







# Character-Voice Embodiment Impacts on the Cognitive Task Performance with the Voice Ownership Illusion.

Yusuke Kunimi<sup>1,2</sup> , Kenta Kimura<sup>1,2</sup> , Keigo Matsumoto<sup>1</sup> ,  
Shinnosuke Takamichi<sup>1,3</sup> , Takuji Narumi<sup>1</sup>  and Masaaki Mochimaru<sup>1,2</sup> 

<sup>1</sup>The University of Tokyo, Japan

<sup>2</sup>National Institute of Advanced Industrial Science and Technology, Japan

<sup>3</sup>Keio University, Japan

## Abstract

Embodying a voice quality different from the innate one by utilizing real-time voice conversion has paid attention to enhance the cognitive abilities and manipulating emotions while social interaction in physical activity. Past research has shown that embodying voice qualities that evoke specific stereotypes can induce a variety of cognitive effects and emotion. However, such an approach has been criticized for its active use of stereotypes and thus reinforces stereotypes about certain groups within society. In contrast, the use of images of well-known characters in stories has the potential to influence thinking and behavior without reinforcing stereotypes of specific social groups. This paper investigate the impact of voice conversion to a animation character voice quality on attitude, behavior, and personality. The results show that animation character-based voice conversion enhanced the planning ability according to the social image of the character.

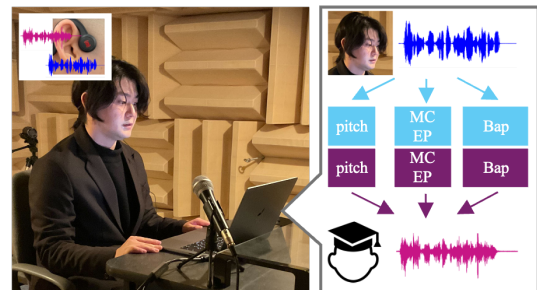
## CCS Concepts

• **Human-centered computing** → *Virtual reality; Mixed / augmented reality;*

## 1. Introduction

Social VR platforms such as VRChat and Cluster provides users to communicate with others using an embodied avatar that gives them various nonverbal cues. Users can customize their avatars' appearance including facial features, body shape, gender and voice quality to suit various communication contexts and styles. It has been suggested that the customization of such non-verbal cues produced by the user affects the quality of communication and the self-perception gained through communication. For instance, previous research has demonstrated that Japanese men experience a sense of liberation from conventional societal norms when using "Bishoujo" avatars, which offer an escape from traditional gender roles and expectations [BG22]. One of the phenomena underlying these effects is the Proteus Effect [YB07]. The Proteus effect is the effect that the thinking and behavior of a user embodying the virtual avatar are influenced by the stereotype evoked by the avatar.

Prior research has shown that both appearance of the avatar and real-time voice conversion to evoke specific stereotypes or self-images based on voice characteristics can elicit effects similar to the Proteus effect [AKT\*21, CJC\*18]. Costa et al. showed that pitch manipulation, which the low pitch of the voice, could be used to increase power-performance. Arakawa et al. demonstrated that age bias could be reduced by converting the voice to one with geriatric-like characteristics [AKT\*21]. It has been suggested that the Proteus effect affects not only temporary thinking



**Figure 1:** Voice Conversion System: The Voice ownership illusion induced character-voice embodiment with voice conversion system via bone-conduction earphone. pitch = Fundamental Frequency, MCEP = Mel Cepstral Coefficients, Bap = Band-Averaged Aperiodicity.

and behavior changes but also has an ongoing effect on the Big-5 personality traits after the experience, particularly when both visual and auditory feedback are modified with an avatar and voice conversion [STOI23]. Voice features are known to affect emotions, and the real-time voice conversion also affects emotions [AJH\*16]. These effects suggest that the real-time voice conversion may have a stronger influence on self-image than visual feedback.

In this study, the phenomenon in which a voice with different characteristics from the converted voice is perceived as if it were one's own voice is called the voice ownership illusion (VOI). VOI induces voice embodiment that embodying a voice quality different from the innate one by utilizing real-time voice conversion. Furthermore, we refer to the "Parakeet effect" as the effect of perceiving such a particular voice quality as one's own while read out sentences, which manifests itself in thinking and behavior. We investigated whether such a methodology secures a sense of ownership and agency over a voice that is different from the own voice, as well as whether one's thinking and behavior are altered by the perception of one's voice as being of a particular voice quality, as in the Proteus effect.

Latan pointed out that the almost all study of the Proteus effect exploits stereotypes about certain social groups and may serve to re-fix stereotypes in society, and strongly criticized that we should not be unaware of the usage of such stereotypes [Cla20]. In fact, it was reported that American men were strongly influenced by gender stereotypes when they experienced Body Ownership Illusion with female avatars [LYB\*19, PGB20]. In contrast, the use of images of well-known characters in stories has the potential to influence thinking and behavior without reinforcing stereotypes of specific social groups. In particular, non-existent characters, such as those in animated films, may be created to reflect stereotypes of a particular group, but new characters can be created to avoid such stereotypes. It is not difficult to choose a well-known character that has little connection to stereotypes of a particular social group. Then, this paper investigate the impact of voice conversion to a animation character voice quality on attitude, behavior, and personality. RQ1: Does a character-voice embodiment impacts attitude, behavior, and personality to align with voice quality on the animation character in stories with voice ownership illusion? RQ2: How does the Parakeet effect relate to sense of voice ownership, sense of speech agency, and self-esteem? The detailed hypotheses followed below:

1. **RQ1-1:** Implicit association was change to similar with character.
2. **RQ1-2:** Planning ability was enhanced by character image.
3. **RQ1-3:** Personality factor of Big-5 was change to similar with character.
4. **RQ2-1:** Sense of Voice Ownership and Sense of Speech Agency related on the Parakeet effect.
5. **RQ2-2:** Self-esteem related on the planning ability.

## 2. Related Work

In this study, we investigated how the Parakeet effect influences attitude, planning ability, and personality in the context of voice ownership illusion. In this section, the Parakeet effect is addressed in contrast with the Proteus effect. First, related works of the Proteus effect and the body ownership illusions were surveyed. Then, works on Parakeet effect and voice ownership illusions were mentioned.

