TRAMES, 2019, **23**(**73/68**), 2, 239–257

THE EFFECTS OF CULTURE ON VOICE LIKABILTY

Hille Pajupuu, Rene Altrov, and Jaan Pajupuu

Institute of the Estonian Language, Tallinn

Abstract. This study investigated the effects of culture on voice likability assessments. A total of 32 Finns and 32 Estonians rated the likability of 40 Finnish and 40 Estonian female and male voices. The voices represented two phonogenres: poetry and interview. The results showed that Finns and Estonians liked the same voices, but the listeners preferred Finnish voices reading poetry to Estonians, and Estonian voices in interviews to Finnish. The gender and age of the speaker only had a low impact on likability ratings. An analysis of acoustic correlates of voice likability was also conducted, which showed that likable and unlikable voices were differentiated by a set of frequency, energy and spectral parameters, but not tempo parameters.

Keywords: voice likability, culture, GeMAPS, phonogenre, speech style, speaker characteristics, paralinguistics

DOI: https://doi.org/10.3176/tr.2019.2.08

1. Introduction

The voice conveys a lot of information about the speaker, which is why the voice has an important role in communication. Even if we cannot see the speaker, for instance in a phone conversation, we can create an image of them: what their native language is, along with their age, gender, emotional state (whether they are sad or happy, bored or excited), intentions, social status, character and even appearance. People have preferences as to which voices they like or do not like. People with likable voices are considered socially attractive: friendly, competent, self-assured and trustworthy (see McAleer et al. 2014, Schweitzer et al. 2017, Ueda et al. 2013). Many professions necessitate a pleasant voice, for example politicians, news presenters, customer support persons, teachers and voice actors. The last decade has seen a noticeable increase in devices that use the voice for communication and information transfer (e.g. smartphones, reading assistants, car

applications). One criterion for choosing voices for technical solutions is their likability to a wide range of people, whether it is a human or synthesised voice.

Likability we take to mean "how much we like a speaker based on the sound of her/his voice and manner of speaking" (Burkhardt et al. 2011). Schuller and Batliner (2014) consider likability a long-term personality trait. Previous studies have shown that although listeners' ratings may differ on an absolute scale, they concur in terms of which voices are likable or not (see Altrov et al. 2018, Ding et al. 2018, Goy et al. 2016, Obuchi 2017). A likable voice is describable by acoustic parameters. Depending on the field, studies have used either a classical set of features (e.g. voice pitch, energy, speaking rate) or a choice among all possible parameters for a subset optimised by the discriminatory power. Due to the studies' different cultural backgrounds, different aims and different parameter choices, the results are not always comparable and therefore generalisations about the acoustics of likable voices are difficult to form.

Despite a marked increase in interest in the last few decades in the recognition of speaker traits and states from voices, there is still little research and knowledge about voice likability and its acoustics (see Schuller et al. 2015). Some studies have addressed cross-gender perception of voice likability/attractiveness and determined relevant acoustic parameters (e.g. Babel et al. 2014, Bruckert et al. 2006, Collins 2000, Fraccaro et al. 2013, Zuta 2009). Other studies have originated from various technical applications that use voices, for example studying a likable voice for speech synthesis (e.g. Coelho et al. 2008, Ding et al. 2018, Hinterleitner et al. 2014, Syrdal et al. 1998) or classifying voices based on likability (e.g. Coelho et al. 2011, Montacié and Caraty 2012, Pinto-Coelho et al. 2013, Schuller et al. 2012, 2015). Research has also gone into the relation between speaker age and voice likability (e.g. Deal and Oyer 1991, Gampel and Ferreira 2017, Goy et al. 2016) and handling questions about how to assess and annotate voice likability for speech corpora (e.g. Baumann 2017, Gallardo 2016, Gallardo et al. 2017, Schuller and Batliner 2014:170). A few studies have focused on the connections between culture, language and voice likability (e.g. Biadsy et al. 2008, Dahlbäck et al. 2007, Ding et al. 2017, 2018, Trouvain and Zimmerer 2017).

In our study we tried to determine what the influence of culture is on voice likability. That is, how voice likability is perceived across cultures: whether people within a single culture perceive the same voices as likable and the same voices as unlikable, and whether people from different cultures like the same voices. More precisely, we were interested in which voices were perceived as likable by Finns and Estonians, who are geographically close and whose languages belong to the Finnic branch of the Uralic language family.

1.1. Cross-cultural studies on voice likability

There are remarkably few cross-cultural and different language studies on the likability perception of speech, but a few studies can be found on adjacent subjects (see Schuller et al. 2013, 2015). Dahlbäck et al. (2007) studied assessments by Americans and Swedes on a speaker's knowledge of the topic, voice likability and

information quality in an infosystem intended for tourists, which spoke to them in English with either an American or Swedish accent. The listeners preferred voices that shared their accent. Researchers explained this with the similarity-attraction effect – people trust those who are similar to them. Biadsy et al. (2008) came to similar findings in their study on the charisma of voices speaking native and foreign languages. In their research, American, Swedish and Palestinian listeners had to rate political speech in Standard American English and Americans and Palestinians had to rate Palestinian Arabic speech from the aspect of charisma on a five-point Likert scale. Both experiments revealed that listeners gave native speech higher and non-native speech lower charisma ratings.

Trouvain and Zimmerer (2017) came to contrary results while studying how voice attractiveness ratings were affected by speaking in another language. Germans, who assessed speech read by French and Germans (both groups reading in both languages), held French voices to be more attractive than German voices, both in the case of French and German speech. French-accented German speech was perceived as more attractive than the Germans' own native speech and French with a German accent. Therefore foreign-accented speech can be perceived as more attractive than native-accented speech and speakers of a foreign language can be perceived as more attractive than speakers of the listeners' native language. The authors held these results to mirror "the stereotypical picture of French as a popular and sympathetic language for German speakers".

