EasyChair Preprint

№ 2181

Investigations of Verbal Cues and Self-Voice Perception Model

Aibao Zhou, Yanbing Hu, Xiaoyong Lu, Yu Li, Xianya Zhang and Pan Tao

December 16, 2019

Investigations of Verbal Cues and Self-Voice Perception Model

1st Aibao Zhou
School of Psychology
Northwest Normal University
Lanzhou, China
zhouab@nwnu.edu.cn

2nd Yanbing Hu
School of Psychology
Northwest Normal University
Lanzhou, China
hybpsy2018@163.com

3rd Xiaoyong Lu
School of Physics and Electronic Engineering
Northwest Normal University
Lanzhou, China
luxy@nwnu.edu.cn

4thYu Li
School of Psychology
Northwest Normal University
Lanzhou, China
2562109082@qq.com

5th Xianya Zhang
School of Psychology
Northwest Normal University
Lanzhou, China
1835288258@qq.com

6th Tao Pan
Department of information engineering
Lanzhou Resources and Environment
Voc-Tech College
Lanzhou, China
pant_revt@126.com

***Abstract*—It is of great significance to explore the influence mechanism of verbal cues on self-voice perception under different conditions. The aim of this study is to probe the effects of verbal and non-verbal conditions on implicit task (experiment 1) and explicit task (experiment 2) of self-voice. Binaural auditory channels were used to present two voices successively, and the speakers included the participants, familiar and strangers. In the experiment 1, participants were asked to judge whether the pair of stimuli were pronounced by the same speaker. In the experiment 2, participants were asked to identify whether one of the stimuli was their own voice. The results show that there is no difference in individual voice perception of self (vs. non-self) under different conditions of verbal cues in the implicit task. In the explicit task, there is less accuracy in recognizing the voice of self than that of non-self. It is also found that there exists no difference in individual self-voice perception at different levels of factors.**

Keywords—self-voice; perception mechanism; voice perception; verbal cues; self-disadvantage effect

I. INTRODUCTION

Self-voice perception (the capacity to recognize physical and mental aspects of oneself) is a highly developed ability in humans that underlies a range of social and interpersonal functions, such as the theory of mind and introspection [1]. Voice is one of the symbols of individual identification, which contains not only emotional information, but also identity information and speech information [2]. From the voice, individuals can not only know what the speaker wants to express, but also know identity information such as the age and gender of the speaker, and even know the immediate emotions of the speaker. Self-voice perception is crucial for an individual's self-awareness and self-monitoring. Its destruction may have a harmful effect on mental health and may have a negative impact on a person's quality of life [3, 4].

Individual self-monitoring of voice uses neural feedback mechanisms to distinguish self from non-self [5]. The ability of an individual to distinguish between self and non-self vocal cues is a fundamental aspect of self-awareness which assists in self-monitoring in verbal communication [6]. Previous studies have illuminated the connection identifying non-self identities to the degree of familiarity (1) individuals have different brain mechanisms for non-self identity perception with different familiarity degrees. In other words, individuals' right superior temporal sulcus are activated when they recognize the voice of acquaintances (vs. the voice of strangers) [7]; (2) individuals' familiarity of the speaker can affect their judgment of voice identity [8].

Previous studies have shown that there are two different cognitive processing modes of self-voice perception. The results show that there is no significant difference in the accuracy of self-voice perception (vs. non-self) in the implicit task. In the explicit task, the accuracy of self-voice perception is significantly lower than that of non-self voice perception [9, 10]. The reason lies in the fact that there are two conduction pathways among individuals in the perception of their own voice: bone conduction and air-conduction. The individual perceives the self-voice in the recording only includes the air-conducted voice. The absence of bone conduction changes the characteristics of voice, so there is a difference the cognitive ability between individuals who listen to recorded self-voices and those who listen to self-voices on a daily basis [11].According to self-monitoring theory, individuals use a day-to-day perceptual representation of their own voices to monitor if they are self-voice in recordings [12].Therefore, due to the mismatch between the self-voice in the recording and in daily life, there is a disadvantage effect in the individual perception of the self-voice (vs. non-self) in the recording. By filtering out the voice signal, that is, only keeping the frequency higher than the third resonance peak, it is found that individuals have a processing advantage in self-voice perception [13]. Therefore, the lack of acoustic information may reduce the influence of bone conduction on the individual's perception of the self-voice in the recording, and thus reduce the degree of an individual's monitoring of the self-voice in the recording.