### 2.1. Proteus Effect and Body Ownership Illusions

Body ownership illusions occur when a virtual appearance is perceived as one's own body, typically stimulated by synchronized

visual and other sensory feedback. A well-known study, the rubber hand illusion (RHI), was the first experiment to demonstrate body ownership illusion [BC98]. When tactile and visual feedback were applied, participants perceived a rubber hand as their own. Building on the RHI, the virtual body ownership illusion is a well-researched area where a virtual appearance is perceived as one's own body when visual and other sensory feedback are synchronized [LTMB07, Ehr07, MS13].

The Proteus effect [YB07] is a well-known phenomenon in which cognitive ability and emotion are influenced by body ownership illusion accordance to virtual appearance. For instance, child avatars lead to an overestimation of the object size compared with adult avatars [BGS13]. Also, racial bias against black individuals decreased when participants used black avatars, stimulated by visual feedback [PSAS13]. However, gender bias is enhanced with female avatars by visuo-motor feedback [LYB\*19, PGB20]. These studies suggested that the possibility of avoiding the use and re-fixation of stereotypes for a specific group. By contrast, the use of images of well-known characters in stories has the potential to influence thinking and behavior without reinforcing stereotypes of specific social groups. In particular, non-existent characters, such as those in animated films, may be created to reflect stereotypes of a particular group, but new characters can be created to avoid such stereotypes. For example, Einstein avatar enhanced participants' planning ability and reduce implicit age-bias [BKS18], and superhero avatar enhanced helping behavior [RBB13]. These studies suggested that well-known characters avatars can influence cognition associations with the image of character in stories without reinforcing stereotypes of specific social groups, even while maintaining body ownership illusions.

### 2.2. Parakeet Effect and Voice Ownership Illusions

Voice ownership illusions induce voice embodiment, where a others voice is perceived as one's own through synchronized auditory and other sensory feedback. Zeng [ZMMJ11] investigated how a recorded other voice, when converted and synchronized with motor feedback, could be perceived as one's own voice. Also, Banakou [BS14] explored the perception of a converted voice as one's own when combined with synchronized auditory (recorded voice), visual (lip movements with an avatar), and tactile (vibrated thyroid cartilage) feedback. To expand on these studies, researchers have investigated the perception of a converted voice as one's own using real-time voice conversion systems during speaking. Ohata [OAI22] examined how manipulating the pitch of a voice decreases sense of voice ownership and sense of speech agency compared to one's own voice, using real-time voice conversion systems. Furthermore, previous studies have suggested that a converted voice as a different gender reduced social presence and body ownership in social VR [KPL23]. From these studies, it is evident that voice ownership illusions which a different or converted voice is perceived as one's own can induce voice embodiment, influencing sense of speech agency, sense of voice ownership, and related aspects.

Previous studies investigated that the voice quality of converted voice impacts on cognition and emotion within the context of voice ownership illusions. Costa et.al indicated that power-

performance increased when the pitch of the converted voice was lowered [CJC\*18]. Additionally, manipulating the emotional tone enhanced emotional response through a low-latency voice conversion system [AJH\*16].

We named the Parakeet effect as the influence of voice embodiment on cognition and ability, based on voice quality of personal information within the context of voice ownership illusions. One study suggested that using an elderly-voice decrease implicit age bias with a real-time voice conversion system [AKT\*21]. Also, Ogawa et al. [OBN24] explored that users experienced enhanced teleoperation with a social robot when stimulated a converted voice to match a customer service robot. These studies provided evidence that voice ownership illusions impact cognition and emotion based on voice characteristics. Generally, almost study has been suggested that voice quality of personal information can be categorized with stereotype which age and gender from voice features. Also, various voice qualities such as beauty and attractiveness could not be categorized as stereotypes. Therefore, animation character voice embodiment can be used to investigate what impacts cognitive abilities association with character image without specific stereotype of voice quality of personal information.

### 3. Voice Conversion System

In this study, we investigate the Parakeet effect, focusing on how an animated character-voice influences cognitive ability in relation to the images of character in stories using voice ownership illusion. An overview of our voice conversion system is shown in Figure 1, which details the voice conversion system.

#### 3.1. Voice Conversion Algorithm

We implemented a voice conversion system using a low-latency voice conversion algorithm [ATS19] on a laptop. This algorithm estimates voice characteristics such as pitch (fundamental frequency) and other parameters of the target character. The converted character voice is synthesized from the estimated parameters (pitch, Mel Cepstral Coefficients, Band-Averaged Aperiodicity and so on.). Details of the algorithm are described by [ATS19]. Our voice conversion system was implemented by Python 3.11 on MacBook Pro M1 (Memory: 16 GB). Voice data streaming was performed using the Python library (sound device). To ensure compatibility with various audio interface, we used the internal microphone on the MacBook Pro M1 to decrease audion input/output latency. Speech disorders can occur with a latency of approximately 200 ms [HA84]. Therefore, we facilitated VOI using wired bone-conduction earphones based on [AKT\*21, OBN24].

#### 3.2. System Latency

Previous studies have shown that speech disorders often occur, because of latency in voice feedback [HA84]. Additionally, sense of body ownership and the sense of agency are effected by visual feedback with latency [WSH\*16, RL20, IS16]. Therefore, we measured the voice feedback latency to assess its impact on VOI. To consider the mixed noise in the converted voice, we measured the sound wave using an impulse signal generated by clapping hands

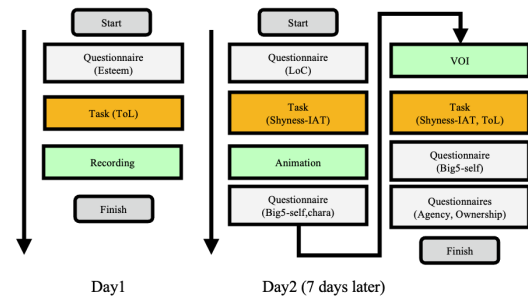


Figure 2: Experimental procedure

after speech during voice conversion. The latency of the voice feedback was calculated by comparing with the recorded raw and converted sounds by 10 samples. As a result, the latency of the voice feedback was 152.6 ms (SE = 4.74 ms).