Studies by Ding et al. (2017, 2018) confirmed that there are prosodic features in voices that direct listeners to prefer the same voices among both native and non-native speech. The aim of these studies was to find a likable donor voice for speech synthesis. In the first study, Chinese and Germans rated Chinese voices (speaking Mandarin) and German voices, while in the second study, Chinese and Germans rated German voices. The results of both studies showed a strong correlation between both German and Chinese ratings for both native and non-native voices. Therefore, listeners of different cultural backgrounds perceived similar voices as likable, whether the speech was in their native language or a foreign one.

Previous studies have given contradictory results concerning the influence of culture and language on the voice likability perception. With our study we wished to determine whether the Finnish and Estonian listeners' voice preference depends on the language heard or whether Finnish and Estonian listeners prefer the same voices irrespective of language and culture.

1.2. On the connections between gender and voice likability

Researchers of voice likability have been interested in whether likability ratings are affected by the gender of the speaker. The connection between voice likability and gender is still somewhat open. A study with Californian English speakers and listeners by Babel et al. (2014) revealed that while listeners found the same voices attractive, female voices were perceived as more attractive. In a study by Altrov et al. (2018), Estonian women and men rated the voice likability of Estonian female

and male voices. Raters preferred female voices. A further study conducted in a Chinese-speaking context also showed a significant preference toward female voices (Chang et al. 2018).

In contrast, in a study by Deal and Oyer (1991), English male voices were assessed as being more pleasant than female voices. In a study by Jokisch et al. (2018), where the charisma of German male and female politicians of different ages was rated, male voices also received higher scores. A study by Ueda et al. (2013) on Japanese voice likability showed that speaker gender had no significant effect on rating.

Although the studies are for the most part incomparable, the contradictory results hint that speaker gender might have a different effect on voice likability assessment in different cultures. With our study we wished to add knowledge on the importance of gender in assessing female and male voice likability as exemplified by Finnish and Estonian cultures.

1.3. On the connections between age and voice likability

Voice likability perception may also be influenced by the age of the speaker and listener, varying from one culture to another. Previous research that has considered the effect of age on voice likability can roughly be divided in two – studies that confirmed the effect age has on voice likability ratings and studies that found no effect of age on voice likability ratings.

The study by Deal and Oyer (1991) showed that age has an effect on likability. In their study, five groups of different-aged North American English-speaking listeners rated the likability of speakers of different ages. The results showed that younger speakers were rated as more likable. Weiss and Burkhardt (2012) also drew the same conclusions in their study, where German voices of three different age groups – youths, adults and seniors – were listened to, and where speakers from the younger group were more positively assessed than those from the older group. Goy et al.'s (2016) study also supported the effect of age on likability. They had English-speaking listeners of different ages rate younger and older voices for likability and suitability for voicing audiobooks. Comparing the ratings by younger and older listeners, they found that younger raters gave older voices lower scores. However, both groups considered voices rated as likable and suitable for reading out audiobooks to be more natural and louder, whether the voice was young or old.

Ueda et al. (2013) obtained converse results to the aforementioned studies. Their study of Japanese voice likability involved both female and male listeners in two age groups: young and middle-aged. The voices of four actors were assessed (two men and women in their twenties and two in their forties). Results showed that listeners' age and gender did not affect likability ratings. Neither was the effect of age on voice likability confirmed in Gampel and Ferreira's (2017) study in Brazil. They let listeners rate the likability of older teachers (over 65 years of age). The results revealed that likability was not tied to speaker age, but associated with the acoustic parameters of expressivity: for men, likability correlated with

loudness and variations in the fundamental frequency and loudness; for women, with variations in loudness.

The study by Altrov et al. (2018) evidences both tendencies – that in some cases there is a link between voice likability and age, but in some cases there is no connection. They have examined how Estonian female and male listeners of different ages rated the voice likability of different-aged male and female voices. The results revealed that in the case of female voices, likability ratings fell with rising age, while in male voices there was no connection between age and likability.

However, much like the results from studies on gender and voice likability are not comparable, neither are the results on age and voice likability, as the studies have often been carried out with different aims, and with differently organised listener and speaker age groups. In our study we were interested in whether for Estonian and Finnish listener groups, voice likability is dependent on the age of the speaker.

1.4. On the connections between phonogenre and voice likability

Some studies have shown that voice likability may depend on situation-specific speech style, also known as 'phonogenre'. A study by Ueda et al. (2013) looked at the effect that manner of speaking had on likability and credibility ratings in Japanese. Men and women of different ages had to listen to sentences by four speakers in four speech styles: as if talking to a person, cordial, mechanical and indifferent. Results showed that phonogenre had a significant impact on likability scores. The listeners most preferred voices speaking 'as if talking to a person', followed by voices speaking cordially and mechanically. Indifferent-sounding voices received the lowest ratings. Likability strongly correlated with credibility.

Altrov et al. (2018) studied the likability of Estonian voices in three phonogenres (radio commentaries, talk shows and lectures) and established that likability is connected to phonogenre. The listeners liked lecture voices the least. Acoustically, lecture voices were differentiated from other phonogenres by a significantly higher fundamental frequency.

In our research we wished to find out whether phonogenre plays a role when evaluating the voice likability of people from another culture. For this reason we observed two phonogenres: poetry and interview.

