Phonogenre affecting voice likability

Rene Altrov, Hille Pajupuu, Jaan Pajupuu

Institute of the Estonian Language, Estonia
rene.altrov@eki.ee, hille.pajupuu@eki.ee, jaan@pajupuu.eu

Abstract

A pleasant voice is an asset not only in various professions but also in speech technologies. This article addresses the correlation between voice likability and phonogenres. Men and women of different age groups were asked to evaluate voice likability in 5-second speech passages arranged into two web-based listening tests, presenting 50 female voices and 60 male voices, respectively. The passages represented three phonogenres: radio commentaries, lectures and radio talk shows. The results demonstrated the impact of a phonogenre on voice likability scores.

The phonogenres were analysed acoustically using the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS). Out of the 88 parameters, 27 turned out to be relevant for describing phonogenres (six parameters coincided with male and female voices). The analysis revealed that one of the three phonogenres—lectures—displayed a consistent difference from the other two. Lecturing voices were considered the least likable, irrespective of the age and gender of the speaker as well as the listener.

Index Terms: voice likability, phonogenre, speaking style, paralinguistics

1. Introduction

Voices contain cues of the speaker’s emotions, intentions, social status, character and even appearance. Whether we like a voice or not may be decisive in our social interactions. Previous studies have shown that people with likeable voices are considered more trustworthy [1], [2]. A pleasant voice is a must for various professions and occupations (lecturers, politicians, salespeople, newsreaders, show hosts, customer support staff). In technologies utilising speech (smart phones, screen readers, e-books, cars, etc.), likability for the target group is one of the criteria of voice selection (see also [3], [4]). Yet, voice likability and its acoustic features are still a little-explored area [4], [5].

Our goal was to find out why some voices are preferred over others: Does voice likability depend on the speaker’s age and gender, or phonogenre (situation-dependent speaking style [6])? Does likability ratings depend on the age and gender of the listener?

There are relatively few studies available on voice likability. Attractiveness of the opposite gender (e.g., [7], [8], [9]) has attracted more attention, while attractiveness and voice likability may, but need not, be correlated. Although studies of voice pleasantness have been conducted in different cultural environments such as Scotland, Japan, Portugal and Germany [1], [2], [3], [5], [10], little attention has been paid to cultural effects on likability ratings [11]. Some studies have even failed to mention the cultural environment, which prevents comparing results and making generalisations.

According to several earlier studies, listener gender and age are not significant factors in likability ratings [3], [5], [12], [13]. This may imply that there is a tendency for people within a culture, irrespective of their age or gender, to like or dislike a particular voice. In German cultural space, also, speaker gender has been observed not to affect the ratings [5]. In Brazil, it has been found that voice likability does not depend on speaker age [14].

Some studies have correlated likability with speaking style, for example, using either factual [15] or command style [16]. The results read that likable speakers exhibit no command styles. As for presenter speaking style, it has been shown that speakers with a fast speaking rate or high liveliness are perceived as pleasant [17]. In Japan, it was found that different manners of speaking (to a listener, cordial, mechanical or indifferent) affect likability, while people of different ages and gender preferred different manners of speaking [2]. Based on these studies, it may be assumed that speaking style has a considerable role in voice likability ratings.

The acoustic parameter sets associated with voice likability differ depending on whether the aim of the study is analysis, synthesis or classification (e.g., [4], [5], [18]).

Our study was conducted in Estonia. The research question was how men and women of different ages rate the likability of male and female voices from three different phonogenres.

2. Method

2.1. Corpus collection

First, we created a voice corpus, which presently contains 60 male (aged 27–81 years) and 50 female voices (aged 25–71 years): 22050 Hz, 16 bit, Mono, Estonian. The voices were collected from radio broadcasts and the Lecture Corpus of Tallinn University of Technology [19].

Three phonogenres, all addressing a large audience, were represented in the voice corpus, which mainly differed in their level of preparation:

1) Radio commentaries. These were short opinion stories presented either by professional radio staff or by other professionals who did not appear on the radio daily. The speaker read a prepared written text meant for a large audience (radio listeners), but without direct contact; monologues.

2) Lectures/conference presentations. These were delivered by a lecturer or conference speaker. They were semi-prepared—the speaker prepared the talk, but did not...
follow the text literally; these were meant for a large audience in direct contact; monologues.

(3) Talk shows. These were comprised of guests’ talking. They were spontaneous—no text was prepared to follow; these were meant for a large audience (radio listeners) without direct contact; dialogues with the host.

