 |
| SD | 1.3 | 1.04 | 1.12 | 1,0 | 1.21 | 0.99 | 0.96 | 0.77 | 1.30 | 0.91 | 0.86 | 0.65 |
| Sig. | .06 | | .029 | | <.001 | | .389 | | <.001 | | .055 | |

Bonferroni $p \leq 0.008$.

*Table 8: Differences between the evaluations of 30 Brazilian actresses' samples by Brazilian and Finnish listeners. Paired samples T-test for pitch range, related samples Wilcoxon signed rank test for other characteristics.*

| | Voice quality | | Voice production | | Sonority | | Pitch | | Pitch range | | Nasality | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** |
| Mean | 6.39 | 5.88 | 5.09 | 4.77 | 5.52 | 5.87 | 5.17 | 5.19 | 5.70 | 6.01 | 5.07 | 4.92 |
| SD | 1.05 | 1.05 | 1.02 | 0.90 | 1.21 | 1.09 | 1.07 | 0.85 | 1.2 | 0.85 | 0.74 | 0.59 |
| Sig. | .002 | | .130 | | .146 | | .866 | | .000 | | .754 | |

Bonferroni $p < 0.008$.

*Table 9: Differences between the evaluations of 30 Brazilian non-actresses' samples by Brazilian and Finnish listeners. Related samples Wilcoxon signed rank test for voice production, paired samples T-test for other characteristics.*

| | Voice quality | | Voice production | | Sonority | | Pitch | | Pitch range | | Nasality | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** | **Bra** | **Fin** |
| Mean | 4.85 | 4.95 | 4.62 | 5,30 | 4.07 | 5.15 | 5.4 | 5.59 | 4.08 | 5.09 | 5.08 | 4.72 |
| SD | 0.85 | 0.81 | 1.09 | 1.07 | 0.66 | 0.70 | 0.83 | 0.69 | 0.79 | 0.72 | 0.99 | 0.72 |
| Sign. | .531 | | <.001 | | <.001 | | 0.206 | | <.001 | | .033 | |

Bonferroni $p < 0.008$.