1.5. On the acoustics of voice likability

As regards voice likability, most attention has been given to voice pitch, the acoustic counterpart of which is fundamental frequency (F0). Studies have shown that within English, men with voices a little lower than average and women with voices higher than average are perceived as attractive (see Babel et al. 2014, Riding et al. 2006, Xu et al. 2013), while a higher voice is also associated with youth (e.g. Zuta 2009). In evaluating the likability of the voices of young American women, it was found that likable voices are high, but also exhibit a fast speech rate and vocal fry (Parker and Borrie 2018). In a study by Collins (2000),

Dutch women considered men with a deep voice (i.e. a low-frequency voice) attractive. Bruckert et al. (2006) studied French women, who judged male voices with a temporally increasing pitch more pleasant than voices with a constant or decreasing pitch. Pleasantness and mean pitch were correlated in their study: men with low-pitched voices were more appreciated than men with high-pitched voices. In a study by Weiss and Burkhardt (2010) on German voices, male voices with a low pitch and female voices with energy spread over the spectrum and lower third central moment were classified as likable, as were speakers with a higher articulation rate and lower spectral centre of gravity (darker sound). Yet the importance of mean pitch has not become evident in all German voice likability studies. A study by Zuta (2007) found that attractive male voices feature modulation in pitch (bigger standard deviation in F0) and are not nasal (lacking a dip at around 2.8 kHz). Research by Schweitzer et al. (2017) on female voices showed that most parameters that had been connected to attractiveness in previous studies did not carry weight in their study. For example, they did not find absolute pitch or pitch range to be connected to likability. Instead of phonetic-prosodic realisation, likability was determined by lexical content. As many previous studies had shown that men prefer women with a higher-pitched voice and women prefer lower male voices, a study by Fraccaro et al. (2013) tested whether deliberate manipulation affects vocal attractiveness. Results showed that deliberately exaggerated sex-typical pitch (i.e. lowered voice pitch for men and heightened voice pitch for women) might not increase attractiveness. Yet changing pitch in a sex-atypical direction (rising men's pitch and lowering women's pitch) may lower attractiveness.

Likability might not be describable by isolated parameters of the acoustic signal, and might instead be revealed in a combination of parameters (see Niebuhr et al. 2018, Warhurst et al. 2017, Zuta 2007) and therefore all computed acoustic parameters might lack a specific perceptible counterpart. This has primarily been shown in automatic classification of voices based on likability, where hundreds and thousands of acoustic parameters were in use (see, for example, Schuller et al. 2015). Relying on the many findings of speech analysis, Eyben et al. (2016) have recommended using a minimalistic standard parameter set for the acoustic analysis of speech (GeMAPS) in paralinguistic voice analysis tasks. This allows for replication of findings and makes results from individual researchers or groups more comparable. In our analysis of voice likability we use eGeMAPS, an extended parameter set, which, in addition to frequency-related parameters, energy/amplitude-related parameters and spectral parameters, includes temporal parameters (see Eyben et al. 2016).

To cross-culturally study the effect culture has on voice likability we looked at voices from two cultures – Estonian and Finnish – and searched for an answer to the following questions:

1. Do Finns and Estonians prefer similar voices?
2. Is there a preference for own-language or foreign-language voices?
3. Does likability depend on the speaker's gender?

4. Does likability depend on the speaker's age?
5. Does likability depend on phonogenre?
6. Which acoustic parameters distinguish between likable and unlikable voices?

2. Method

2.1. Material

The material comprised Finnish and Estonian female and male voices taken from the media (see Table 1). Each group had 20 voices, which were equally divided between two phonogenres: 1) interview (a spontaneous conversation with an interviewer) and 2) poetry (read out from text or quoted by heart).

Table 1. Age profiles of Finnish and Estonian female and male voices

	Finnish female voices $N = 20$	Estonian female voices $N = 20$	Finnish male voices $N = 20$	Estonian male voices $N = 20$
max	77	71	80	75
Q3	59	58	61	59
median	46	45	47	45
Q1	40	38	39	39
min	22	22	24	25

2.2. Listening tests

To rate voice likability we conducted two web-based listening tests. In the first test, 20 Finnish female voices and 20 Finnish male voices had to be listened to and rated, and in the second, 20 Estonian female voices and 20 Estonian male voices. Each voice lasted 5 seconds. The passages chosen for listening were not dominated by emotional content. All passages were distinct. Voices from interviews and poetry were presented in a mixed order. Likability had to be rated on a seven-point Likert scale, where 1 = *not likable at all* … 7 = *very likable*, without taking into account sentence content or transmission quality.

There were four groups of raters: Finnish women, Finnish men, Estonian women, Estonian men. Each group had 16 raters, of whom three were between the ages of 20 and 29, three were between 30 and 39, three were between 40 and 49, three were between 50 and 59, and four were aged 60 or older.

2.3. Method

Before analysis, all scores for each rater were normalised:

$$y = \frac{x - \bar{X}}{s},$$

where x is score, $\bar{X}$ is mean of scores and s is standard deviation of scores.

To find out whether raters assess voices similarly within their group – also known as 'inter-rater reliability' – the intra-class correlation coefficient (ICC2k) for the following groups was calculated: all raters together; all men together; all women together; Finnish men; Finnish women; Estonian men; Estonian women.

A Welch Two Sample *t*-test was used to determine whether language, speaker gender and phonogenre affect voice likability ratings (see R Core Team 2017).

Pearson's correlation coefficient was used to measure the possible relationship between speakers' age and their likability scores.

2.4. Acoustic analysis

OpenSMILE software was used for the acoustic analysis of the voices (Eyben et al. 2013). A total of 88 parameters, which form the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS), were extracted from the speech (Eyben et al. 2016).

To find acoustic parameters that distinguish between likable and unlikable voices, the Welch *oneway.test* was used (R Core Team 2017). The test was run separately for Finnish and Estonian female voices and for Finnish and Estonian male voices.