2.2. Listening tests

Two web-based listening tests were run to assess voice likability. One of them presented 50 female voices, the other 60 male voices, each for 5 seconds. The passages represented different phonogenres, which were not dominated by emotional contents: for example, “…and to start from somewhere, to incorporate funds, private capital, donations and so on...” All passages were different. Every voice could be repeated as many times as necessary. A voice could score likability, ignoring the absolute values of scores. The possible parameters distinguishing between groups.

Inter-rater reliability was found by using intraclass correlation (see [13]). Pearson’s correlation coefficient was used to measure the relation between speaker age and voice likability, ignoring the absolute values of scores. The possible differences between the likability of male versus female voices was gauged by a paired t-test. Welch’s ANOVA test was used to establish the differences between the ratings based on phonogenres.

2.3. Acoustic analysis

Vocal stimuli were analysed with regard to 88 acoustic parameters. Using openSMILE software, we extracted the parameters defined in the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS), recommended for various areas of automatic voice analysis, including paralinguistic speech analysis [20]. ANOVA was used to find the parameters distinguishing between groups.

Source data for statistical analysis is available at GitHub https://github.com/pajupuujh/phonogenre

3. Results and discussion

3.1. Inter-rater reliability

To find out if listeners assess voices similarly within their group (also known as inter-rater reliability), the intra-class correlation coefficient (ICC2k) for all four groups was calculated. In addition, the coefficient was also calculated for the following listener groups: all under 35 years; all over 35 years; all women; all men; all listeners together. All the ICC values were greater than .80, which proves that all listeners acted similarly (see Tables 1 and 2).

Consequently, Estonian culture belongs to those where voice likability is perceived similarly by men and women, young and old (see also [3], [5], [12], [13]). Future research will show whether this is a universal tendency.

Table 1: Intra-class correlation coefficients with lower and upper bounds for female voice likability in the listener groups.

<table>
<thead>
<tr>
<th>Listener groups</th>
<th>ICC</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women under 35</td>
<td>.89</td>
<td>.84</td>
<td>.93</td>
<td>.0001</td>
</tr>
<tr>
<td>Women over 35</td>
<td>.93</td>
<td>.89</td>
<td>.95</td>
<td>.0001</td>
</tr>
<tr>
<td>Men under 35</td>
<td>.90</td>
<td>.85</td>
<td>.94</td>
<td>.0001</td>
</tr>
<tr>
<td>Men over 35</td>
<td>.91</td>
<td>.87</td>
<td>.94</td>
<td>.0001</td>
</tr>
<tr>
<td>All under 35</td>
<td>.95</td>
<td>.92</td>
<td>.96</td>
<td>.0001</td>
</tr>
<tr>
<td>All over 35</td>
<td>.96</td>
<td>.94</td>
<td>.97</td>
<td>.0001</td>
</tr>
<tr>
<td>All women</td>
<td>.95</td>
<td>.93</td>
<td>.97</td>
<td>.0001</td>
</tr>
<tr>
<td>All men</td>
<td>.95</td>
<td>.93</td>
<td>.97</td>
<td>.0001</td>
</tr>
<tr>
<td>All</td>
<td>.98</td>
<td>.96</td>
<td>.98</td>
<td>.0001</td>
</tr>
</tbody>
</table>

Table 2: Intra-class correlation coefficients with lower and upper bounds for male voice likability in the listener groups.

<table>
<thead>
<tr>
<th>Listener groups</th>
<th>ICC</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women under 35</td>
<td>.88</td>
<td>.83</td>
<td>.92</td>
<td>.0001</td>
</tr>
<tr>
<td>Women over 35</td>
<td>.93</td>
<td>.91</td>
<td>.96</td>
<td>.0001</td>
</tr>
<tr>
<td>Men under 35</td>
<td>.95</td>
<td>.94</td>
<td>.97</td>
<td>.0001</td>
</tr>
<tr>
<td>Men over 35</td>
<td>.90</td>
<td>.86</td>
<td>.93</td>
<td>.0001</td>
</tr>
<tr>
<td>All under 35</td>
<td>.93</td>
<td>.90</td>
<td>.95</td>
<td>.0001</td>
</tr>
<tr>
<td>All over 35</td>
<td>.96</td>
<td>.94</td>
<td>.97</td>
<td>.0001</td>
</tr>
<tr>
<td>All women</td>
<td>.95</td>
<td>.94</td>
<td>.97</td>
<td>.0001</td>
</tr>
<tr>
<td>All men</td>
<td>.93</td>
<td>.91</td>
<td>.96</td>
<td>.0001</td>
</tr>
<tr>
<td>All</td>
<td>.97</td>
<td>.96</td>
<td>.98</td>
<td>.0001</td>
</tr>
</tbody>
</table>