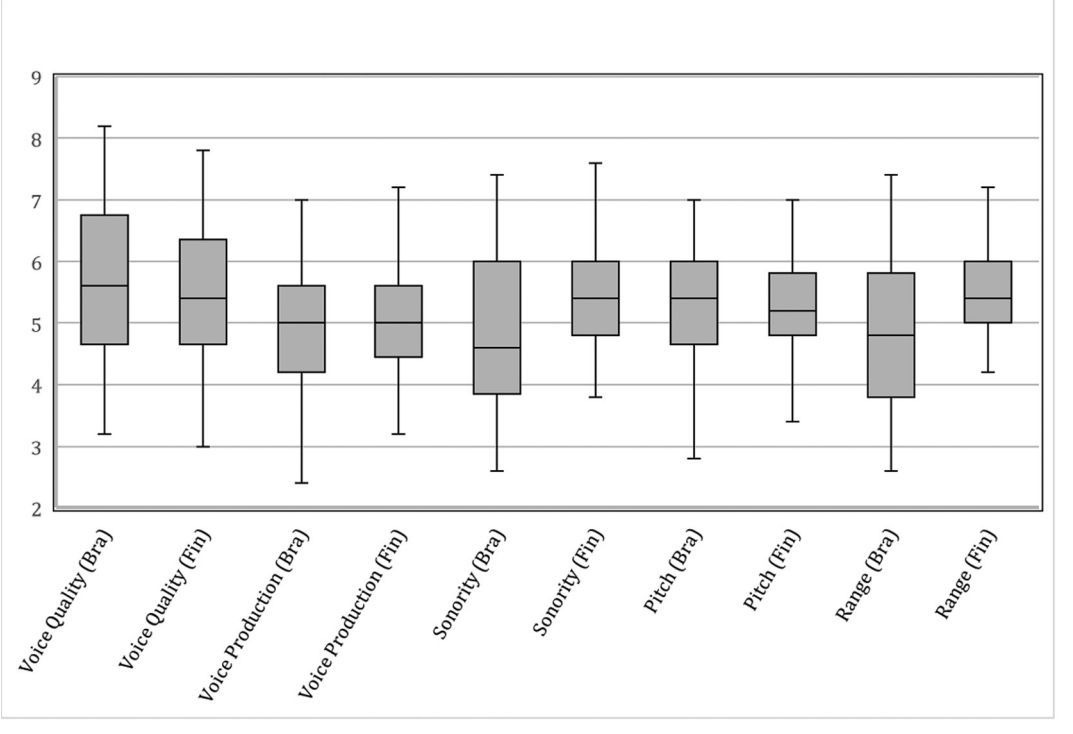


*Figure 1: A boxplot presentation of the evaluations by Finnish and Brazilian listeners for voice quality, voice production, sonority, pitch and pitch range (N = 60).*

*Table 10: Differences between actresses' and non-actresses' samples as evaluated by Finnish and Brazilian raters. Independent samples T-test for characteristics that showed normal distribution.*

| | Finnish | Brazilian |
|---|---|---|
| | Sig. (two-tailed) | |
| Voice quality | 0.000 | 0.000 |
| Sonority | 0.003 | 0.000 |
| Pitch | 0.150 | 0.578 |
| Pitch range | 0.000 | 0.000 |

compared to Brazilian listeners (Bra = 5.70, Fin = 6.01, and Bra = 4.08, Fin = 5.09, respectively). There were no significant differences in overall pitch and nasality ratings.

Table 10 shows differences between actresses and non-actresses as rated by Finns and Brazilians. Both groups evaluated the actresses' voices as better in quality, more sonorous and with a larger pitch range. Voice production and nasality did not differentiate the groups as rated by Finns (Mann-Whitney U-test, $p > 0.05$). A similar result was obtained in Brazilian ratings for nasality, but listeners rated the voice production of non-actresses as somewhat breathier (mean rating <5, $p = 0.08$, Mann-Whitney U-test).

## Discussion

The aim of this study was to examine whether raters from different cultural and language backgrounds assess normal/supranormal voices differently. We compared Finnish and Brazilian voice experts' ratings of Brazilian actresses' and non-actresses' text samples, read in Brazilian-Portuguese. The Finnish and Brazilian-Portuguese languages were chosen as they differ in many ways, for example, in pitch contour and intonation and phoneme repertoire (Barbosa and Albano 2004; Suomi et al. 2008).

Overall, the mean intra-rater results (Table 3) of all Finnish listeners and Brazilian-Portuguese listeners indicate that the raters were relatively consistent in their judgements of different voice characteristics. Based on the inter-rater reliability results (Table 4) a reasonable consensus occurred in both groups in the assessment of all 60 voice samples for voice quality, voice production, pitch and pitch range. Characteristics like voice quality, voice production and pitch have been shown to be among the most reliably rated in normal and supranormal voices (Bele 2005). The low inter-rater reliability for nasality among the Finnish raters ($p < 0.5$) suggests that the assessment of nasality caused special difficulties for Finnish-speaking listeners. This is most likely since the Finnish language does not contain nasal vowels. Thus, our results of this particular characteristic seem to support the hypothesis that unfamiliar voice characteristics would be more difficult to distinguish (Esposito 2010; Kreiman and Gerratt 2010; Englert et al. 2022).

Inter-rater reliability was better in both groups of listeners and for most voice characteristics when evaluating actresses as opposed to evaluating non-actresses. The reason may be that the actresses may use different voice characteristics to a larger extent. Furthermore, as suggested by Bele (2007), when there is significant variation in both loudness and pitch, the resonance

characteristics of the voice may become more noticeable. The evaluations by Finnish and Brazilian raters correlated significantly ($p < 0.001$) (see Table 6) in all other characteristics except nasality ($r = 0.16$, $p = 0.225$). Pitch range correlated very strongly ($r = 0.82$), voice quality and pitch strongly (0.76 and 0.62, respectively), and also voice production and sonority received moderate correlations between Finnish and Brazilian listeners (0.51 and 0.59). Overall, the correlations between the two groups of raters were stronger for actresses than for non-actresses (Tables 8 and 9). This again seems to suggest that for both groups of listeners it was easier to rate the actresses' samples.

