Influence of Personal Traits on Impressions of One’s Own Voice

Hikaru Yanagida, Yusuke Ijima, Naohiro Tawara

NTT Corporation, Japan

hikaru.yanagida@ntt.com, ijima@m.ieice.org, naohiro.tawara@ieee.org

Abstract

The aim of this paper is to identify personal traits, e.g., age group, gender, personality traits, and values, that influence the perception of one’s own voice. Previous studies have shown that the perception of one’s own voice is different from those of others. Although studies have also shown that the perception of one’s own voice is associated with the listener’s personal traits, only a limited personal trait was considered in previous studies. In this paper, a large-scale subjective experiment using 162 Japanese participants was conducted to evaluate the perceptual voice impressions of their own and others’ voices. We next analyzed relationships between the obtained voice impressions and their personal traits. As a result, we found that the perception of one’s own voice was affected by multiple personal traits, such as age group, gender, personality traits, and values, that influence the perception of one’s own voice. While the previous studies have focused on one personal trait defined by authors, this study focuses on finding personal traits that impact the perception of one’s own voice from various personal traits listed in Table 1. To determine the relationships between personal traits and perception, it would be desirable to analyze the voices of more speakers. In this paper, we first conducted a large-scale subjective experiment using 162 Japanese participants in their 20s to 60s via crowdsourcing to obtain perceptual scores of one’s own and others’ voices. To identify factors that influence the perception of one’s own voice, we then analyzed relationships between the obtained scores and personal traits. The obtained effect sizes showed that the perception of one’s own voice is affected by multiple personal traits, such as age group, gender, personality traits, and values, that influence the perception of one’s own voice.

1. Introduction

Speech signals convey both linguistic and nonlinguistic information such as a speaker’s age, gender, and emotional state [1]. These are known to affect the perceived impression or behavior of the listener [2, 3, 4]. For example, mate selection [3] and leader election [4] are affected by personality impressions from speech. However, the aforementioned studies mainly focused on a rating of others’ voices. Text-to-speech (TTS) synthesis and voice conversion (VC) have enabled natural soundings of one’s own voice [5]. This indicates that there will be more applications using our own voices, for instance, comprehending or memorizing something by listening to our own voices. However, unlike others’ voices, the perception of one’s own voice has not yet been fully clarified.

In previous studies regarding the perception of one’s own voice, two different tendencies have been reported. The first is that negative reactions are produced by listening to one’s own voice [6, 7]. These studies have indicated that the difference between the bone and air conduction sounds would produce these negative reactions. In contrast, the second is that one’s own voice is more attractive than others’ voices because of the implicit egotism and the familiarity effect [8, 9, 10, 11]. Pucher et al. [12] also reported that blind children, compared with visually impaired children, feel a higher immersion towards a synthetic voice similar to their own. These different tendencies indicate that reactions towards one’s own voice would be different in a number of situations such as applications and listeners’ attributions. Several studies have shown that individual differences, e.g., gender [11], personality traits such as self-positivity-bias [10], and social anxiety [13], affect the perception of one’s own voice. However, because these studies focused on only one personal trait in subjective experiments and data analysis, it was not clear what personal traits significantly influenced the perception of one’s own voice. Furthermore, because these experiments considered mainly younger adults (university students), the various attributes, such as age group, of real listeners have not been covered.

In this study, our aim is to identify the personal traits, e.g., age group, gender, personality traits, and values, that affect the perception of one’s own voice. While the previous studies have focused on one personal trait defined by authors, this study focuses on finding personal traits that impact the perception of one’s own voice from various personal traits listed in Table 1. To determine the relationships between personal traits and perception, it would be desirable to analyze the voices of more speakers. In this paper, we first conducted a large-scale subjective experiment using 162 Japanese participants in their 20s to 60s via crowdsourcing to obtain perceptual scores of one’s own and others’ voices. To identify factors that influence the perception of one’s own voice, we then analyzed relationships between the obtained scores and personal traits. The obtained effect sizes showed that the perception of one’s own voice is affected by multiple personal traits, such as age group, gender, personality traits, and values, that influence the perception of one’s own voice.

2. Method

2.1. Participants

A total of 229 native Japanese participants were recruited online. The participants were in their 20s to 60s. All were registrants in the “Human Information Database 2021” by NTT Data Institute of Management Consulting [14]. The feature of “Human Information Database” is that the registrants’ various personal traits are registered by answering questionnaires in advance. This feature enables us to analyze the relationship between personal traits and behaviors for participants. Table 1 lists the personal traits used for the following analysis.