To detect the effect of the parameters, the raw values for each parameter were normalised and confidence intervals (CIs, 95%) for the mean values for likable and unlikable voice groups were calculated. If the CI range of the group mean was fully above zero or fully below zero, then this parameter was considered significantly distinctive for this group.

3. Results

3.1. Do Finns and Estonians prefer similar voices?

In order to find out whether the raters prefer the same voices, inter-rater reliability was assessed for likability for each listener group using intra-class correlation coefficients (ICC2k). The ICC values were bigger than 0.8 in all groups, showing that the members of each group behaved similarly: they considered the same voices likable and the same voices unlikable (see Table 2).

Table 2. Inter-class correlation coefficients for Finnish and Estonian voice likability ratings

Rater groups	Finnish voice likability	Estonian voice likability
All raters	0.96****	0.94 ****
All men	0.92****	0.88 ****
All women	0.92****	0.90 ****
Finnish men	0.84****	0.73 ****
Finnish women	0.86****	0.79 ****
Estonian men	0.87****	0.85 ****
Estonian women	0.84****	0.86 ****

Note. **** $p < 0.0001$

3.2. Is there a preference for own-language or foreign-language voices?

We were interested in whether listeners prefer voices speaking in their native language or a foreign language. The Welch Two Sample *t*-test showed that Finns gave Estonian voices significantly higher scores than Finnish voices (see Table 3). For Estonians there was no statistically significant difference in rating Finnish and Estonian voices.

Table 3. Finnish and Estonian raters' mean scores for Finnish and Estonian voices

Rater groups	*t*	*df*	*P*	Mean scores for Finnish voices	Mean scores for Estonian voices
Finns	4.00	2479.3	0.0001	–0.08	0.08
Estonians	1.36	2544.1	0.1751	–0.03	0.03

3.3. Does likability depend on the speaker's gender?

Using the Welch Two Sample *t*-test, we determined which get higher ratings – female or male voices. It emerged that a statistically significant difference only appeared for Finnish raters, who gave Finnish female voices significantly higher scores than Finnish male voices (see Table 4).

Table 4. Finnish and Estonian raters' mean scores for Finnish and Estonian female and male voices

Rater groups	*t*	*df*	*P*	Mean scores for female voices	Mean scores for male voices
Finnish ratings for Finnish voices	3.15	1231.4	0.0017	0.01	–0.17
Finnish ratings for Estonian voices	1.62	1237.3	0.1058	0.12	0.04
Estonian ratings for Finnish voices	–0.16	1259.7	0.8698	–0.03	–0.02
Estonian ratings for Estonian voices	–0.19	1260.2	0.8466	0.02	0.03

3.4. Does likability depend on the speaker's age?

To ascertain whether likability depends on speaker age, we correlated the scores and ages. It became apparent that older speakers received only slightly lower scores (there was a very weak negative correlation between score and age); see Table 5.

Table 5. Correlation between score and speaker age

Subsets	Correlation coefficient
All voices rated by all raters	–0.15****
All voices rated by Estonians	–0.15****
All voices rated by Finns	–0.14****
All voices rated by Estonian men	–0.15****
All voices rated by Estonian women	–0.15****
All voices rated by Finnish men	–0.14****
All voices rated by Finnish women	–0.14****
Estonian male voices rated by Estonian men	–0.06****
Estonian male voices rated by Estonian women	–0.05****
Estonian male voices rated by Finnish men	–0.09****
Estonian male voices rated by Finnish women	0.01***
Estonian female voices rated by Estonian men	–0.29****
Estonian female voices rated by Estonian women	–0.19****
Estonian female voices rated by Finnish men	–0.22****
Estonian female voices rated by Finnish women	–0.18****
Finnish male voices rated by Estonian men	–0.15****
Finnish male voices rated by Estonian women	–0.21****
Finnish male voices rated by Finnish men	–0.20****
Finnish male voices rated by Finnish women	–0.14****
Finnish female voices rated by Estonian men	–0.10***
Finnish female voices rated by Estonian women	–0.16****
Finnish female voices rated by Finnish men	–0.03****
Finnish female voices rated by Finnish women	–0.24****

Note. ****$p < 0.0001$

3.5. Does likability depend on phonogenre?

We determined the extent to which phonogenre affects voice likability. The results of the Welch Two Sample *t*-test revealed that phonogenre had a significant effect on voice likability: voices reading poetry received scores that were significantly different from voices in interviews. In the case of Finnish voices, raters preferred voices reading poetry over those in interviews; for Estonian voices, the preference was for voices in interviews (see Table 6 and Figures 1 and 2).

Table 6. Finnish and Estonian raters' mean scores for Finnish and Estonian voices reading poetry and voices in interviews

Rater groups	*t*	*df*	*p*	Mean scores for voices reading poetry	Mean scores for voices in interviews
Finnish ratings for Finnish voices	3.85	1236.0	0.0001	0.03	–0.19
Estonian ratings for Finnish voices	6.22	1265.4	0.0001	0.14	–0.19
Finnish ratings for Estonian voices	–1.95	1249.0	0.0514	0.03	0.13
Estonian ratings for Estonian voices	–3.81	1262.8	0.0001	-0.08	0.13

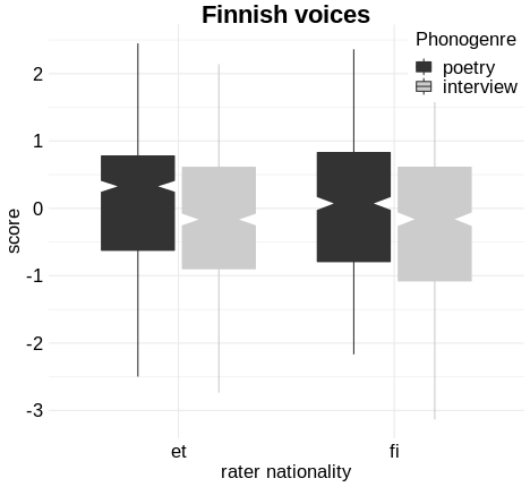

Figure 1. Comparison of Estonian (et) and Finnish (fi) raters' mean scores for Finnish voices reading poetry and Finnish voices in interviews.