From the macro point of view, there are two basic processes of identity processing. Firstly, individuals can distinguish between different identities; Secondly, individuals can perceive their identity constancy from different physical environments [14].Non-verbal (vs. verbal) contains less information, which may reduce the individual's perception of the difference between the self-voice in the recording and the daily self-voice. Therefore, it is necessary to consider non-speech in voice perception. From the perspective of voice perception research, both verbal and non-verbal cues have an impact on individual voice perception processing, that is, individual voice perception processing is likely to be a collaborative processing process of bottom-up acoustic analysis and top-down voice processing [15].There are two main deficiencies in previous studies :(1) the research focus is confined to the perception of different voice identities, ignoring the

perceptual constancy of voice perception; (2) the influence of verbal cues on cognitive processing of implicit and explicit tasks in self-voice perception is not considered .

This study adopted [9] both implicit and explicit task perception paradigm of voice to discuss individual self-voice perception mechanism of different verbal cues. In experiment 1, participants were asked to judge whether the pair of stimuli were pronounced by the same speaker. In experiment 2, the participants were asked to identify if one of the stimuli was their own voice. This paper puts forth the study hypotheses are as follows：(1) In the implicit task: There is no difference between the self and non-self in the verbal cues；(2) In the explicit task: Individuals have processing advantages in non-self identity (vs. self-voice perception) in verbal condition. There is no difference between self-identity and non-self identity in non-verbal condition. (3) Individual self-voice perception (vs. non-self voice perception) has perceptual constancy.

II. EXPERIMENT 1

A. Participants and procedure

Thirty college students from a certain university participated in this experiment. Among them, 12 male students, aged (20.7±2.20), had no hearing impairment and obtained mandarin proficiency certificate. All the participants were of Han ethnicity whose mother tongue was Mandarin, without the knowledge of the purpose of the experiment. In this study, 2 college students (1 male and 1 female) were randomly selected to receive experimental materials from strangers. According to [9] in the study of implicit task content difference of voice perception to various stimuli (ηp2 = 0.35), the effect of using G*Power3.1, set Power to 95%, the alpha level of 0.05, calculate sample size of 15 participants.

2.2. Stimuli

Prepare: First, a week before the experiment, the participants were recorded in stereo mode in a selected recording room using a portable digital recording device (Roland r-26). Recording process is divided into two sections, the first stage to let the participants sit in front of the screen to 60 cm, lips to 10 cm, the distance to the microphone asked participants as clearly as possible, read quietly appeared on the screen of the text, a text after the need to pause and then to the next text recording, such as was non-self accordance with the requirements for the recording, insist on the recording again. At the same time, the participants were required to bring a person of the same sex who had been with them for at least one year, and record them. The previous procedure was repeated, and the voice material was defined as the voice of acquaintances. Then, two randomly selected college students (who did not know each each) were recorded. The recording procedure was the same as the recording process of the participants, whose voice material was defined as the voice of strangers.

According to Candini, et al. [9], stimuli all belong to the same semantic system. For example, the words used all belong to "animal". The words used in this study contain two Chinese characters and are high-frequency words as the language condition. According to the research of Conde, et al. [6], vocalization /a/ and /i/ are taken as non-verbal conditions in this study. Voice material is recorded in stereo with sampling rate of 44100Hz and 16bit. After recording, the loudness of the voice materials was standardized by Praat 5.3.56 software (root mean square amplitude -RMS=70dB). The duration of voice stimulation was (588.63±58.21). The duration of non-verbal stimulation was (442.37±45.26). Each voice stimulus consists of three types: (A) participants' voice (B) acquaintances' voice, (C) strangers' voice.

B. Procedure

In the experiment, a 21-inch computer monitor with a screen resolution of 1024×768 and a refresh rate of 100Hz was used to display all stimuli using E-prime2.0 (Psychology Software Tool @1996-2012). The participants sat 60cm in front of the computer screen. At the start of the first present the fixation "+" in the middle of the screen (the time for 500 ms), then in what participants wearing headphones (audio technica ATH - FC700) in turn in two voices, the voice interval is 500ms, voice broadcast, requirement of button, button for "yes" to a response, a button for "no" (two buttons "J" and "F", respectively), the balance between button. Each trial run is 3000ms.