Finnish raters evaluated a larger pitch range (Tables 7–9). This result is to be expected. It is plausible that it is difficult for a listener to evaluate the degree of normal or large pitch range in an unfamiliar linguistic context. Furthermore, compared to other languages, the intonation range in Finnish is relatively narrow and pitch intervals are small (Hakulinen 1979; Iivonen 1998). The voice quality of the actresses was rated somewhat lower by the Finnish raters (Bra = 6.39, Fin = 5.88). Non-actresses were evaluated as having firmer voice production (Bra = 4.62, Fin = 5.30) and also more sonority (Bra = 4.07, Fin= 5.15). Both groups of raters gave better evaluations for actresses in voice quality and sonority, and they also evaluated the pitch range larger in the actresses (Table 10). This is to be expected as actors are experienced voice users, typically with intensive voice training.

The overall tendency seemed to be that Finnish listeners rated Brazilian non-actresses' voice samples more favourably/less severely than Brazilian Portuguese speakers did. Some studies indicate that voice characteristics of less familiar voices are evaluated less severely (Anders et al. 1988). Das et al. (2020) suggest that when listeners encounter unfamiliar voices, they perceive voice quality not as separate attributes but as an overall pattern within the multidimensional vocal landscape.

However, that does not explain why the opposite occurred for the actresses, as Finnish listeners rated their voice quality as somewhat less good than the Brazilian listeners. The actresses clearly had more voice quality variation in their samples compared to the non-actresses. This large, expression-related variation might have made the voice quality rating more difficult, particularly in samples consisting of a foreign language the rater does not understand. It could be assumed, based on previous research, that Brazilian-Portuguese listeners had a natural advantage and were better able to distinguish and interpret different voice quality features while hearing a familiar language (see, e.g., Waaramaa et al. 2021; Englert et al. 2022). However, as voice perception is naturally affected by understanding and interpretation, not just hearing and evaluation (Brownell 2018), it could also be speculated that if one listens to what is said, they hear less clearly how it is said, which may affect the evaluations. Perhaps Brazilian-Portuguese just sounds more pleasant to Brazilian-Portuguese listeners. What is considered an aesthetically pleasant voice in the same social and cultural context may influence both the ways the voice is used and how it is being perceived (Olwage 2004).

Several studies have compared the auditory assessment of voice disorders with various standardized tools and found it to be quite consistent when the listeners are 'trained' practitioners (Chan and Yiu 2006; Kreiman et al. 2007; Iwarsson and Petersen 2012; Brinca et al. 2015). Specific consensus training will increase the inter-rater reliability even if the listeners are students and do not have years of clinical experience (Chan and Yiu 2006). However, certain cultural differences have been found in international comparisons.Yamaguchi

et al. (2003) compared GRBAS ratings by Japanese and American clinicians and students in speech pathology for a set of dysphonic voice samples. According to the results, no significant differences were found in three out of five characteristics. Cultural differences were found in the evaluations of asthenia and strain. In a study by Ghio et al. (2011), GRBAS ratings were compared between Italian and French clinicians. The ratings did not differ in the overall severity (G) and breathiness (B), while Italians underestimated the roughness (R) of the samples.

In the evaluation of normal and supranormal voices, no standardized perceptual tools exist, and as far as we know, no intercultural comparisons have been made. We chose voice characteristics from an earlier study where Norwegian actors and non-actors were evaluated (Bele 2002). No anchor samples were used in the present study, as that could have affected the detection of possible cultural differences in the assessments. We had only ten raters in total, so the results may be regarded as preliminary. However, our results seem to be in line with those results comparing the evaluation of dysphonic voices by experts from different countries; they indicate that certain key characteristics are rated similarly, while some variables show differences. Further investigations will focus on the results obtained by different groups of listeners – for example, professional and novice listeners from different genders, age, cultural and linguistic backgrounds – and address a wider variety of languages. Future studies should also be designed with a more mixed-methods approach and include interviews with the raters to disclose the underlying reasons for their evaluations.

## Conclusion

This study compared Brazilian and Finnish voice experts' perceptual ratings of Brazilian actresses' and non-actresses' voices. Differences were found for ratings of nasality and pitch range. Overall voice quality and suitability of the pitch for the speaker, on the other hand, showed strong correlation among Finnish and Brazilian voice experts. The rating procedure tested seems to be useful for speech and voice coaches in evaluating normal and supranormal speaking voices potentially independent of culture. In line with earlier studies comparing the evaluation of dysphonic voices by experts from different countries, our results showed that certain key characteristics are rated similarly, while some variables show differences.