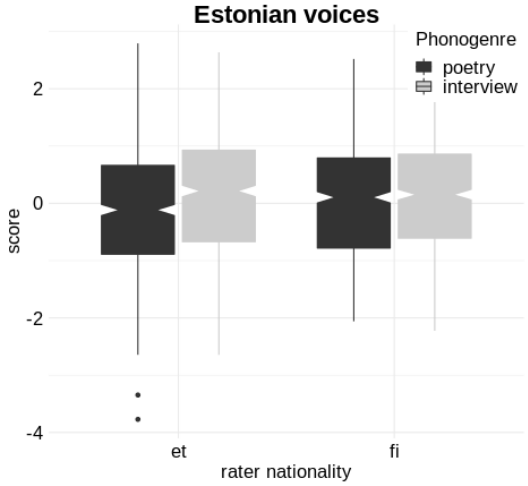

Figure 2. Comparison of Estonian (et) and Finnish (fi) raters' mean scores for Estonian voices reading poetry and Estonian voices in interviews.

3.6. Which acoustic parameters differentiate between likable and unlikable voices?

Based on the Welch *oneway.test*, 11 out of 88 eGeMAPS parameters were statistically significant for differentiating between likable and unlikable female voices: one energy parameter, four frequency parameters and six spectral parameters. For differentiating between likable and unlikable male voices, six parameters were significant: two energy parameters, one frequency parameter and three spectral parameters (see Tables 7 and 8).

Table 7. ANOVA of acoustic parameters for likable and unlikable female voices

eGeMAPS parameter	Description	*F*-statistic *F*(2, 40) for female voices	↑	↓
Energy-/amplitude-related parameters				
logRelF0.H1.A3_sma3nz_amean	Ratio of energy of the first F0 harmonic (H1) to the energy of the highest harmonic in the third formant range (A3)	6.74*	L	UL
Frequency-related parameters				
F1frequency_sma3nz_stddevNorm	*SD* of the first formant (F1) frequency	10.78**	UL	L
F3bandwidth_sma3nz_stddevNorm	*SD* of the third formant (F3) bandwidth	7.83**	L	UL
F1bandwidth_sma3nz_amean	Mean of the first formant (F1) bandwidth	7.78**	L	UL
F2bandwidth_sma3nz_amean	Mean of the second formant (F2) bandwidth	6.12*	L	UL
Spectral (balance) parameters				
mfcc1V_sma3nz_amean	Mean Mel-Frequency Cepstral Coefficient 1 of voiced regions	8.23**	L	UL
hammarbergIndexV_sma3nz_amean	Mean Hammarberg index (the ratio of the strongest energy peaks in the 0–2 kHz vs 2–5 kHz regions) of voiced regions	7.99**	L	UL
slopeV0.500_sma3nz_amean	Mean Spectral Slope 0–500 Hz (linear regression slope of the logarithmic power spectrum) of voiced regions	4.19*	UL	L
slopeUV0.500_sma3nz_amean	Mean Spectral Slope 0–500 Hz (linear regression slope of the logarithmic power spectrum) of unvoiced regions	4.16*	UL	L
mfcc1_sma3_amean	Mean Mel-Frequency Cepstral Coefficient 1	4.15*	L	UL
hammarbergIndexUV_sma3nz_amean	Mean Hammarberg index (the ratio of the strongest energy peaks in the 0–2 kHz vs 2–5 kHz regions) of unvoiced regions	4.14*	L	UL

Note. * $p < 0.05$, ** $p < 0.01$. Groups: L – likable voices, UL – unlikable voices. High (↑) and low (↓) denote groups that have parameter with CI range of mean fully above 0, or fully below 0, respectively.

Table 8. ANOVA of acoustic parameters for likable and unlikable male voices

eGeMAPS parameter	Description	*F*-statistic *F*(2, 40) for male voices	↑	↓
Energy-/amplitude-related parameters				
loudness_sma3_amean	Mean loudness	7.93**	UL	L
loudness_sma3_percentile50.0	The 50th percentile of loudness	7.38**	UL	L
Frequency-related parameters				
F3bandwidth_sma3nz_stddevNorm	*SD* of the third formant (F3) bandwidth	5.98*	UL	L
Spectral (balance) parameters				
spectralFluxUV_sma3nz_amean	Mean spectral flux (difference of the spectra of two consecutive frames) of unvoiced regions	4.48*	UL	L
spectralFlux_sma3_amean	Mean spectral flux (difference of the spectra of two consecutive frames)	4.19*	UL	L
spectralFlux_sma3_stddevNorm	*SD* of the spectral flux (difference of the spectra of two consecutive frames)	4.02*	UL	L

Note. * $p < 0.05$, ** $p < 0.01$. Groups: L – likable voices, UL – unlikable voices. High (↑) and low (↓) denote groups that have parameter with CI range of mean fully above 0, or fully below 0, respectively.

4. Discussion

The goal of our research was to find out whether culture determines which voices are preferred. We looked into Finnish and Estonian likability ratings for Finnish and Estonian female and male voices. Intra-class correlation results revealed similarities in Finnish and Estonian rating behaviour: whatever the speakers' language, the same voices were preferred (see Table 2). From this we can conclude that there is something in voices that makes them cross-culturally likable or unlikable. Yet as Finnish and Estonian are related languages and spoken by neighbouring peoples who are in close contact, shared voice likabilities might stem from being accustomed to hearing the other culture's voices and shared likability standards that might have developed over time. All the same, we cannot dismiss a universal tendency to prefer certain types of voices to others. This had previously been shown in Chinese-German cross-cultural studies by Ding et al. (2017, 2018), which revealed that whatever the language, the raters' preference was for the same voices. More cross-cultural studies on voice likability would add clarity on this issue.