## 4. Methods

To create the voice conversion model, we must record the voice of the participants. Consequently, the experiments were divided into two days, as illustrated in Figure 2. This section describes overview of the experimental design.

### 4.1. Character Voice

To address the reinforcement stereotype, we used the voices of well-known animation characters from the top 20 animations in Japan (<https://www.nrc.co.jp/report/210218.html>). To respect voice actor rights, we obtained approval from the voice actors to create a voice conversion model for use it in the experiment. The selected animation character is renowned in Japan. The animation featured a character who easily solved difficult problems and assisted a child in a friendly manner. We aimed to investigate whether the voice ownership illusion could impact attitudes toward sociable and enhance planning abilities.

### 4.2. Design and Procedure

The experiment employed a 2×1 within-subject design with pre and post condition measures. This procedures is illustrated in Figure 2. On the first day, participants completed questionnaires and tasks, and their voices were recorded while reading sentences from the ATR503 (A01-50, B01-45) dataset [KTS\*90] to develop a voice conversion model. Seven days later (Day 2), participants completed another questionnaires, tasks and experienced VOI after watching an animation.

### 4.3. Participants

We only included participants of the same gender (male) as the animation character. In addition, we restricted participation to students from a The University of Tokyo to avoid sampling bias. Individuals with hearing impairments or claustrophobia were also excluded.

**Table 1:** Participants classified by Voice similarity and task

	Character	Others
Shyness-IAT	8	22
ToL Task	5	18

All participants were native Japanese speakers and know an animation character. To determine the sample size, we calculated the sample size using G\*Power 3.1 [FELB07] [AJH\*16] (t-Test, matched pairs, two-way; Effect Size = 0.7,  $\alpha$  = 0.05, Power (1- $\beta$ ) = 0.6). The 13 participants were needed.

As a results, we conducted the experiments with 32 participants recruited through social networking services and emails. The mean age of the participants was 31.063(SE= 1.502). We obtained complete data from the number of participants that was shown in Table 1. The participants received a 5,000 yen Amazon gift card after completing the experiment.

#### 4.4. Voice Ownership Illusion

The Voice Ownership Illusion (VOI) is an experience in which a converted voice is perceived as one's own, facilitated through bone-conduction earphones using a real-time voice conversion system. We introduced the voice conversion system to the participants and provided them with instructions regarding its use. The participants practiced operating the system and adjusted the volume of the voice feedback when reading sentences. After the practice session, the participants read aloud 67 sentences from the speech of the character that they had seen in the animation.

#### 4.5. Shyness-Implicit Association Test (IAT)

The Shyness-Implicit Association Test (Shyness-IAT) is a task that measures implicit attitudes based on response times [GMS98]. The participants sorted the words from the center of the screen to the edges based on their attitudes. The IAT score was calculated by measuring the response times during sorting. The IAT is a well-known task for investigating the Proteus effect that demonstrates changes in gender bias [LYB\*19], age bias [BKS18], and racial attitudes [PSAS13] following visual feedback. In this study, we used a sociable character to investigate the changes in shyness and sociability using the Shyness-IAT [ABM02, AF11] after VOI. The IAT was implemented using the Millisecond Test Library and Inquisit 6 on a MacBook M1 Pro, and the IAT scores were calculated accordingly [GNB03]. Also, Positive IAT score indicated self-attitude as shyness, and negative IAT score indicated self-attitude as sociable.

#### 4.6. Tower of London Task

The Tower of London (ToL) task is a well-known task of assessing planning ability [SBW82]. Previous studies have investigated the planning ability in relation to Einstein avatar embodiment effects [BKS18]. Therefore, we used the ToL task to investigate the enhancement of the planning ability as part of the Parakeet effect following VOI. The ToL task involved three balls (red, blue, and green) and poles of different heights. Participants solved the ToL task by moving the balls from their initial positions to the target

positions. The initial and target positions were obtained from the ISO Problems [BB02, KUS11]. The participants were allowed up to three trials per problem and could reset the ball positions. The ToL Score is calculated as follows:

$$ToLScore = \sum_{i=1}^{16} (3 - n_i).$$

is the count of resets and failures for each problem. The total ToL Score is the sum of all 16 problems, with a perfect score of 48 points. Additionally, the D-Score was calculated to measure the enhancement of planning ability. The D-Score is calculated as follows:

$$dScore = ToLScore(Post) - ToLScore(Pre).$$

The ToL task was implemented using the Millisecond Test Library and Inquisit 6 on a MacBook M1 Pro.

#### 4.7. Sense of Speech Agency and Sense of Voice Ownership

Sense of Speech Agency (SoSA) and Sense of Voice Ownership (SoVO) are crucial for the Proteus effect, which is based on the sense of embodiment [WSH\*16, RL20]. Because VOI is not a well-established research area, no validated questionnaires currently exist for measuring the SoSA and SoVO. We developed a questionnaire on SoSA and SoVO, based on [GFP18]. Also, the participants also evaluated the difficulty of speaking while using the voice conversion system. Due to imperfections in the voice conversion algorithm, users assessed the voice similarity in the converted voice quality when using bone-conduction earphones. The questionnaire is presented in Table 2. A 7-point Likert scale (-3: Strongly Disagree, 3: Strongly Agree) was used for this evaluation.

SoSA refers to the perception that the feedback voice is controlled by one's own utterance. SoSA was quantified as follows:

$$SoSA = SoSA1 + SoSA2 + SoSA3 - SoSA4.$$

SoVO refers to the perception that the feedback voice is one's own, specifically a character voice. We observed a similar tendency between voice similarity and VO2. This suggested that the participants perceived the voice similarity and SoVO2 similarly, as indicated in Table 2. Therefore, we used SoVO1 to investigate the sense of voice ownership.

#### 4.8. General Questionnaire

Previous studies have suggested that planning ability is related to self-esteem. Therefore, we measured self-esteem using the RSES-J [MG07]. Additionally, we assessed locus of control in relation to the sense of agency [JAAL18]. Previous study has demonstrated that voice and body embodiment can alter Big-5 personality factors to resemble the character [STOI23]. Thus, we investigated the changes in personality traits to align with the character stimulated by the converted voice feedback. We measured the Big-5 personality factors using the TIPI-J [GRS03, OAC12].