## References


Aalto, A. and Parviainen, K. (1985), *Auta Ääntäsi, 5. laajennettu painos*, Keuruu: Otava.

Aderhold, E. (2021), *Sprecherziehung des Schauspielers: Grundlagen und Methoden*, Berlin: Henschelverlag.

Anders, L. C., Hollien, H., Hurme, P., Sonninen, A. and Wendler, J. (1988), 'Perception of hoarseness by several classes of listeners', *Folia phoniatrica*, 40:2, pp. 91–100.

Andreeva, B., Demenko, G., Möbius, B., Zimmerer, F., Jügler, J. and Oleskowicz-Popiel, M. (2014), 'Differences of pitch profiles in Germanic and Slavic languages', *Proceedings of Interspeech 2014: 15th Annual Conference of the International Speech Communication Association*, Singapore, 14–18 September, Baixas: ISCA, pp. 1307–11.

Barbosa, P. A. and Albano, E. C. (2004), 'Brazilian Portuguese', *Journal of the International Phonetic Association*, 34:2, pp. 227–32.

Bauer, H. (1973),'Klinik der Stimmstörungen', P. Biesalski, G. Böhme, F. Frank and R. Luchsinger (eds), *Biesalski, Phoniatrie und Pädaudiologie*, Stuttgart: Georg Thieme Verlag, pp. 165–93.

Behlau, M. and Murry,T. (2011),'International and intercultural aspects of voice and voice disorders', in D. E. Battle (ed.), *Communication Disorders in Multicultural and International Populations*, 4th ed., Missouri: Elsevier, pp. 174–207.

Bele, I. V. (2002),'Professional speaking voice, a perceptual and acoustic study of male actors' and teachers' voices', Ph.D. thesis, Oslo: University of Oslo.

Bele, I. V. (2005), 'Reliability in perceptual analysis of voice quality', *Journal of Voice*, 19:4, pp. 555–73.

Bele, I. V. (2007), 'Dimensionality in voice quality', *Journal of Voice*, 21:3, pp. 257–72.

Bele, I. V., Laukkanen, A.-M. and Sipilä, L. (2010),'Effects of a three-week vocal exercise program using the Finnish Kuukka exercises on the speaking voice of Norwegian broadcast journalism students', *Logopedics, Phoniatrics, Vocology*, 35:4, pp. 150–65, https://doi.org/10.3109/14015430903578779.

Berry, C. (1995), *Your Voice and How to Use It*, 2nd ed., London: Virgin Books.

Best, C. T. (2019),'The diversity of tone languages and the roles of pitch variation in non-tone languages: Considerations for tone perception research', *Frontiers in Psychology*, 10, https://doi.org/10.3389/fpsyg.2019.00364.

Blankenship, B. (2002),'The timing of nonmodal phonation in vowels', *Journal of Phonetics*, 30:2, pp. 163–91.

Brinca, L., Batista, A. P., Tavares, A. I., Pinto, P. N. and Araújo, L. (2015), 'The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli', *Journal of Voice*, 29:6, https://doi.org/10.1016/j.jvoice.2015.01.007.

Brownell, J. (2018), *Listening: Attitudes, Principles, and Skills*, 6th ed., New York: Routledge.

Chan, I. L. and Chang, C. B. (2019),'Perception of nonnative tonal contrasts by Mandarin–English and English–Mandarin sequential bilinguals', *Journal of the Acoustical Society of America*, 146:2, pp. 956–72.

Chan, K. M. K. and Yiu, E. M.-L. (2006),'A comparison of two perceptual voice evaluation training programs for naive listeners', *Journal of Voice*, 20:2, pp. 229–41, https://doi.org/10.1016/j.jvoice.2005.03.007.

Cooper, M. (1973), *Modern Techniques of Vocal Rehabilitation*, Springfield, IL: Charles C. Thomas.

Das, A., Zhao, G., Levis, J., Chukharev-Hudilainen, E. and Gutierrez-Osuna, R. (2020), 'Understanding the effect of voice quality and accent on talker similarity', *Interspeech 2020*, virtual conference, 25–29 October, Shanghai, China, pp. 1763–67.