In addition to correlation, some studies have compared the mean scores for likability/charisma given by listeners to native and second-language speakers. Depending on the culture, results have been varied: non-native or foreign-accented speech has received both lower likability scores (Biadsy et al. 2008, Dahlbäck et al. 2007) and higher likability scores (Trouvain and Zimmerer 2017). This has been explained by a preference and trust for the similar, and the prestige of the other language. A comparison of the mean likability scores given to Finnish and Estonian voices revealed that scores given by Finns to Estonian voices were significantly higher than the scores they gave to Finnish voices (i.e. Finns preferred foreign-language voices). For Estonian listeners, there were no differences stemming from Finnish or Estonian voices (see Table 3). The reason why some cultures place higher value on voices speaking their native languages and some foreign languages, and why some are unaffected by language, is difficult to find, but probably depends on some culture-specific values or rules of behaviour.

We also observed other factors that might affect the perception of likability differently depending on culture.

Gender. Stemming from previous research, we assumed that voice likability raters might have gendered preferences and that a preference for male or female voices might depend on the listeners' culture. In our study, we also looked at how Finns and Estonians evaluate Finnish and Estonian female and male voices. Likability assessments of Estonian voices did not reveal a gender-specific preference. This result differed from the results in the study by Altrov et al. (2018), where only the likability of Estonian voices was assessed and where Estonian raters preferred female voices to male voices. In our study, Estonian raters also lacked a gendered preference in rating Finnish voices, but Finnish raters had a significant preference for female Finnish voices (see Table 4). Therefore judgments on Finnish voice likability by Finnish raters coincided with previous research, which had shown a preference for female voices (e.g. Altrov et al. 2018, Babel et al. 2014, Chang et al. 2018). As Finns did not have a gender preference for Estonian voices and neither did Estonians, in contrast to the previous study by Altrov et al. (2018), we cannot claim that a preference for female or male voices is determined solely by culture. Likability assessments might also have been influenced by the set of voices used in these studies. More clarity on this question may arise once there are more studies on cross-cultural voice likability.

Age. In answer to the question of whether the age of a Finnish or Estonian speaker might affect likability ratings given to their voice amongst the Finns and Estonians, we can say, based on our study, that age has only a marginal effect on voice likability. Both Finns and Estonians gave older speakers only slightly lower scores (see Table 5). This finding differs from those studies where the effect of age had been clear: the voices of younger speakers had been rated as significantly more likable than the voices of older speakers (see Deal and Oyer 1991, Goy et al. 2016, Weiss and Burkhardt 2012). Yet, the findings of our study are supported by some previous studies carried out in Brazil and Japan, where a significant connection was not found between speaker age and voice likability (see Gampel

and Ferreira 2017, Ueda et al. 2013). Thus, the results of different studies have shown that the effect of age on voice likability is not universal: in some cultures younger voices may be preferred to older voices, but there are cultures where the listener may find both young and old voices equally pleasant.

Phonogenre. There are few studies on the connection between situation-specific speech style and voice likability, but they point toward a relationship (see, for example, Altrov et al. 2018, Ueda et al. 2013). In our study, two phonogenres were represented: poetry and interview. In the case of Estonian voices, interviews were rated as significantly more likable than voices reading poetry by both Finns and Estonians. In the case of Finnish voices, the listeners preferred voices reading poetry to interviews (see Table 6 and Figures 1 and 2). Apparently there are differences within phonogenres related to culture, and the Finnish performance of poetry is preferred by listeners to the Estonian performance of poetry. As research so far has shown the impact of phonogenre on voice likability, more attention should be focused on this issue.

Acoustics. Of 88 eGeMAPS parameters, 11 differentiated likable female voices from unlikable ones, and six differentiated likable from unlikable male voices. All these parameters were among spectral, frequency and energy parameters (see Tables 7 and 8). Interpretation (finding of a perceptual equivalent) by parameter is neither meaningful nor possible for all parameters, but we can now say that voice likability is a combination of several acoustic parameters (see Niebuhr et al. 2018, Warhurst et al. 2017, Zuta 2007). Voice likability was not determined by speech tempo (tempo parameters were missing among the differentiating parameters). Nor did we find evidence that listeners prefer low or high voices, as frequency parameters related to fundamental frequency were not significant (cf. for example, Babel et al. 2014, Parker and Borrie 2018, Riding et al. 2006). Most differentiating parameters were related to voice quality and timbre. We can say that likable male voices were quieter then unlikable ones.

The limitation of our acoustic study is that the number of analysed voices was relatively small (40 female and 40 male), resulting in few statistically significant acoustic parameters distinguishing between likable and unlikable voices. This kind of acoustic analysis could be repeated in the future on a larger set of speakers.

As far as we know, this is the first voice likability study to deploy GeMAPS parameters to differentiate between likable and unlikable voices. Although Altrov et al. (2018) had used GeMAPS in their study, their focus had been on determining the acoustic differences between likable- and unlikable-sounding phonogenres, so we do not yet have studies with which to compare our results.