2.2. Speech materials

Each participant recorded their own voice, which consisted of five neutral Japanese sentences and 23 Japanese words. We asked the participants to sit in front of their own computer in their own room. They were first presented with voice materials synthesized by our internal TTS system, and then they read the materials as in the presented example. Windows sound recorder software in their own computers and their own microphones were used for the voice recording. We also asked the participants to keep their recording environments as quiet as possible. In contrast to previous works, the participants recorded their own voices, which were not considered in previous studies. In contrast to previous works, the participants recorded their own voices, which were not considered in previous studies.

Index Terms: one’s own voice, voice attractiveness, voice familiarity, personal traits
Table 1: Personal traits included in “Human Information Database” used for the analysis.

<table>
<thead>
<tr>
<th>Item</th>
<th>Num. of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age groups</td>
<td>1</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
</tr>
<tr>
<td>Big Five (TIPI-J) [15]</td>
<td>5</td>
</tr>
<tr>
<td>Interpersonal Circumplex (IPPC) [16]</td>
<td>8</td>
</tr>
<tr>
<td>Dark Triad [17]</td>
<td>3</td>
</tr>
<tr>
<td>Dichotomous Thinking [18]</td>
<td>3</td>
</tr>
<tr>
<td>Dispositional Greed [19]</td>
<td>1</td>
</tr>
<tr>
<td>Psychological Entitlement [20]</td>
<td>1</td>
</tr>
<tr>
<td>Resilience [21]</td>
<td>3</td>
</tr>
<tr>
<td>Behavioral adjustment [22]</td>
<td>1</td>
</tr>
<tr>
<td>Optimism [23]</td>
<td>1</td>
</tr>
<tr>
<td>Ten basic values [24]</td>
<td>10</td>
</tr>
<tr>
<td>Satisfaction with Life [25]</td>
<td>4</td>
</tr>
<tr>
<td>Four factors of happiness [26]</td>
<td>2</td>
</tr>
<tr>
<td>Unmitigated Communion [27]</td>
<td>1</td>
</tr>
<tr>
<td>Approval motivation [28]</td>
<td>2</td>
</tr>
<tr>
<td>Prosocial behavior [29]</td>
<td>1</td>
</tr>
<tr>
<td>Grit-S [30]</td>
<td>2</td>
</tr>
<tr>
<td>Procrastination [31]</td>
<td>1</td>
</tr>
<tr>
<td>Subjective age</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>52</td>
</tr>
</tbody>
</table>

voice in their own room.

After the recording, we applied DNN-based speech enhancement to all recorded speech to reduce noises such as ambient sounds, stationary noises, and reverberation from the recorded voices. The pre-trained DeepFilterNet2 [32] was used as the speech enhancement method. We excluded participants with misreading, unintelligible pronunciations, and inadequate recording environments by proof listening. As a result, a total of 190 participants were used for the following experiment.

2.3 Procedure

The voice impression evaluations of participants were conducted on a web browser system in their own room. There was a minimum of one week between the recording and evaluation. In the evaluation, it would be desirable that each participant rated all other participants’ voices in the same manner as the previous works [9, 10, 11]. However, since the number of participants is larger than those of the previous works, evaluating all other participants’ voices would be impractical. Therefore, the participants were presented with 21 recorded voices consisting of three types of speakers; their own voice, ten common speakers for all participants, and ten selected speakers for each participant. The evaluation voices were presented to each participant in random order. The speech content of all voices was the same neutral sentence (“There are several blue hardcover books on the desk.” in English). The participants were instructed to evaluate the five voice impressions (attractiveness, familiarity, intelligibility, confidence, and similarity to their own voice) of the voice on the basis of a 0-100 visual analog scale (VAS). Then, we transformed the 101-point scale into a 7-point scale in seven equal intervals.

Participants rated their own voices only once, and the obtained score was denoted as “self-by-self.” The ten common speakers were chosen from all participants to obtain impression scores of the common speakers evaluated by each participant. Since the common speakers are desirable to have various speaker characteristics, we select them by the following procedure using speaker clustering based on x-vectors. We first extracted x-vectors from five neutral sentences recorded by participants and calculated their average for each participant. Then, we performed speaker clustering by the k-means algorithm with ten clusters. Finally, the closest speakers with centroids of each ten clusters were chosen as the ten common speakers (six males and four females) whose characteristics are different from each other. The average score of these ten speakers was denoted as “other-by-self.” These ten participants, selected as the common speakers, did not participate in the evaluation.