**Table 2:** Questionnaire of sense of speech agency and sense of voice ownership

Questionnaire			sentence
Sense of Speech Agency	SoSA1	Manip Voice	"It felt like I could control the character voice as if it was my own voice"
	SoSA2	Self Produced	"The utterance of the character voice were caused by my utterance"
	SoSA3	Produced Change	"I felt as if the utterance of the character voice were influencing my own utterance"
	SoSA4	Speaking	"I felt as if the character voice was speaking by itself"
Sense of Voice Ownership	SoVO1	Own Voice	"I felt as if the character voice was my voice"
	SoVO2	Other Voice	"It felt as if the character voice I heard was someone else"
	SoVO3	Two Voices	"It seemed as if I might have more than one voice"
Voice Similarity			"Did you felt like character voice from headphone?"
Speech Disorder			"Did you felt like speech disorder with voice feedback from headphone?"

## 5. Results

The results are presented in Table 3 and Table 4. We investigated the significant differences in the task scores between the pre and post conditions.

This study is an explorative study. In an unpublished paper [Cla20] by Yee et al., a reversal effect in the predicted direction was observed without user's knowledge of stereotype of avatars. Thus, we divided participants with complete data into two groups based on voice similarity, such as character voice (Character) and other voices (Others) in order to investigate the relationships between voice similarity scores (Character: Voice similarity Score over 5 points, Others: Voice similarity Score under 4 points.). Number of the participants was shown in Table 1.

### 5.1. Shyness-IAT Change

The results are shown in Figure 3. We conducted two-tailed Paired t-test, which indicated that shyness-IAT was not significantly more sociable in the post condition than in the pre-condition ( $t(29) = 0.964$ ,  $p = .172$ , Cohen's  $d = 0.166$ ). We divided voice similarity into character voice (Character) and other voices (Others) for analysis based on [Cla20]. We conducted a one-tailed Wilcoxon rank sum test. The results was shown in Figure 4. The results showed that the difference in Shyness-IAT with voice similarity was marginally significantly more sociable in the character voice than in the other voice ( $p = .078$ , Cohen's  $r = 0.259$ ).

### 5.2. ToL Change

The results are shown in Figure 5. We conducted two-tailed Paired t-test, which indicated that ToL Score was significantly higher in the post-condition than in the pre-condition ( $t(22) = 2.450$ ,  $p = .021$ , Cohen's  $d = 0.450$ ). We also investigated cosine similarity with voice similarity. This results was shown in Figure 6. Another one-tailed Student t-test showed that the D-Score with voice similarity was significantly higher in the character than in the others ( $t(6.4805) = 2.091$ ,  $p = .039$ , Cohen's  $d = 1.049$ ).

### 5.3. Big5 Change

The changes in Big-5 personality traits are presented in Table 4 and Table 5. We calculated the cosine similarity that between personality of character and participants with pre and post condition.

**Table 3:** Overview of questionnaire results.

Task	Mean	SE	
Voice Similarity	-1.067	0.342	
Speech Disorder	1.967	0.227	
SoSA	SoSA1	-1.567	0.266
	SoSA2	-0.967	0.316
	SoSA3	0.467	0.335
	SoSA4	-0.567	0.298
SoVO	SoVO1	-1.633	0.256
	SoVO2	1.167	0.292
	SoVO3	-0.233	0.313
Big-5(Pre)	Extraversion	4.117	0.246
	Agreeableness	4.700	0.210
	Conscientiousness	3.317	0.243
	Neuroticism	4.317	0.229
	Openness	5.017	0.182
Big-5(Post)	Extraversion	4.267	0.233
	Agreeableness	4.617	0.252
	Conscientiousness	3.367	0.250
	Neuroticism	4.467	0.217
	Openness	4.867	0.208
Big-5(Character)	Extraversion	6.433	0.177
	Agreeableness	5.367	0.192
	Conscientiousness	3.833	0.254
	Neuroticism	2.067	0.162
	Openness	6.383	0.157
Self-Esteem	26.967	0.422	
Locus of Control	21.333	0.599	

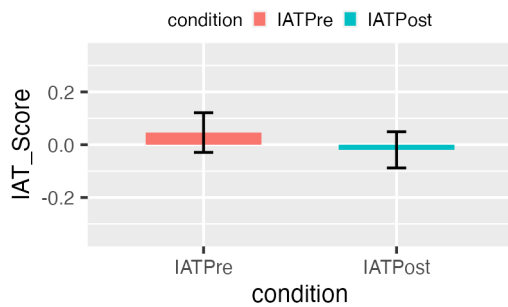
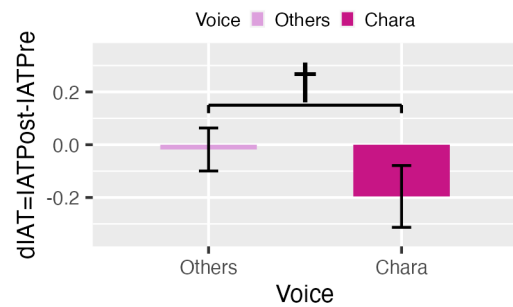
We conducted a Wilcoxon rank sum test, which revealed that the cosine similarity with post-condition was marginally significantly lower than the cosine similarity with pre-condition ( $p = .059$  Cohen's  $r = 0.244$ ). In addition, we conducted a two-tailed Wilcoxon rank sum test, which revealed that the cosine similarity with voice similarity. We conducted a Paired t-test on the results of character, which revealed that the cosine similarity with post-condition was not significantly higher than the cosine similarity with pre-condition ( $t(7) = 0.477$ ,  $p = .648$ , Cohen's  $d = 0.132$ ). Also, We conducted a Wilcoxon rank sum test on the results of others, which revealed that the cosine similarity with post-condition was marginally

**Table 4:** Overview of analysis task results. n.s.: not significant, †: Marginally Significant,  $p < 0.10$ , \*: Significant,  $p < 0.05$ 