3.2. Likability versus speaker age and gender

To compare the likability results for male and female voices, two mean scores were computed for each evaluator, one for male voices and the other for female voices. Those means were submitted to a paired t-test. The test showed that female voices were preferred over the male ones: \( M_{female\ voice} = 4.26 \), \( M_{male\ voice} = 4.08, t = 4.85, df = 80, p < .0001 \); mean of the differences = .18 (see Figure 1).

There are too few studies available to enable cross-culture comparison. According to a study done on the German cultural environment, speaker gender did not have a significant effect on likability scores [5].

Figure 1: Speaker gender versus voice likability scores
Pearson’s correlation coefficient was used to measure the possible relationship between speaker age and their likability scores, which showed a slight correlation for female speakers \((r = .23, p < .0001)\) and no correlation for male speakers \((r = .08, p < .0001)\); see Figure 2.

This result confirms the observation from [14] that voice likability does not depend on speaker age. It is not impossible that a substantial increase in the number of speakers can reveal more likable age groups.

![Figure 2: Speaker age versus voice likability scores](image1)

### 3.3. Likability of phonogenres

Regarding phonogenres, the best scores were earned by the voices of radio commentaries and the lowest scores were given to those of lectures and conference presentations (see Table 3, Figure 3).

Radio commentaries are specific to the use of a prepared text being read aloud. In lectures and presentations, however, a prepared text is presented freely and directly to the audience.

![Figure 3: Voice likability across phonogenres](image2)

<table>
<thead>
<tr>
<th>Voices in phonogenres</th>
<th>min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radio commentaries</td>
<td>2.6</td>
<td>4.2</td>
<td>4.6</td>
<td>5.0</td>
<td>5.9</td>
</tr>
<tr>
<td>(prepared)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talk shows (spontaneous)</td>
<td>2.5</td>
<td>3.6</td>
<td>4.2</td>
<td>4.6</td>
<td>5.5</td>
</tr>
<tr>
<td>Lectures, presentations (semi-prepared)</td>
<td>1.8</td>
<td>2.7</td>
<td>3.5</td>
<td>4.4</td>
<td>5.1</td>
</tr>
</tbody>
</table>

The significance of differences between mean values of phonogenre groups was also confirmed by Welch’s ANOVA test, where \(F = 240.63, p < .0001\).

### 3.4. Acoustics of the phonogenres

The ANOVA was used to analyse the 88 acoustic parameters measured by eGeMAPS, which included frequency, energy, spectral and temporal descriptors (see [20]). Male and female speakers were handled separately.

A total of 13 parameters were found to have significantly different mean values for phonogenre groups of female voices, and 20 parameters in the case of male voices, with six of the parameters matching. Temporal parameters turned out to be not significant (see Table 4).

To detect the effect of the parameters, the raw values for each parameter were normalised and confidence intervals (95%) for the mean values of each phonogenre group were calculated (see [21]). If the CI range of the group mean was above zero or below zero, then this parameter was considered significantly distinctive for this group (see Table 4).

The results in Table 4 reveal that lectures (semi-spontaneous speech) consistently stand out from the other phonogenres. For male voices, 19 out of 20 parameters were distinctive. For female voices, 9 out of 13 parameters were distinctive but no energy-related parameters had a classifying effect.

The results lead to the conclusion that speaking style depends on the situation, which can affect voice likability in either direction. This indicates that voice pleasantness is not a person’s stable characteristic, being, at least to some extent, influenced by the speaking situation.

The limitation of this study was that different phonogenres were represented by different speakers. In the future, the same voices should be examined in different speaking situations. Also, more phonogenres should be involved.
The results can be used in applications for voice likability estimation and phonogenre detection.

5. Acknowledgements

This study was supported by the Estonian Ministry of Education and Research (IUT35-1), and by the European Union through the European Regional Development Fund (Centre of Excellence in Estonian Studies).

4. Conclusions

This study examined how men and women of different ages evaluate the likability of male and female voices representing three phonogenres: radio commentaries (prepared speech), lectures (semi-prepared speech) and radio talk shows (spontaneous speech). The results showed that the lowest likability scores were given to lecturing voices, irrespective of the gender or age of the listener or of the speaker.
6. References