To obtain voice impression scores of each participant evaluated by other participants, ten selected speakers for each participant were also chosen. To do so, we first obtained cosine similarities of x-vectors between one participant and other participants as speaker similarities. Then, ten speakers were selected from the other participants sorted by speaker similarities at equal intervals. In other words, each participant evaluated voices of other participants whose similarities ranged from similar to dissimilar to their own voice. As a result, each participant’s voice was rated by nine other participants on average. The average score from each participant was denoted as “self-by-other.”

After the evaluation, the participants answered questionnaires regarding the frequency of listening to their own recorded voice. The aim is to analyze the relationship between perceived voice impressions and frequency of listening to one’s own recorded voice since it has also been reported that the frequency of listening to one’s own recorded voice influences the recognition of one’s own voice [34]. The questionnaires consisted of two questions described as follows.

1. How many times do you upload videos including your own voice to social networking services (SNSs) in a week?
2. How many times do you record your own voice such as for presentation practice in a week?

Finally, a total of 162 participants who completed the experiment were used for the following analysis (78 males and 84 females; 18 males and 16 females in their 20s, 17 males and 15 females in their 30s, 18 males and 19 females in their 40s, 16 males and 22 females in their 50s, and 9 males and 12 females in their 60s).

3. Results

3.1 Overall results of attractiveness and familiarity

To investigate whether overall participants tend to enhance the ratings of their own voices, we conducted paired-sample t-tests. Due to the space limitation, only attractiveness and familiarity out of five voice impressions were considered in the following analysis.

Figure 1 shows the mean rating scores of attractiveness and familiarity. Paired-sample t-tests between “self-by-self” and “other-by-self” and between “self-by-self” and “self-by-other” were conducted. For attractiveness ratings, there was no significant difference in both between “self-by-self” ($M = 3.40, SD = 1.41$) and “self-by-other” ($M = 3.37, SD = 0.86; t(161) = 0.21, p = 0.84$) and “self-by-self” and “other-by-self” ($M = 3.33, SD = 0.96; t(161) = 0.59, p = 0.55$).

The x-vector extractor, FastResNet-34 with self-attentive pooling [33], was trained on the basis of angular prototypical loss using our over 8,000 internal Japanese speakers.

---

1https://github.com/Rikorose/DeepFilterNet  
2The authors confirmed that speaker characteristics of recorded speech were not changed by the speech enhancement.

---
On the other hand, for the familiarity ratings, the scores in “self-by-self” ($M = 4.61, SD = 1.54$) were significantly higher than those in “self-by-other” ($M = 3.44, SD = 0.72$; $t(161) = 8.69, p < 0.01$) and those in “other-by-self” was significantly higher than those in “other-by-self” ($M = 3.50, SD = 1.01; t(161) = 8.21, p < 0.01$).

### 3.2. Relationship between personal traits and obtained scores

#### 3.2.1. Data analysis

To identify personal traits affecting impressions of one’s own voice, we analyzed the between-group effect size (Cohen’s d) by dividing the participants into two groups on the basis of personal traits. We also conducted $t$-test between the two groups. Gender was divided by male and female, and the frequency of listening to their own recorded voice was divided by whether they were listening to their own recorded voices more than once a week or not. The age groups were respectively divided. For other traits, we divided participants into two groups by one of the following three division points; 25th, 50th, and 75th percentile from the bottom of the personal trait score. We excluded personal traits whose correlation coefficients to the other personal traits were larger than 0.4.

The two scores were calculated to examine whether the ratings by oneself were different from those of others and whether ratings by oneself were different from those by others. The discrepancy scores [10, 13] were calculated by subtracting the “other-by-self” score from the “self-by-self” one (“self-by-self – other-by-self”) and by subtracting the “self-by-other” score from the “self-by-self” one (“self-by-self – self-by-other”).

#### 3.2.2. Results

Figure 2 shows the top five traits that have larger effect sizes for both attractiveness and familiarity. The result of the $t$-test between groups divided by division point showed that all differences were significant except for Unassuming-Ingenious (IPIP-IPC) in Fig. 2a, age group in Fig. 2b, and Gregarious-Extraverted (IPIP-IPC) and Unmitigated Communion in Fig. 2c. For the “self-by-self – other-by-self” score for attractiveness (Fig. 2a), Gregarious-Extraverted had the largest effect size, followed by Openness (Big Five), Frequency of listening to their own recorded voice, Procrastination, and Unassuming-Ingenious. Participants whose Gregarious-Extraverted or Openness scores were low (below the 25th percentile) or whose Procrastination scores were high (the 75th percentile or higher) tended to lower their self-voice evaluation. On the other hand, participants who frequently listened to their own voice (once a week or more) tended to overestimate their self-voice evaluation. For the “self-by-self – self-by-other” score for attractiveness (Fig. 2b), participants whose Openness or Gregarious-Extraverted or Narcissism (Dark Triad) scores were low were likely to lower their self-voice evaluation. Participants whose Procrastination scores were low tended to overestimate their self-voice evaluation. Although the difference was not significant between the age groups ($p = 0.063$), participants whose age groups were high (60s or over) also tended to overestimate their self-voice evaluation.

