Simulating the inner voice: A study of sound parameters

Ding DING[a,1], Mark A. NEERINCX[a,b], Willem-Paul BRINKMAN[a]
[a] *Delft University of Technology, Netherlands*
[b] *TNO Perceptual and Cognitive Systems, Netherlands*

Abstract. Introduction: Inner voice is estimated to occur at least a quarter of people's conscious waking life. Much research work asserts that inner voice plays various important roles in cognitive functions, such as self-regulation, self-reflection, and so on. Virtual cognitions are a stream of simulated thoughts people can hear while emerged in a virtual reality environment that intend to mimic the inner voice and thus simulating the effect of an inner voice. Presenting and manipulating virtual cognitions in learning and training may be a useful intervention method to affect people's behavior and beliefs. Exposing people to virtual cognitions, presented as an inner voice, requires the simulation of such voice and therefore understanding of the underlying sound parameters. Many researchers believe that there is a relationship between people's inner and outer voice, even suggesting that people's inner voice resembles the features of their own outer voice. The work presented here, therefore, explored people's perception of their simulated inner voice by considering several core sound parameters of their outer voice. **Methods:** Using a specially developed audio recording and modification software tool, 15 participants (11 males, 4 females) set key sound parameters to match their own voice recording with their perception of either their own inner or their outer voice. After reading aloud nine sentences, they modified seven sound parameters of the recordings: pitch, speed, echo and volume of sound with the frequency band (20-320Hz, 320-1280Hz, 1280-5120Hz, and 5120-20480Hz). **Conclusion:** The result of the study indicates that people's sound perception is different between inner and outer voice. Also, individual variations were found for the perception of inner and outer voice differences. For developers who want to simulate inner voice in a virtual environment, these findings suggest that inner voice has its own distinct characteristics compared to an outer voice. The volume setting for the frequency band of 1280-5120Hz can be based on group perception, whereas for speed and echo settings it might require individualization.

Keywords. Virtual cognitions, Inner voice, Inner speech, Sound parameters

1.Introduction

Possibilities are, when you are reading this first sentence, you are hearing your own voice speaking in your head even if you are not saying anything out loud. This phenomenon is commonly called "inner voice", "inner speech" or referred to as "verbal stream of consciousness". Heavey and Hurlburt [1] found that in their sample, around a quarter of people's conscious waking life contains an inner voice. Much research work asserts that inner voice has a positive effect on many cognitive functions, such as self-regulation [2], self-reflection [3], and so on. Meanwhile, the stream of consciousness, already proposed by psychologist William James [4], refers to a continuous succession

[1] Ding Ding, Interactive Intelligence Group, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands; E-mail: d.ding-1@tudelft.nl.

of thoughts in the conscious mind. It is also a narrative technique, intended to mirror people's internal psychological world and the way internal thoughts form in the mind. James Joyce's Ulysses [5] casts the thoughts and conscious experiences of characters in words in a first-person perspective, just as capturing the inner voice of characters. Based on these considerations, we propose creating virtual cognitions that work as a kind of inner voice or personalized voice-overs when people are in a virtual environment. Like virtual environment aims at replicating an environment by artificially creating sensory experiences, virtual cognitions aim at replicating thoughts by artificially creating cognitive experiences. Some replication successes have already been reported. However, these studies focus on replicating the physical body in virtual reality. For example, the rubber hand illusions or virtual body transfer illusions let people regard parts, or even their entire virtual human body as their own [6, 7]. It is interesting, therefore, to examine possible parallels for virtual cognitions to elicit an internalized mind illusion. Presenting and manipulating virtual cognitions may be a useful way to affect people's behavior and beliefs for training or therapeutic purposes.