Dejonckere, P. H., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier, L. and Millet, B. (1998), 'Reliability and clinical relevance of perceptual evaluation of pathological voices', *Revue De Laryngologie – Otologie – Rhinologie*, 119:4, pp. 247–48.

DeVore, K. and Cookman, S. (2009), *The Voice Book: Caring for, Protecting, and Improving Your Voice*, Chicago: Chicago Review Press.

Ding, H., Hoffmann, R. and Jokisch, O. (2017), 'Prosodic correlates of voice preference in Mandarin Chinese and German: A cross-linguistic comparison', in *Konferenz Elektronische Sprachsignalverarbeitung*, Dresden: TUDpress, pp. 83–90.

Doty, N. D. (1998), 'The influence of nationality on the accuracy of face and voice recognition', *American Journal of Psychology*, 111:2, pp. 191–214.

Eadie, T. L. and Baylor, C. R. (2006), 'The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice', *Journal of Voice*, 20:4, pp. 527–44.
Englert, M., v. Latoszek, B. B. and Behlau, M. (2022), 'The impact of languages and cultural backgrounds on voice quality analyses', *Folia phoniatrica et logopaedica*, 74:2, pp. 141–52.
Erickson, D., Kawahara, S., Rilliard, A., Hayashi, R., Sadanobu, T., Li, Y., Daikuhara, H., de Moraes, J. and Obert, K. (2020), 'Cross cultural differences in arousal and valence perceptions of voice quality', in *Speech Prosody 2020*, Tokyo, Japan, 25–28 May, pp. 720–24.
Esposito, C. M. (2010),'The effects of linguistic experience on the perception of phonation', *Journal of Phonetics*, 38:2, pp. 306–16.
Forsmark, K. (1977), *Sa du nåt? En bok om röst- og talvård*, Stockholm: LiberLäromedel.
Garellek, M. and Keating, P. (2011), 'The acoustic consequences of phonation and tone interactions in Jalapa Mazatec', *Journal of the International Phonetic Association*, 41:2, pp. 185–205.
Ghio, A., Weisz, F., Baracca, G., Cantarella, G., Robert, D., Woisard, V., Fussi, F. and Giovanni, A. (2011), 'Is the perception of voice quality language-dependant? A comparison of French and Italian listeners and dysphonic speakers', in *12th Annual Conference of the International Speech Communication Association*, Florence, Italy, 27–31 August, pp. 525–28.
Gordon, M. and Applebaum, A. (2010), 'Acoustic correlates of stress in Turkish Kabardian', *Journal of the International Phonetic Association*, 40:1, pp. 35–58.
Gordon, M. and Ladefoged, P. (2001), 'Phonation types: A cross-linguistic overview', *Journal of Phonetics*, 29:4, pp. 383–406.
Hakulinen, L. (1979), *Suomen kielen rakenne ja kehitys* ('The structure and development of the Finnish language'), 4th ed., Helsinki: Otava.
Hartelius, L., Theodoros, D., Cahill, L. and Lillvik, M. (2003), 'Comparability of perceptual analysis of speech characteristics in Australian and Swedish speakers with multiple sclerosis', *Folia phoniatrica et logopaedica*, 55:4, pp. 177–88.
Hillman, R. (2003), 'Overview of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) instrument developed by ASHA Special Interest Division 3', in *The 32nd Annual Symposium on the Care of the Professional Voice*, 4 June, Philadelphia, PA: AJSLP.
Hirano, M. (1989),'Objective evaluation of the human voice: Clinical aspects', *Folia phoniatrica*, 41:2&3, pp. 89–144, https://doi.org/10.1159/000265950.
Iivonen, A. (1998), 'Intonation in Finnish', in *Intonation Systems: A Survey of Twenty Languages*, Cambridge: Cambridge University Press, pp. 314–30.
Ilomaki, I., Laukkanen, A.-M., Leppanen, K. and Vilkman, E. (2008), 'Effects of voice training and voice hygiene education on acoustic and perceptual speech parameters and self-reported vocal well-being in female teachers', *Logopedics, Phoniatrics, Vocology*, 33:2, pp. 83–92, https://doi.org/10.1080/14015430701864822.
Iwarsson, J. and Petersen, N. R. (2012), 'Effects of consensus training on the reliability of auditory perceptual ratings of voice quality', *Journal of Voice*, 26:3, pp. 304–12.
Järvikivi, J., Vainio, M. and Aalto, D. (2010),'Real-time correlates of phonological quantity reveal unity of tonal and non-tonal languages', *PloS One*, 5:9, https://doi.org/10.1371/journal.pone.0012603.