For the two discrepancy scores for familiarity, the effect sizes were small compared with those of the attractiveness. Frequency of listening to their own recorded voice, Procrastination, Gregarious-Extraverted, Agreeableness (Big Five), and Unmitigated Communion influenced the “self-by-self – other-by-self” score (Fig. 2c). Frequency of listening to their own recorded voice, Procrastination, Unmitigated Communion, Preference for dichotomy (Dichotomous Thinking), and Gregarious-Extraverted also influenced the “self-by-self – self-by-other” score (Fig. 2d).

### 4. Discussions

#### 4.1. Overall results

The results of the $t$-test showed that the familiarity ratings of one’s own voice were significantly higher than those of others and the familiarity of one’s own voice was significantly higher than one’s own voice rated by others. In addition, there was no significant difference for the attractiveness ratings.

For the familiarity ratings, the previous study [12] has reported that familiarity ratings of all acquaintance voices including one’s own voice suggested to be high. However, they did not distinguish one’s own voice from voices of acquaintance, familiarity ratings of only one’s own voice has not been examined. In contrast, we confirmed that the familiarity of one’s own voice is higher than those of others’ voices.

For the attractiveness ratings, the attractiveness of one’s own voice was comparable to those of others’ voices, and the results were not consistent with the previous works [8, 9]. One reason for this that the trend of self-evaluation is different among countries. For example, it was reported that people in the United States and China have higher self-esteem than Japanese people [35]. Since the previous works indicate that self-enhancement of their own voice would be produced by self-esteem, our experiment conducted in Japan would obtained different results from previous studies conducted in the United States and China.

#### 4.2. Personal traits affecting impressions of one’s own voice

The effect size showed that the impression of one’s own voice was affected by Gregarious-Extraverted (IPIP-IPC) and Pro-
Participants whose personal traits fall into the category

division point
cohens d
Mean voice attractiveness ratings

-1.0
-0.5
0.0
0.5
1.0
1.5

(a) “self-by-self – other-by-self” scores for attractiveness.

Participants whose personal traits not fall into the category

division point
cohens d
Mean voice attractiveness ratings

-1.0
-0.5
0.0
0.5
1.0
1.5


There are a number of different tendencies between the evaluations of attractiveness and familiarity. For instance, Openness (Big Five) and Narcissism (Dark Triad) influenced only attractiveness ratings, and Unmitigated Communion influenced only familiarity ratings. It is reported that Openness [37] and Narcissism [38] are related to self-evaluation, and Unmitigated Communion [39] is related to other-evaluations. This indicates that familiarity ratings would be affected by personal traits related to interpersonal relationships than attractiveness ratings. This is because the familiarity ratings of one’s own voice were higher than those of attractiveness, and we speculate how evaluating not our own but others’ voices influenced the discrepancy scores of familiarity.

Our results also indicated that participants with the frequency of listening to their own recorded voice also have high self-voice ratings. One reason for this would be to reduce negative feelings to their own voice by this frequency. Previous studies have indicated that the difference between one’s own recorded voice (air conduction sounds) and one’s own voice when they are speaking (mixture of air and bone conduction sounds) would produce these negative reactions [6, 7]. However, it has also been reported that the frequency of listening to their own recorded voice influenced the recognition of their own voice [34]. From these results, the frequency of listening to their own recorded voice would reduce negative feelings to their own recorded voice and influence their voice ratings.

5. Conclusion

The purpose of this paper was to identify the personal traits that influence the impression of one’s own voice. We conducted a large-scale subjective experiment using 162 Japanese participants in their 20s to 60s via crowdsourcing to obtain perceptual scores of one’s own and others’ voices. We then analyzed relationships between the obtained scores and personal traits to identify factors that influence the perception of one’s own voice. The obtained effect sizes of personal traits showed that multiple personal traits related to extraversion, self-esteem, and frequency of listening to one’s own recorded voice influence the impression of one’s own voice. In future work, we will examine the difference between the perception of our own voices and that of others by setting specific contexts to confirm that the result of this study can be generalized to a specific behavior.
6. References