Task		Mean	SD	Normality	Homoscedasticity	Test	P Value	Effect Size
Shyness IAT	Pre	0.046	0.075	Yes	—	Paired t-test	n.s	Cohen's d = 0.166
	Post	-0.019	0.069					
dIAT (PostIAT-PreIAT)	Character	-0.018	0.081	No	Yes	Wilcoxon rank sum test	†	Cohen's r = 0.259
	Others	-0.196	0.117					
Tower of London	Pre	34.870	1.375	Yes	—	Paired t-test	*	Cohen's d = 0.450
	Post	37.783	1.326					
dScore (PostToL-PreToL)	Character	1.722	1.234	Yes	Yes	Student t-test	*	Cohen's d = 1.049
	Others	7.200	2.311					
Speech-Agency and dScore	Lower-Agency	0.300	5.832	Yes	Yes	Student t-test	*	Cohen's d = 0.888
	Higher-Agency	4.923	4.681					
Self-Esteem and dScore	Lower-Esteem	1.286	1.669	Yes	No	Welch t-test	*	Cohen's d = 0.828
	Higher-Esteem	5.600	1.002					

**Table 5:** Overview of analysis Big-5 results with cosine similarity. n.s.: not significant, †: Marginally Significant,  $p < 0.10$ , \*: Significant,  $p < 0.05$ .

Task		Mean	SD	Normality	Test	P Value	Effect Size
All	Chara-Pre	0.908	0.011	No	Wilcoxon signed-rank test	†	Cohen's r = 0.244
	Chara-Post	0.899	0.012				
Character	Chara-Pre	0.922	0.015	No	Paired t-test	n.s	Cohen's d = 0.132
	Chara-Post	0.916	0.019				
Others	Chara-Pre	0.903	0.014	Yes	Wilcoxon signed-rank	†	Cohen's r = 0.251
	Chara-Post	0.893	0.014				

**Figure 3:** Results of Shyness-IAT change. Error bars represent the standard errors.**Figure 4:** Results of Shyness-IAT change with voice similarity. Error bars represent the standard errors.

significantly lower than the cosine similarity with pre-condition ( $p = .095$ , Cohen's  $r = 0.251$ ).

#### 5.4. Voice Embodiment and ToL

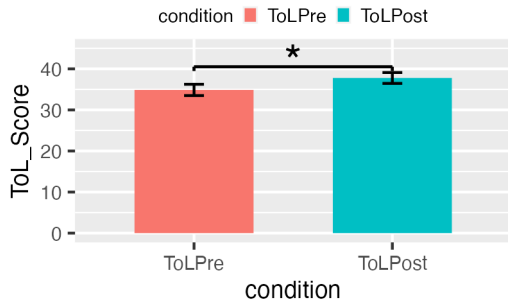
We also investigated the correlations among SoSA, SoVO, Shyness-IAT, and ToL scores. The results are shown in Table 6.

It shown that spearman correlation indicated that there was association between SoSA and dScore. The relationship between SoSA and dScore was analyzed by classifying the participants into higher and lower SoSA groups based on the mean of the results. The results are shown in Figure 7 and Figure 8. We conducted two-

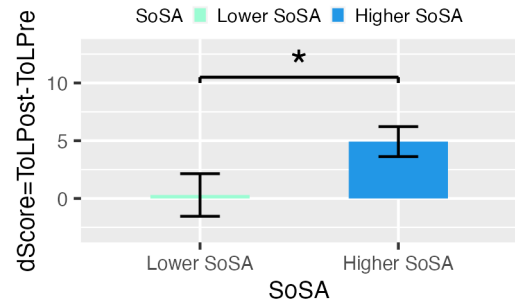
tailed Student t-test, which indicated that dScore was significantly higher in the Higher SoSA than in the Lower SoSA ( $t(21) = 2.111$ ,  $p = .0469$ , Cohen's  $d = 0.888$ ).

**Table 6:** Questionnaire of sense of speech agency and sense of voice ownership: CC: correlation coefficient. n.s.: not significant, \*: Significant,  $p < 0.05$ 

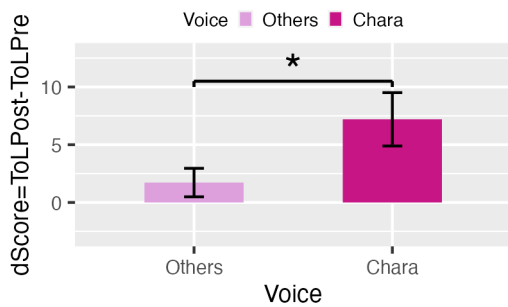
	dIAT		dScore	
	CC	P value	CC	P value
SoSA	0.089	n.s	0.342	*
SoVO	-0.110	n.s	-0.009	n.s



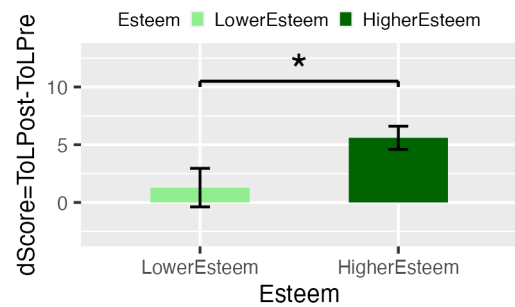
**Figure 5:** Results of ToL change with trial. Error bars represent the standard errors.



**Figure 7:** Results of relationship between sense of speech agency and ToL change. Error bars represent the standard errors.



**Figure 6:** Results of ToL change with voice similarity. Error bars represent the standard errors.



**Figure 8:** Results of relationship between self-esteem and ToL change. Error bars represent the standard errors.

### 5.5. Self-Esteem and ToL

The relationship between self-esteem and ToL scores was analyzed by classifying the participants into higher and lower self-esteem groups based on the mean of the results. The results are presented in Table 4 and Figure 8. We conducted a two-tailed Welch t-test, which showed that the dScore was significantly higher in the higher self-esteem group than in the lower self-esteem group ( $t(20.261) = 2.216$ ,  $p = .038$ , Cohen's  $d = 0.828$ ).

### 5.6. Free Description

Free description was provided by the participants after these experiments. We classified it divided into character and others.

*Character 1:* Due to the own voice feedback being delayed and not matching utterances, we think it was difficult to read long sentences. I enjoyed that my own voice was transformed like character's voice with my own utterance, and I felt a little like transforming into the character. Therefore, I feel like my articulation changed for pronunciation with the change in voice quality to represent the character's voice.