5. Summary

The present study aimed to explore the effect of culture on voice likability assessments, and found that voice likability is a trait that might not be limited solely to cultural tenets of pleasantness, but rather crosses cultures. In the example

of Finns and Estonians we saw that both cultures found the same voices likable and unlikable. Yet ratings of voice likability might be affected by culturally different situational speech styles – phonogenres. For example, listeners preferred Finnish voices reading poetry over Estonian voices doing the same, and Estonian voices in interviews over Finnish voices in interviews. Voice likability assessments might also be affected by speaker gender and age in culturally different ways. In our study, the connection between voice likability and gender and age was barely there, but in studies on other cultures this connection had been apparent. The use of eGeMAPS in acoustic analysis revealed a set of frequency, energy and spectral parameters that differentiated likable voices from unlikable ones. The outcomes of this study can be taken into account in the creation of paralinguistic databases and paralinguistic information processing, such as the prediction of voice likability.

Acknowledgements

The authors would like to thank the Finns and Estonians who participated in rating the likability of the voices.

This study was supported by the Estonian Ministry of Education and Research (IUT35-1), by the European Union through the European Regional Development Fund (Centre of Excellence in Estonian Studies) and by the National Programme for Estonian Language Technology 2018–2027.

Address:
Hille Pajupuu
Eesti Keele Instituut
Roosikrantsi 6
10119 Tallinn
Estonia
E-mail: hille.pajupuu@eki.ee
Tel.: +372 51 39 408

References

Altrov, Rene, Hille Pajupuu, and Jaan Pajupuu (2018) “Phonogenre affecting voice likability”. *Proceedings of the 9th International Conference on Speech Prosody 2018*, 177–181.

Babel, Molly, Grant McGuire, and Joseph King (2014) “Towards a more nuanced view of vocal attractiveness”. *PLoS ONE* 9, 2, e88616.

Baumann, Timo (2017) “Large-scale speaker ranking from crowdsourced pairwise listener ratings”. *Proceedings of Interspeech 2017*, 2262–2266.

Biadsy, Fadi, Andrew Rosenberg, Rolf Carlson, Julia Hirschberg, and Eva Strangert (2008) “A cross-cultural comparison of American, Palestinian, and Swedish perception of charismatic speech”. *Proceedings of Speech Prosody 2008*, 579–582.

Bruckert, Laetitia, Jean-Sylvain Liénard, André Lacroix, Michel Kreutzer, and Gérard Leboucher (2006) "Women use voice parameters to assess men's characteristics". *Proceedings of the Royal Society of London B: Biological Sciences* 273 (November 2005), 83–89.

Burkhardt, Felix, Björn Schuller, Benjamin Weiss, and Felix Weninger (2011) "'Would you buy a car from me?' – On the likability of telephone voices". *Proceedings of Interspeech 2011*, 1557–1560.

Chang, Rebecca Cherng-Shiow, Hsi-Peng Lu, and Peishan Yang (2018) "Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan". *Computers in Human Behavior* 84, 194–210.

Coelho, Luis, Daniela Braga, and Carmen Garcia-Mateo (2008) "Voice pleasantness: on the improvement of TTS voice quality". *V Jornadas En Tecnología Del Habla*, 211–214.

Coelho, Luis, Daniela Braga, Miguel Sales Dias, and Carmen Garcia-Mateo (2011) "An automatic voice pleasantness classification system based on prosodic and acoustic patterns of voice preference". *Proceedings of Interspeech* 2011, 2457–2460.

Collins, Sarah A. (2000) "Men's voices and women's choices". *Animal Behaviour* 60, 6, 773–780.

Dahlbäck, Nils, QianYing Wang, Clifford Nass, and Jenny Alwin (2007) "Similarity is more important than expertise: accent effects in speech interfaces" *Proceedings of CHI 2007 – Conference on Human Factors in Computing Systems, April 28–May 3, San José, CA, U.S.A*, 1553–1556.

Deal Leo V. and Herbert J. Oyer (1991) "Ratings of vocal pleasantness and the aging process". *Folia Phoniatr* (Basel) 43, 44–48.

Ding, Hongwei, Rüdiger Hoffmann, and Oliver Jokisch (2017) "Prosodic correlates of voice preference in Mandarin Chinese and German: a cross-linguistic comparison". *28. Konferenz Elektronische Sprachsignalverarbeitung 2017, Saarbrücken*, 83–90.

Ding, Hongwei, Rüdiger Hoffmann, and Oliver Jokisch (2018) "Voice preferences in German: a cross-linguistic comparison of native and Chinese listeners". *Proceedings of the 29th Conference on Electronic Speech Signal Processing. ESSV2018*.

Eyben, Florian, Klaus Scherer, Bjorn Schuller, Johan Sundberg, Elisabeth Andre, Carlos Busso, Laurence Devillers, Julien Epps, Petri Laukka, Shrikanth Narayanan, and Khiet Truong (2016) "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing". *IEEE Transactions on Affective Computing* 7, 2, 190–202.

Eyben, Florian, Felix Weninger, Erik Marchi, and Björn Schuller (2013) "Likability of human voices: a feature analysis and a neural network regression approach to automatic likability estimation". *Proceeding of the 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS) 2013*, 1–4.

Fraccaro, Paul J., Jillian J. M. O'Connor, Daniel E. Re, Benedict C. Jones, Lisa M. DeBruine, and David R. Feinberg (2013) "Faking it: deliberately altered voice pitch and vocal attractiveness". *Animal Behaviour* 85, 1, 127–136.

Gallardo, Laura Fernandez (2016) "Recording a high-quality German speech database for the study of speaker personality and likability". *Tagung Phonetik und Phonologie im deutschsprachigen Raum*, 43–46.

Gallardo, Laura Fernandez, Rafael Zequeira Jimenez, and Sebastian Möller (2017). "Perceptual ratings of voice likability collected through in-lab listening tests vs. mobile-based crowdsourcing". *Proceedings of Interspeech 2017*, 2233–2237.