Jónsdottir, V., Laukkanen, A.-M. and Siikki, I. (2003), 'Changes in teachers' voice quality during a working day with and without electric sound amplification', *Folia phoniatrica et logopaedica*, 55:5, pp. 267–80, https://doi.org/10.1159/000072157.

Keating, P., Kuang, J., Garellek, M., Esposito, C. M. and Khan, S. u. D. (2023), 'A cross-language acoustic space for vocalic phonation distinctions', *Language*, 99:2, pp. 351–89.

Kempster, G. B., Gerratt, B. R., Verdolini, A. K., Barkmeier-Kraemer, J. and Hillman, R. E. (2009), 'Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol', *American Journal of Speech-Language Pathology*, 18:2, pp. 124–32, https://doi.org/10.1044/1058-0360(2008/08-0017).

Koo, T. K. and Li, M. Y. (2016), 'A guideline of selecting and reporting intraclass correlation coefficients for reliability research', *Journal of Chiropractic Medicine*, 15:2, pp. 155–63, https://doi.org/10.1016/j.jcm.2016.02.012.

Kreiman, J., Gerratt, B. R. and Ito, M. (2007), 'When and why listeners disagree in voice quality assessment tasks', *Journal of the Acoustical Society of America*, 122:4, pp. 2354–64.

Kreiman, J., Gerratt, B. R. and Khan, S. u. D. (2010), 'Effects of native language on perception of voice quality', *Journal of Phonetics*, 38:4, pp. 588–93.

Ladefoged, P. and Maddieson, I. (1998), 'The sounds of the world's languages', *Language*, 74:2, pp. 374–76.

Laukkanen, A.-M. (2003), 'On the goals and bases of vocal exercises', in R. Dal Vera (ed.), *Film, Broadcast and e-Media Coaching and Other Contemporary Issues in Professional Voice and Speech Training Presented by the Voice and Speech Review*, Cincinnati, OH: Voice and Speech Trainers Association, Incorporated, pp. 255–62.

Lennes, M. (2010), 'Finnish sound structure: Phonetics, phonology, phonotactics and prosody', *Puhe ja kieli*, 30:2, pp. 121–24.

Leppänen, K., Laukkanen, A.-M., Ilomäki, I. and Vilkman, E. (2009), 'A comparison of the effects of voice massage and voice hygiene lecture on self-reported vocal well-being and acoustic and perceptual speech parameters in female teachers', *Folia phoniatrica et logopaedica*, 61:4, pp. 227–38, https://doi.org/10.1159/000228000.

Lessac, A. (1967), *The Use and Training of the Human Voice: A Practical Approach to Speech and Voice Dynamics*, Mountain View, CA: Mayfield Publishing.

Lewis, J. A. (2002), *Social Influences on Female Speakers' Pitch*, Berkeley, CA: University of California Press.

Loakes, D. and Gregory, A. (2022), 'Voice quality in Australian English', *JASA Express Letters*, 2:8, https://doi.org/10.1121/10.0012994.

Master, S., De Biase, N., Chiari, B. M. and Laukkanen, A.-M. (2008), 'Acoustic and perceptual analyses of Brazilian male actors' and nonactors' voices: Long-term average spectrum and the "actor's formant"', *Journal of Voice*, 22:2, pp. 146–54, https://doi.org/10.1016/j.jvoice.2006.09.006.

Master, S., De Biase, N. G. and Madureira, S. (2012), 'What about the "actor's formant" in actresses' voices?', *Journal of Voice*, 26:3, pp. e117–22.

Mennen, I., Schaeffler, F. and Docherty, G. (2007), 'Pitching it differently: A comparison of the pitch ranges of German and English speakers', in *ICPhS 2007 Proceedings of 16th International Congress of Phonetic Sciences,* Saarbrücken, 6–10 August, pp. 1769–72.