This description includes positive and negative comments on character voice embodiment with VOI.

*Character 2:* It is a difficult task to read sentences and listen to

the converted own voice in real-time. I think that voice actors form the characteristics of characters with various voice acts. I found that the converted voice had a similar voice quality and speech style to the character for the first time.

The voice similarity of the character enhanced voice embodiment and articulation of voice quality with VOI.

Otherwise, others wrote the following:

*Others 1:* I feel the voice quality is unknown, which seems to inhibit the articulation of utterance.

*Others 2:* I feel uncomfortable that my own voice was transformed into an unknown character voice, making it difficult to read sentences. I cannot read the sentence in the speech style of the character.

*Others 3:* A rare noise in the voice was noticeable while using the voice conversion system.

These comments suggested that voice quality and voice similarity impacts on the sense of immersion in VOI.

## 6. Discussion

We discussed the results in relation to the following hypotheses outlined below: RQ1: The results indicated that VOI potentially influenced implicit sociability and enhanced planning ability. RQ2: We

confirmed that SoSA and self-esteem are related to planning ability. Future studies should consider voice similarity and self-esteem in the context of the Proteus effect [BKS18].

### 6.1. Shyness-IAT Change (RQ1-1)

The change in the Shyness-IAT indicated that voice similarity is a crucial factor in VOI. Given the quality of voice conversion, these results suggested that voice similarity should be incorporated into the experimental design of VOI in future studies. The Shyness-IAT show a marginally significantly change in voice similarity to the image of character in accordance with character stories with VOI. From these results, we should considered imbalance in the sample size in future study. We calculated the sample size for within-subject design with pre-post conditions to investigate the effect of VOI. However, we also used a between-subject design based on voice similarity. Therefore, there was an imbalance results in the sample size due to owing variations in the perception of voice similarity. These reasons suggested that we should improve the similarity of the voice conversion algorithm to achieve a balanced proportion of the sample size in the experimental design. Based on these aspects, the voice conversion algorithm should be updated to better investigate the changes in the Shyness-IAT in future study.

## 6.2. ToL Change

### 6.2.1. RQ1-2

The change in ToL scores indicated a significant effect on planning ability based on the difference between Pre and Post condition ToL score. In addition, the planning ability was significantly effected by voice similarity. These results suggested that character voice embodiment enhanced planning ability following VOI. This indicated that VOI might have a similar effect to that of the Einstein avatar in terms of the Proteus Effect [BKS18].

### 6.2.2. RQ2-1

There is a positive correlation between SoSA and the planning ability. This results suggested that higher SoSA has a more significantly affect planning ability than lower SoSA. SoSA might influence the Proteus Effect in conjunction with body ownership illusion. This finding indicated that the first indication that SoSA enhanced the cognitive ability similarly to body ownership illusion in VOI. It is recommended that future experimental designs include questionnaires on SoSA and SoVO for VOI. Because a validated questionnaire is not yet available. Therefore, it should develop to explore the detailed relationship between the parameters(such as voice similarity, voice quality, latency and so on.) of VOI and Parakeet Effect in future study.

### 6.2.3. RQ2-2

We investigated the relationship between self-esteem and planning ability. Previous research suggested that higher self-esteem has a more significantly affect planning ability than lower self-esteem. However, our findings showed a different trend [BKS18]: lower self-esteem had a more pronounced effect on planning ability than higher self-esteem. Due to the unbalanced sample size related to

voice similarity, we were could not confirm the relationship between planning ability and self-esteem within the same experimental design related to the Proteus Effect [BKS18]. Future studies should explore this discrepancy and investigate the trends observed in the Proteus Effect.

### 6.3. Big5 Change (RQ1-3)

Regarding the Big-5 personality traits, the cosine similarity between the participants and the character did not change significantly. Therefore, this study did not support the hypothesis that the Big-5 traits would change with stimulated converted voice. However, this study supported the hypothesis in line with the commonly accepted theory.

## 6.4. Limitation and Future Work

From these results, we recognize the limitations in assessing cognitive ability due to voice similarity and the latency of voice feedback. Whereas voice similarity might enhance the planning ability, we were could not conduct investigations with a balanced experimental design that considered voice similarity because of the current quality of the voice conversion system. Therefore, we should improve the voice similarity of the converted voice using a voice conversion system, and investigate the Parakeet Effect based on balanced experimental design. Additionally, the latency of voice feedback effects the perception of speech disorders. Latency might also be related to agency and ownership, as observed in body ownership illusions [RL20]. The latency should be reduced by refining the voice conversion algorithm. To address these issues, we should improve both voice similarity and latency in the voice conversion system. We identified Stream-VC [YKL\*24], which provides high-quality voice similarity and low latency for voice conversion. Stream-VC uses HuBERT [HBT\*21] to generate high-quality converted voices with minimal participant data, similar to few-shot learning with soft speech units. Future studies should investigate the effectiveness of VOI using low-latency and high-quality voice conversion systems.

## 7. Conclusion

We investigated the Parakeet effect, focusing on how voice embodiment impacts implicit attitude, planning ability, and personality. As a result, we confirmed that the planning ability was significantly enhanced by the voice ownership illusion when using a character's voice through a real-time voice conversion system. Furthermore, we also confired that voice similarity was a crucial factor for the Parakeet effect. Based on these findings, the voice conversion algorithm should be updated to further investigate the Parakeet effect, considering the quality of voice conversion system.

## 8. Acknowledgement

This work was supported by by JST Moonshot Research Development (JPMJMS2013) and JST SPRING (JPMJSP2108).