Gampel, Deborah and Leslie Piccolotto Ferreira (2017) "How do adolescent students perceive aging teachers' voices?" *Journal of Voice* 31, 4, 512.e9-512.e16.

Goy, Huiwen, Kathleen M. Pichora-Fuller, and Pascal van Lieshout (2016) "Effects of age on speech and voice quality ratings". *The Journal of the Acoustical Society of America* 139, 4, 1648–1659.

Hinterleitner, Florian, Christiana Manolaina, and Sebastian Möller (2014) "Influence of a voice on the quality of synthesized speech". *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 99–104.

Jokisch, Oliver, Viktor Iaroshenko, Michael Maruschke, and Hongwei Ding (2018) "Influence of age, gender and sample duration on the charisma assessment of German speakers". *Proceedings of the 29th Conference on Electronic Speech Signal Processing. ESSV2018*.

McAleer, Phil, Alexander Todorov, and Pascal Belin (2014) "How do you say 'hello'? Personality impressions from brief novel voices". *PLoS ONE* 9, 3, 1–10.

Montacié, Claude and Marie-José Caraty (2012) "Pitch and intonation contribution to speakers' traits classification". *Proceedings of Interspeech 2012*, 526–529.

Niebuhr, Oliver, Radek Skarnitzl, and Lea Tylečková (2018) "The acoustic fingerprint of a charismatic voice – initial evidence from correlations between long-term spectral features and listener ratings". *Proceedings of the 9th International Conference on Speech Prosody 2018*, 359–363.

Obuchi, Yasunari (2017) "Personalized quantification of voice attractiveness in multidimensional merit space". *Proceedings of Interspeech 2017*, 2223–2227.

Parker, Michelle A. and Stephanie A. Borrie (2018) "Judgments of intelligence and likability of young adult female speakers of American English: the influence of vocal fry and the surrounding acoustic-prosodic context". *Journal of Voice* 32, 5, 538–545.

Pinto-Coelho, Luis, Daniela Braga, Miguel Sales-Dias, and Carmen Garcia-Mateo (2013) "On the development of an automatic voice pleasantness classification and intensity estimation system". *Computer Speech and Language* 27, 1, 75–88.

Riding, David, Deryle Lonsdale, and Bruce Brown (2006) "The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women". *Journal of Nonverbal Behavior* 30, 2, 55–61.

R Core Team (2017) "R: a language and environment for statistical computing". Available online at <https://www.R-project.org/>. Accessed on February 4, 2019.

Schuller, Björn W. and Anton M. Batliner (2014) *Computational paralinguistics: emotion, affect and personality in speech and language processing*. Chichester, UK: John Wiley and Sons.

Schuller, Björn, Stefan Steidl, Anton Batliner, ElmarNöth, Alessandro Vinciarelli, Alessandro, Felix Burkhardt, Rob van Son, Felix Weninger, Florian Eyben, Tobias Bocklet, Gelareh Mohammadi, and Benjamin Weiss (2012) "The Interspeech 2012 speaker trait challenge". *Proceedings of Interspeech 2012*, 254–257.

Schuller, Björn, Stefan Steidl, Anton Batliner, Felix Burkhardt, Laurence Devillers, Christian Müller, and Shrikanth Narayanan (2013) "Paralinguistics in speech and language: state-of-the-art and the challenge". *Computer Speech and Language* 27, 1, 4–39.

Schuller, Björn, Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob van Son, Felix Weninger, Florian Eyben, Tobias Bocklet, Gelareh Mohammadi, and Benjamin Weiss (2015) "A survey on perceived speaker traits: personality, likability, pathology, and the first challenge". *Computer Speech and Language* 29, 1, 100–131.

Schweitzer, Antje, Natalie Lewandowski, and Daniel Duran (2017) "Social attractiveness in dialogs". *Proceedings of Interspeech 2017*, 2243–2247.

Syrdal, Ann K., Alistair Conkie, and Yannis Stylianou (1998) "Exploration of acoustic correlates in speaker selection for concatenative synthesis". *Proceedings of International Conference on Spoken Language Processing (ICSLP 98)*, 2–5.

Trouvain, Jürgen and Frank Zimmerer (2017) "Attractiveness of French voices for German listeners: results from native and non-native read speech". *Proceedings of Interspeech 2017*, 2238–2242.

Ueda, Hiroshi, Yasunori Arita, and Katsumi Watanabe (2013) "Effects of different manners of speaking on voice likeability, credibility, and intentionality ratings". *Proceedings of the 2013 International Conference on Biometrics and Kansei Engineering (ICBAKE)*, 117–120.

Warhurst, Samantha, Catherine Madill, Patricia McCabe, Sten Ternström, Edwin Yiu, and Robert Heard (2017) "Perceptual and Acoustic Analyses of Good Voice Quality in Male Radio Performers". *Journal of Voice* 31, 2, 259.e1-259.e12.

Weiss, Benjamin and Felix Burkhardt (2010) "Voice attributes affecting likability perception". *Proceedings of Interspeech 2010*, 1934–1937.

Weiss, Benjamin and Felix Burkhardt (2012) "Is 'not bad' good enough? Aspects of unknown voices' likability". *Proceedings of Interspeech 2012*, 510–513.

Xu, Yi, Albert Lee, Wing Li Wu, Xuan Liu, and Peter Birkholz (2013) "Human vocal attractiveness as signaled by body size projection". *PLoS ONE* 8, 4, e62397.

Zuta, Vivien (2007) "Phonetic criteria of attractive male voices". *Proceedings of the 17th International Congress of Phonetic Sciences*, 1837–1840.

Zuta, Vivien (2009) "Voice pleasantness of female voices and the assessment of physical characteristics". *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 5641 LNAI,* 116–125.