The trial consists of two voice stimuli emitted by the same individual or by two different individuals. Therefore, three identical voice stimuli were (AA-BB-CC) and three different voices were (AB-AC-BC), and the gender of the participant matched the gender of the voice stimulus. Half of the same and half of the different stimuli are made up of the same words and the same vocalization (e.g. /a/-/a/), the non-self half is made up of different words and different vocalization (e.g. /a/-/i/). Words and vocalization exist in four blocks, each of which contains 2(stimulus)*2(verbal cue)*6(voice owner). There are 24 experimental conditions and they are presented randomly. The two blocks (words and vocalization) are balanced between the participants.

In experiment 1, the participants were asked to judge whether the two voice stimuli in sequence were from the same individual or not. The experiment consisted of 8 practice trials and 96 formal trials. The whole experiment lasts about 15 minutes.

III. RESULTS

A. Implicit task: self vs. non-self

Due to the non-keystroke response of one participant in the implicit task, invalid data were eliminated, leaving 29 valid participant data (12 male). SPSS21.0 software was used for statistical analysis of the data.

Based on previous studies, we defined the self-condition as containing at least one participant's voice (AA-AB-AC), and the non-self-condition as excluding the participant's voice (BB-CC-BC) [9 163, 10 164, 16]. For the implicit task of self-voice perception, the variance analysis of repeated measures was conducted in participants with 2(self vs. non-self)×2(words vs. vocalization)×2(same stimulus vs. different stimulus). Results show that no significant main effect of identity, $F(1,28)=0.11$, $p>0.05$, stimulus content of main effect significantly, $F(1,28)=10.97, p<0.01$, $\eta_p^2=0.32$, the main effect of verbal cues significant $F(1,28)=6.20, p<0.05$, $\eta_p^2=0.18$ The interaction of the identity and stimulate the content significantly, $F(1,28)=18.25, p<0.001$, $\eta_p^2=0.40$. Simple effect, found that under the condition of non-self, the same time is more than stimulate different accuracy, $F(1,28)=22.99, p<0.001$, $\eta_p^2=0.45$. Under the self-condition, there was no significant difference in the accuracy of the same stimulation (vs. different stimuli), $F(1,28)=0.41$, $p>0.05$.

B. Implicit task: same owner-different owner

In order to better understand the relationship between different factors and self-voice perception, we will separate the identity (same speaker/different speaker). The reason is that there is one variation owner in the condition of same voice owner (AA-BB-CC) and two variation owners in the condition of different voice owners (AB-AC-BC).

C. Implicit task: same owner

Variance analysis of repeated measures of 3(self-self vs. familiar-familiar vs. stranger-stranger) × 2(words vs. vocalization)

× 2(same stimuli vs. different stimuli) in participants. The Results show that no significant main effect of identity, F(1,28) =0.66， p>0.05, stimulate the content of main effect significantly, F(1,28)=13.21， p<0.01， η_p^2=0.32, verbal cues no significant main effect of F(1,28)=0.20， p>0.05. Stimulate the interaction content and verbal cues, F(1,28) =6.22， p<0.05， η_p^2=0.18. Simple effect, found in the same condition, the words of the time is greater than the vocalization of the time, F(1,28)= 12.32， p<0.05， η_p^2=0.17.

D. Implicit task: different owner

Variance analysis of repeated measures of 3(self-familiar vs. self-stranger vs. familiar-stranger)×2(word vs. vocalization)×2(same owner vs. different owner) in participants. Results show that no significant main effect of identity, F(1,28)=0.33,p>0.05, no significant main effect to stimulate content, F(1,28)=0.78,p>0.05, the main effect of verbal cues significant F(1,28)=12.17,p<0.01, η_p^2=0.30.

E. Discussion

The first result confirms the hypothesis that there was no difference between self and non-self voice identification in the implicit task. At the same time, the results show that both the stimulus content and the verbal cue can affect the individual's voice identification in the implicit task. Specifically, there is a significant advantage in the accuracy of the same stimulus (vs. different stimulus) and the word (vs. vocalization), that is, the same stimulus and the verbal condition are more conducive to the individual's voice identification. From the interaction between identity and stimulus content, we can find that the self voice (vs. non-self voice) displays more perceptual constancy in the implicit task of individuals. It can be seen from the results of the same condition of the voice owner in the implicit task that the stimulus content (vs. verbal cue) has a greater impact on the individual's perception of voice identity.