## References

- [ABM02] ASENDORPF J., BANSE R., MÜCKE D.: Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology* 83, 2 (09 2002), 380–393. doi:10.1037//0022-3514.83.2.380. 4
- [AF11] AIKAWA A., FUJII T.: Using the implicit association test (iat) to measure implicit shyness. *The Japanese Journal of Psychology* 82, 1 (2011), 41–48. doi:10.4992/jjpsy.82.41. 4
- [AJH\*16] AUCOUTURIER J.-J., JOHANSSON P., HALL L., SEGNINI R., MERCADIÉ L., WATANABE K.: Covert digital manipulation of vocal emotion alters speakers' emotional states in a congruent direction. *Proceedings of the National Academy of Sciences* 113, 4 (2016), 948–953. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1506552113>, arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.1506552113>, doi:10.1073/pnas.1506552113. 1, 3, 4
- [AKT\*21] ARAKAWA R., KASHINO Z., TAKAMICHI S., VERHULST A., INAMI M.: Digital speech makeup: Voice conversion based altered auditory feedback for transforming self-representation. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (New York, NY, USA, 2021), ICMI '21, Association for Computing Machinery, p. 159–167. URL: <https://doi.org/10.1145/3462244.3479934>, doi:10.1145/3462244.3479934. 1, 3
- [ATS19] ARAKAWA R., TAKAMICHI S., SARUWATARI H.: Implementation of dnn-based real-time voice conversion and its improvements by audio data augmentation and mask-shaped device. In *Proc. 10th ISCA Workshop on Speech Synthesis (SSW 10)* (2019), pp. 93–98. doi:10.21437/SSW.2019-17. 3
- [BB02] BERG W., BYRD D.: The tower of london spatial problem-solving task: Enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology* 24, 5 (09 2002), 586–604. doi:10.1076/jcen.24.5.586.1006. 4
- [BC98] BOTVINICK M., COHEN J.: Rubber hands 'feel' touch that eyes see. *Nature* 391, 6669 (Feb. 1998), 756. URL: <https://doi.org/10.1038/35784>, doi:10.1038/35784. 2
- [BG22] BREDIKHINA L., GIARD A.: Becoming a virtual cutie: Digital cross-dressing in japan. *Convergence* 28, 6 (2022), 1643–1661. URL: <https://doi.org/10.1177/13548565221074812>, arXiv:<https://doi.org/10.1177/13548565221074812>, doi:10.1177/13548565221074812. 1
- [BGS13] BANAKOU D., GROTEN R., SLATER M.: Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences* 110, 31 (2013), 12846–12851. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1306779110>, arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.1306779110>, doi:10.1073/pnas.1306779110. 2
- [BKS18] BANAKOU D., KISHORE S., SLATER M.: Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias. *Frontiers in Psychology* 9 (2018). URL: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.00917>, doi:10.3389/fpsyg.2018.00917. 2, 4, 8
- [BS14] BANAKOU D., SLATER M.: Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *Proceedings of the National Academy of Sciences* 111, 49 (2014), 17678–17683. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1414936111>, arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.1414936111>, doi:10.1073/pnas.1414936111. 2
- [CJC\*18] COSTA J., JUNG M. F., CZERWINSKI M., GUIMBRETIERE F., LE T., CHOUDHURY T.: Regulating feelings during interpersonal conflicts by changing voice self-perception. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), CHI '18, Association for Computing Machinery, pp. 1–13. URL: <https://doi.org/10.1145/3173574.3174205>, doi:10.1145/3173574.3174205. 1, 3
- [Cla20] CLARK O.: How to kill a greek god: A meta-analysis and critical review of 14 years of proteus effect research. *Doctor Thesis* (oct 2020). URL: <https://doi.org/10.31219/osf.io/z5kf8>, doi:10.31219/osf.io/z5kf8. 2, 5
- [Ehr07] EHRSSON H. H.: The experimental induction of out-of-body experiences. *Science* 317, 5841 (2007), 1048–1048. URL: <https://www.science.org/doi/abs/10.1126/science.1142175>, doi:10.1126/science.1142175. 2
- [FELB07] FAUL F., ERDFELDER E., LANG A.-G., BUCHNER A.: G\*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39 (2007), 175–191. doi:10.3758/BF03193146. 4
- [GFP18] GONZALEZ-FRANCO M., PECK T. C.: Avatar embodiment towards a standardized questionnaire. *Frontiers in Robotics and AI* 5 (2018). URL: <https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2018.00074>, doi:10.3389/frobt.2018.00074. 4
- [GMS98] GREENWALD A. G., MCGHEE D. E., SCHWARTZ J. L. K.: Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74, 6 (1998), 1464–1480. URL: <https://api.semanticscholar.org/CorpusID:7840819>. 4
- [GNB03] GREENWALD A. G., NOSEK B. A., BANAJI M. R.: Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of Personality and Social Psychology* 85, 2 (08 2003), 197–216. doi:10.1037/0022-3514.85.2.197. 4
- [GRS03] GOSLING S. D., RENTFROW P. J., SWANN W. B.: A very brief measure of the big-five personality domains. *Journal of Research in Personality* 37, 6 (2003), 504–528. URL: <https://www.sciencedirect.com/science/article/pii/S0092656603000461>, doi:10.1016/S0092-6566(03)00046-1. 4
- [HA84] HOWELL P., ARCHER A.: Susceptibility to the effects of delayed auditory feedback. *Perception Psychophysics* 36, 3 (1984), 296–302. URL: <https://doi.org/10.3758/BF03206371>, doi:10.3758/BF03206371. 3
- [HBT\*21] HSU W.-N., BOLTE B., TSAI Y.-H., LAKHOTIA K., SALAKHUTDINOV R., MOHAMED A.: Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* PP (10 2021), 1–1. doi:10.1109/TASLP.2021.3122291. 8
- [IS16] ISMAIL M. A. F., SHIMADA S.: 'robot' hand illusion under delayed visual feedback: Relationship between the senses of ownership and agency. *PLoS ONE* 11, 7 (2016), e0159619. URL: <https://doi.org/10.1371/journal.pone.0159619>, doi:10.1371/journal.pone.0159619. 3
- [JAAL18] JEUNET C., ALBERT L., ARGELAGUET F., LÉCUYER A.: "do you feel in control?": Towards novel approaches to characterise, manipulate and measure the sense of agency in virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1486–1495. doi:10.1109/TVCG.2018.2794598. 4
- [KPL23] KIM H., PARK J., LEE I.-K.: "to be or not to be me?": Exploration of self-similar effects of avatars on social virtual reality experiences. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4794–4804. doi:10.1109/TVCG.2023.3320240. 2
- [KTS\*90] KUREMATSU A., TAKEDA K., SAGISAKA Y., KATAGIRI S., KUWABARA H., SHIKANO K.: Atr japanese speech database as a tool of speech recognition and synthesis. *Speech Communication* 9, 4 (1990), 357–363. URL: <https://www.sciencedirect.com/science/article/pii/016763939090011W>, doi:10.1016/0167-6393(90)90011-W. 3