Nilsson, T., Laukkanen, A.-M. and Syrjä, T. (2022), 'Effects of sixteen month voice training of student actors applying the Linklater voice method', *Journal of Voice*, 36:5, https://doi.org/10.1016/j.jvoice.2020.08.014.

Ohala, J. J. (1983),'Cross-language use of pitch: An ethological view', *Phonetica*, 40:1, pp. 1–18.
Ohara, Y. (1992), 'Gender-dependent pitch levels: A comparative study in Japanese and English', in K. Hall, M. Bucholtz and B. Moonwomon (eds), *Locating Power: Proceedings of the Second Berkeley Women and Language Conference*, vol. 2, pp. 468–77.
Ohara, Y. (1999), 'Performing gender through voice pitch: A cross-cultural analysis of Japanese and American English', in U. Pasero and F. Braun (eds), *Wahrnehmung und Herstellung von Geschlecht: Perceiving and Performing Gender*, Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 105–16.
Olwage, G. (2004),'The class and colour of tone: An essay on the social history of vocal timbre', *Ethnomusicology Forum*, 13:2, pp. 203–26.
Parsa, V. and Jamieson, D. G. (2001), 'Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech', *Journal of Speech, Language, and Hearing Research*, 44:2, pp. 327–39, https://doi.org/10.1044/1092-4388(2001/027).
Pegoraro Krook, M. I. (1988), 'Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis', *Folia phoniatrica*, 40:2, pp. 82–90.
Perkins, W. H. (1971), *Speech Pathology: An Applied Behavioral Science*, St Louis, MO: The C. V. Mosby.
Podesva, R. J. and Callier, P. (2015), 'Voice quality and identity', *Annual Review of Applied Linguistics*, 35, pp. 173–94, https://doi.org/10.1017/S0267190514000270.
Proctor, D. F. (1968),'The physiologic basis of voice training', *Annals of the New York Academy of Sciences*, 155:1, pp. 208–28.
Richter, B. (2014), *Die Stimme: Grundlagen, künstlerische Praxis, Gesunderhaltung*, Berlin: Henschel Verlag.
Rodenburg, P. (1993), *The Right to Speak: Working with the Voice*, London: Routledge.
Shewell, C. (2013), *Voice Work: Art and Science in Changing Voices*, Bognor Regis: John Wiley & Sons.
Suomi, K., Toivanen, J. and Ylitalo, R. (2008), *Finnish Sound Structure: Phonetics, Phonology, Phonotactics and Prosody (Studia humaniora ouluensia 9)*, Oulu: University of Oulu, https://oulurepo.oulu.fi/bitstream/handle/10024/36099/isbn978-951-42-8984-2.pdf?sequence=1&isAllowed=y. Accessed 24 May 2024.
Sweeney, T., Grunwell, P., Leahy, M., Malone, J. and Sell, D. (1996), 'Nasality and nasal airflow: Definition of terms', *Journal of Clinical Speech and Language Studies*, 6:1, pp. 65–76.
Szakay, A. (2012),'Voice quality as a marker of ethnicity in New Zealand: From acoustics to perception 1', *Journal of Sociolinguistics*, 16:3, pp. 382–97.
Titze, I. R. (2001), 'Acoustic interpretation of resonant voice', *Journal of Voice*, 15:4, pp. 519–28.
Ullakonoja, R. (2007), 'Comparison of pitch range in Finnish (L1) and Russian (L2)', in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 6–10 August, Saarbrücken: Universität des Saarlandes, pp. 1701–04.
Verdolini, K., Druker, D. G., Palmer, P. M. and Samawi, H. (1998), 'Laryngeal adduction in resonant voice', *Journal of Voice*, 12:3, pp. 315–27.
Waaramaa, T., Lukkarila, P., Järvinen, K., Geneid, A. and Laukkanen, A.-M. (2021), 'Impressions of personality from intentional voice quality in