In the implicit task, the results of different voice owner conditions show the opposite result, that is, verbal cues (vs. stimulus content) have greater influence on the individual's perception of voice identity. To some extent, this proves that there are two different basic processes of voice identity perception, that is, the stimulus content has a greater impact on the consistency of individual voice identity, while the verbal cues has a greater impact on individual voice identity. The result of the first experiment are consistent with [9]. It proves that individuals with Chinese cultural background also have implicit processing of self-voice perception. Experiment 2 will further explore whether individuals with Chinese cultural background have explicit self-voice perception processing and whether individuals still have perceptual constancy of self-voice perception.

IV. EXPERIMENT 2

A. Participants and procedure

It is the same as Experiment1.

V. RESULTS

A. Explicit task: self vs. non-self

All 30 participants were operated in accordance with the experimental instructions, and the data of 30 (10 male) valid participants were statistically analyzed by using SPSS21.0 software. The definition of self-non-self is the same as experiment 1. For the explicit task of self-voice perception, the variance analysis of repeated measures in participants was conducted by 2(self vs. non-self)×2(words vs. vocalization)×2(same stimuli vs. different stimuli).

The results show that the identity of main effect significantly, F(1,29)=6.16, p<0.05, η_p^2=0.18;Stimulate the content of main effect significantly, F(1,29)=5.12,p<0.05, η_p^2=0.25, the main effect of verbal cues significant F(1,29)=4.20,p<0.05, η_p^2=0.15. The interaction of the identity and verbal cues significant F(1,29)=5.22, p<0.05,ηp2=0.15. Simple effect, found that under the condition of non-self, words correctly is greater than the vocalization of the time, F(1,29)=12.37, p<0.01,ηp2=0.30. Under the self-condition, there was no significant difference in the accuracy of words (vs. vocalization), F(1,29)=0.003, p>0.05. Simple effect, also found that under the condition of the word, not my accuracy is greater than the accuracy of the self, F(1,29)=12.71,p<0.01, η_p^2=0.31. Under the vocalization condition, there is no significant difference between the accuracy of non-self and self, F(1,29)=0.83, p>0.05.

B. Explicit task: same owner

Variance analysis of repeated measures of 3(self-self vs. familiar-familiar vs. stranger-stranger) ×2(words vs. vocalization) ×2(same stimuli vs. different stimuli) in participants.

Results show that the main effect of verbal cues is significant, F(1,29)=7.48,p<0.05, η_p^2=0.21. Identity and verbal cue interaction significantly, F(1,29)=10.39,p<0.01, η_p^2=0.26. Simple effect, found that under the condition of the familiar- familiar with words correctly is greater than the vocalization, F(1,29)= 10.27,p<0.01, η_p^2=0.26; In the stranger-stranger: words correctly is greater than the vocalization, F(1,29)= 9.53，p<0.01，ηp2=0.25; In the self-self: there was no significant difference in the accuracy of words (vs vocalization), F(1,29)=1.26， p>0.05.Implicit task: same owner.

C. Explicit task: different owner

Variance analysis of repeated measures of 3(self-familiar vs. self-stranger vs. familiar-stranger) ×2(word vs. vocalization) ×2(same owner vs. different owner) in participants. The results showed that Identity, stimulate the interaction content and verbal cues, F(1,29)=5.60， p<0.01， η_p^2=0.16; Simple effect shows that under the condition of different stimulating - words, familiar to strange (vs. self - familiar with) accuracy, F(1,29)=5.74， p<0.05, η_p^2=0.17.

D. Discussion

The results of experiment 2 show that there is a cognitive advantage in the non-self voice perception (vs. self-voice perception) in the explicit task. It is found that the stimulus content and verbal cues in the explicit task will have an impact on the individual self-voice perception. Specifically, different stimuli (vs. identical stimuli) and words (vs. vocalization) were more accurate. In addition, the interaction between identity and verbal cues was found, that is, the accuracy rate of non-self (vs. self) was higher under verbal conditions, but there was no difference between self and non-self under non-verbal conditions. It seems that the processing of self-voice perception in explicit task is different from implicit task.

In the explicit task, under the same condition of voice owner, there is no interaction between self (vs. familiar and stranger), which still indicates that there is cognitive perceptual constancy in the self-voice perception of individuals in the explicit task. Under different voice owner conditions, there is no interaction between the condition non-self (vs. self), which further proves that individuals have cognitive perceptual constancy to self-voice perception.