- [KUS11] KALLER C., UNTERRAINER J., STAHL C.: Assessing planning ability with the tower of london task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment* 24, 1 (08 2011), 46–53. doi:10.1037/a0025174. 4
- [LTMB07] LENGGENHAGER B., TADI T., METZINGER T., BLANKE O.: Video ergo sum: Manipulating bodily self-consciousness. *Science* 317, 5841 (2007), 1096–1099. URL: <https://www.science.org/doi/abs/10.1126/science.1143439>, doi:10.1126/science.1143439. 2
- [LYB\*19] LOPEZ S., YANG Y., BELTRAN K., KIM S. J., HERNANDEZ J. C., SIMRAN C., YANG B., YUKSEL B. F.: Investigating implicit gender bias and embodiment of white males in virtual reality with full body visuomotor synchrony. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, Association for Computing Machinery, pp. 1–12. URL: <https://doi.org/10.1145/3290605.3300787>, doi:10.1145/3290605.3300787. 2, 4
- [MG07] MIMURA C., GRIFFITHS P.: A japanese version of the rosenberg self-esteem scale: Translation and equivalence assessment. *Journal of Psychosomatic Research* 62, 5 (2007), 589–594. URL: <https://www.sciencedirect.com/science/article/pii/S002239990600506X>, doi:https://doi.org/10.1016/j.jpsychores.2006.11.004. 4
- [MS13] MASELLI A., SLATER M.: The building blocks of the full body ownership illusion. *Frontiers in Human Neuroscience* 7 (2013). URL: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2013.00083>, doi:10.3389/fnhum.2013.00083. 2
- [OAC12] OSHIO A., ABE S., CUTRONE P.: Development, reliability, and validity of the japanese version of ten item personality inventory (tipi-j). *The Japanese Journal of Personality* 21, 1 (2012), 40–52. doi:10.2132/personality.21.40. 4
- [OAI22] OHATA R., ASAI T., IMAIZUMI S., IMAMIZU H.: I hear my voice; therefore i spoke: The sense of agency over speech is enhanced by hearing one's own voice. *Psychological Science* 33, 8 (2022), 1226–1239. PMID: 35787212. URL: <https://doi.org/10.1177/09567976211068880>, arXiv:https://doi.org/10.1177/09567976211068880, doi:10.1177/09567976211068880. 2
- [OBN24] OGAWA N., BABA J., NAKANISHI J.: Investigating effect of altered auditory feedback on self-representation, subjective operator experience, and task performance in teleoperation of a social robot. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2024), CHI '24, Association for Computing Machinery. URL: <https://doi.org/10.1145/3613904.3642561>, doi:10.1145/3613904.3642561. 3
- [PGB20] PECK T. C., GOOD J. J., BOURNE K. A.: Inducing and mitigating stereotype threat through gendered virtual body-swap illusions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2020), CHI '20, Association for Computing Machinery, pp. 1–13. URL: <https://doi.org/10.1145/3313831.3376419>, doi:10.1145/3313831.3376419. 2
- [PSAS13] PECK T. C., SEINFELD S., AGLIOTI S. M., SLATER M.: Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition* 22, 3 (2013), 779–787. URL: <https://www.sciencedirect.com/science/article/pii/S1053810013000597>, doi:https://doi.org/10.1016/j.concog.2013.04.016. 2, 4
- [RBB13] ROSENBERG R. S., BAUGHMAN S. L., BAIENSON J. N.: Virtual superheroes: Using superpowers in virtual reality to encourage prosocial behavior. *PLoS ONE* 8, 1 (2013), e55003. URL: <https://doi.org/10.1371/journal.pone.0055003>, doi:10.1371/journal.pone.0055003. 2
- [RL20] ROTH D., LATOSCHIK M. E.: Construction of the virtual embodiment questionnaire (veq). *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3546–3556. doi:10.1109/TVCG.2020.3023603. 3, 4, 8
- [SBW82] SHALLICE T., BROADBENT D. E., WEISKRANTZ L.: Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 298, 1089 (1982), 199–209. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.1982.0082>, arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.1982.0082, doi:10.1098/rstb.1982.0082. 4
- [STOI23] SAKUMA H., TAKAHASHI H., OGAWA K., ISHIGURO H.: Immersive role-playing with avatars leads to adoption of others' personalities. *Frontiers in Virtual Reality* 4 (2023). URL: <https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2023.1025526>, doi:10.3389/frvir.2023.1025526. 1, 4
- [WSH\*16] WALTEMATTE T., SENNA I., HÜLSMANN F., ROHDE M., KOPP S., ERNST M., BOTSCH M.: The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (New York, NY, USA, 2016), VRST '16, Association for Computing Machinery, p. 27–35. URL: <https://doi.org/10.1145/2993369.2993381>, doi:10.1145/2993369.2993381. 3, 4
- [YB07] YEE N., BAIENSON J.: The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research* 33, 3 (July 2007), 271–290. URL: <https://doi.org/10.1111/j.1468-2958.2007.00299.x>, arXiv:https://academic.oup.com/hcr/article-pdf/33/3/271/22324746/jhumcom0271.pdf, doi:10.1111/j.1468-2958.2007.00299.x. 1, 2
- [YKL\*24] YANG Y., KARTYNNIK Y., LI Y., TANG J., LI X., SUNG G., GRUNDMANN M.: Streamvc: Real-time low-latency voice conversion. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2024), pp. 11016–11020. doi:10.1109/ICASSP48485.2024.10446863. 8
- [ZMMJ11] ZHENG Z. Z., MACDONALD E. N., MUNHALL K. G., JOHNSRUDE I. S.: Perceiving a stranger's voice as being one's own: A 'rubber voice' illusion? *PLoS ONE* 6, 4 (2011), e18655. URL: <https://doi.org/10.1371/journal.pone.0018655>, doi:10.1371/journal.pone.0018655. 2