Arabic-speaking and native Finnish-speaking listeners', *Journal of Voice*, 35:2, https://doi.org/10.1016/j.jvoice.2019.09.003.
Wagner, A. and Braun, A. (2003),'Is voice quality language-dependent? Acoustic analyses based on speakers of three different languages', *Language*, 6:4, pp. 1–4, https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_0651.pdf. Accessed 24 May 2024.
Yamaguchi, H., Shrivastav, R., Andrews, M. L. and Niimi, S. (2003), 'A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale', *Folia phoniatrica et logopaedica*, 55:3, pp. 147–57.
Yanushevskaya, I., Gobl, C. and Ní Chasaide, A. (2018), 'Cross-language differences in how voice quality and $f_0$ contours map to affect', *Journal of the Acoustical Society of America*, 144:5, pp. 2730–50.
Yiu, E. M.-L, Murdoch, B., Hird, K., Lau, P. and Ho, E. M. (2008), 'Cultural and language differences in voice quality perception: A preliminary investigation using synthesized signals', *Folia phoniatrica et logopaedica*, 60:3, pp. 107–19.

## Contributor details

Tuuli Nilsson is a voice and speech lecturer in The Theatre Academy, University of Arts in Helsinki. Beyond being trained in various traditional ways of training voice in the Finnish actor-training programme, she is a certified Alexander technique teacher, designated Linklater teacher and associate teacher of Fitzmaurice Voicework. For her Ph.D. she is focusing on the effectiveness of actor-centred voice training methods. She, for example, investigates the acoustic and perceptual effects and voice training-related experiences in actors' voice training performed by applying the Linklater voice method.

Contact details: Speech and Voice Research Laboratory, Faculty of Social Sciences, Tampere University, Tampere FI-33014, Finland.
E-mail: tuuli-mariia.nilsson@uniarts.fi
https://orcid.org/0000-0003-2520-3384

Anne-Maria Laukkanen, Ph.D., is a Finnish speech scientist and voice trainer. She is professor of speech technique and vocology at Tampere University, Finland. Her main research interests are voice quality in speech and singing, effects and bases of vocal exercises, effects and mechanisms of vocal loading and applicability of various research methods in vocological research.

Contact: Speech and Voice Research Laboratory, Faculty of Social Sciences, Tampere Univesity, Tampere FI-33014, Finland.
E-mail: anne-maria.laukkanen@tuni.fi
https://orcid.org/0000-0003-4836-2513

Tiina Syrjä (doctor of arts in theatre and drama) is a university lecturer of voice and speech at the degree programme in theatre arts at Tampere University, Finland.

Contact: Degree Programme in Theatre Arts, Faculty of Information Technology and Communication Sciences, Tampere University, Tampere FI-33014, Finland.
E-mail: tiina.syrja@tuni.fi
https://orcid.org/0000-0002-9545-371X

Suely Master is a professor in the postgraduate programme in performing arts at the Institute of Arts at UNESP. She has served as an assistant professor at the Department of Performing Arts since 1997 and has held the position of pedagogical coordinator for the bachelor's degree in performing arts and the degree in theatre arts since 2023. She is a speech therapist specializing in voice, having graduated from the Escola Paulista de Medicina in 1978, where she also obtained specializations in speech therapy in 1980. She holds a master's degree (1989) and a doctorate in health studies (2005). Recognized as a voice specialist by the Federal Council of Speech Therapy in 1996, she also completed a post-Doctorate in the applied linguistics studies programme at PUC-SP in 2009. Since 2012, she has been a leader of the Research Group Processes, Poetics and Vocal Technique for the Actor and coordinates the ArtInclusiva Extension Project. Having worked in theatre and cinema since 1980, she has developed research into the technique of scenic voice and its implications for expressiveness, especially concerning games and vocal improvisation, which resulted in the proposal 'Physiology of Expressivity'. She completed internships at the Department of Speech Communication and Voice Research, University of Tampere, Finland (2004), the National Center of Voice and Speech, University of Utah, USA (2010), and the Voice Laboratory of the University School of Speech Therapy in Chile, Santiago (Oct 2013). Suely Master also contributes as a reviewer for national and international journals and has presented research projects at various national and international conferences.

Contact: Department of Performing Arts, Instituto de Artes, Universidade Estadual Paulista, Sao Paulo, SP, Brazil.
E-mail: suely.master@unesp.br
https://orcid.org/0000-0001-8428-7501