VI. CONCLUSIONS

This study systematically discusses the mechanism of individual self-voice perception from implicit and explicit tasks, and finds that there are two different cognitive results of individual self-voice perception under verbal condition. In the non-verbal condition, because of the limited voice information, the influence of bone conduction on the self-voice in the recording will be reduced, so there is no difference between the self-voice and non-self-voice in the individual's implicit and explicit task. In addition, this study also finds itself in the individual the implicit task (vs. non-self) voice identification cognitive perceptual constancy, but this study considers only based on acoustic

generator factors within the individual difference. Future research can be conducted by involving different factors (e.g., environmental factors, social background, emotional) to explore the perceptual constancy of recognizing individual identities through voices. self,

ACKNOWLEDGMENT

This research was completed as part of the academic requirements for the National Science Foundation of China (NSFC) under grant No. 31860285 and No. 31660281. Additionally, part of this work is performed in the Scientific Research Project in Higher Education Institutions of Gansu Province (Grant No. 2017A-165).

REFERENCES

[1] C. Rosa, M. Lassonde, C. Pinard, J. P. Keenan, and P. Belin, "Investigations of hemispheric specialization of self-voice perception," *Brain and cognition,* vol. 68, pp. 204-214, 2008.

[2] P. Belin, S. Fecteau, and C. Bedard, "Thinking the voice: neural correlates of voice perception," *Trends in cognitive sciences,* vol. 8, pp. 129-135, 2004.

[3] T. Asai and Y. Tanno, "Why must we attribute our own action to ourselves? Auditory hallucination like-experiences as the results both from the explicit self-other attribution and implicit regulation in speech," *Psychiatry research,* vol. 207, pp. 179-188, 2013.

[4] A. P. Pinheiro, A. Farinha-Fernandes, M. S. Roberto, and S. A. Kotz, "Self-voice perception and its relationship with hallucination predisposition," *Cognitive neuropsychiatry,* pp. 1-19, 2019.

[5] T. Asai and Y. Tanno, "Distinguishing the voice of self from others: the self-monitoring hypothesis of auditory hallucination," *Shinrigaku kenkyu: The Japanese journal of psychology,* vol. 81, pp. 247-261, 2010.

[6] T. Conde, Ó. F. Gonçalves, and A. P. Pinheiro, "Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing," *International Journal of Psychophysiology,* vol. 133, pp. 66-78, 2018.

[7] P. Belin and R. J. Zatorre, "Adaptation to speaker's voice in right anterior temporal lobe," *Neuroreport,* vol. 14, pp. 2105-2109, 2003.

[8] N. Lavan, L. F. Burston, and L. Garrido, "How many voices did you hear? Natural variability disrupts identity perception from unfamiliar voices," *British Journal of Psychology,* 2018.

[9] M. Candini, E. Zamagni, A. Nuzzo, F. Ruotolo, T. Iachini, and F. Frassinetti, "Who is speaking? Implicit and explicit self and other voice perception," *Brain and cognition,* vol. 92, pp. 112-117, 2014.

[10] M. Candini, S. Avanzi, A. Cantagallo, M. Zangoli, M. Benassi, P. Querzani*, et al.*, "The lost ability to distinguish between self and other voice following a brain lesion," *NeuroImage: Clinical,* vol. 18, pp. 903-911, 2018.

[11] D. Maurer and T. Landis, "Role of bone conduction in the self-perception of speech," *Folia phoniatrica,* vol. 42, pp. 226-229, 1990.

[12] P. Langland-Hassan, "Hearing a voice as one's own: two views of inner speech self-monitoring deficits in schizophrenia," *Review of Philosophy and Psychology,* vol. 7, pp. 675-699, 2016.

[13] M. Xu, F. Homae, R.-i. Hashimoto, and H. Hagiwara, "Acoustic cues for the perception of self-voice and other-voice," *Frontiers in psychology,* vol. 4, p. 735, 2013.

[14] N. Lavan, A. M. Burton, S. K. Scott, and C. McGettigan, "Flexible voices: Identity perception from variable vocal signals," *Psychonomic bulletin & review,* pp. 1-13, 2018.

[15] J. M. Zarate, X. Tian, K. J. Woods, and D. Poeppel, "Multiple levels of linguistic and paralinguistic features contribute to voice perception," *Scientific reports,* vol. 5, p. 11475, 2015.

[16] F. Frassinetti, M. Maini, M. Benassi, S. Avanzi, A. Cantagallo, and A. Farnè, "Selective impairment of self body-parts processing in right brain-damaged patients," *Cortex,* vol. 46, pp. 322-328, 2010.