Exposing people to virtual cognitions, presented as an inner voice, requires the simulation of such a voice and therefore understanding of the underlying sound parameters. Much research shows that there is a link between people's inner and outer voices. On one hand, taking a developmental perspective, Vygotsky et al. [2] argue that inner voice is the result of a gradual internalization process of outer voice, while Watson [8] claims that inner voice develops with the reduction of self-directed outer voice. On the other hand, taking a functional perspective, Hickok et al. [9] propose that when people speak, an internal copy of the sound of their voice is created simultaneously with the overt sound. Scott [10] goes a step further, putting forward and testing a theory that the internal copy of people's voice can also be generated without overt sound. He also believes that the mechanism the inner voice makes use of is the one mostly applied for processing outer voice. He sees the inner voice as the results of the internal prediction of the sound of one's own voice. Moreover, Filik and Barber's findings [11] suggest that people's inner voice resembles the features of their outer voice, even their regional accent. The work presented here, therefore, explores people's perception of their simulated inner voice by considering several core sound parameters of their outer voice. Although as described above, the inner voice seems to have a close relation, even similarities, with the outer voice, Brocklehurst and Martin [12] also found that stuttering people believed their inner voice was not stuttered, which means people's inner voice might hold different sound characteristics from outer voice. We, therefore, hypothesize that people's sound parameter settings are different depending on the type of voice – inner or outer voice.

2.Methods

To investigate the sound characteristic of the inner voice, an empirical study was conducted. The study was approved by University Human Research Ethics Committee (ID: 20).

a. Participants

15 participants (11 males, 4 females) were recruited throughout the university campus via e-mail or approached personally. Their ages ranged from 23 to 36 (M = 26.1, SD = 3.52).

b. Procedure

By using a specially developed audio recording and modification software tool, the participants first read aloud nine sentences while their voice was recorded. After that, the experimenter explained the concept of an inner voice to the participants and several examples of the inner voice phenomenon were given to help participants to have a clear understanding of this concept. Next, the participants had 2-3 minutes to recall their inner voice experience. After this, participants listened back to their previously recorded sentences and set key sound parameters to match their recording with their perception of either their own inner or outer voice. They modified seven basic audio effects and common digital audio-processing features [13] of the recordings: pitch, speed, echo and volume of sound with the frequency band (20-320Hz, 320-1280Hz, 1280-5120Hz, and 5120-20480Hz). The modification data of the parameter settings was collected as input for the statistical analysis.

c. Data analysis

To analyze the participants' parameter setting data, multi-level models were used. Models were built in R version 3.4.2. All the experiment data, the R scripts, and output files can be found online. [1] Model 1 was the basic model that only included participants as a random intercept. Model 2 was built on Model 1 and added voice type as a fixed effect. Finally, Model 2 was extended by adding voice type as a random effect (Model 3). All models fitted assumed normal distribution, except models fitted on the echo settings. Here a Poisson distribution was assumed. The analysis compared the ability of the models to fit the data.

6. Results

To investigate the consistency of participants' parameter settings across sentences, Cronbach's alpha was calculated for all nine sentences for each parameter, and for both inner voice and outer voice. The results show consistency for the same parameter, and for participants' settings for each sentence both inner voice and outer voice. Coefficient alpha of inner voice ranged from 0.62 to 0.93 ($M = 0.83; SD = 0.10$). Coefficient alpha of outer voice ranged from 0.69 to 0.93 ($M = 0.87; SD = 0.09$).

Table 1 and Table 2 show the results of multi-level analysis. For parameter speed, Model 3 was the most appropriate ($p < 0.01$), while for the sound volume of frequency band 1280-5120Hz, Model 2 was the most appropriate ($p < 0.05$). Except for these two parameters, Model 1 was not outperformed by the extended model for other parameters. As the results show, none of 95% confidence intervals of the standard deviation for random intercept included zero, indicating that a significant variation between participants in setting the parameters in general.

1. **Table 1.** Multilevel analysis results of the parameters settings for Pitch, Speed, Volume of sound with frequency band (20-320Hz, 320-1280Hz, 1280-5120Hz, and 5120-20480Hz)

[1] Files are stored at the 4TU.Research repository with a DOI: 10.4121/uuid:57d78b85-c9ae-4d9e-81f9-23d065913d52

	df	χ^2	p-value	Lower 95%	mean	Upper 95%
Parameter 1: Pitch						
M1: Random Intercept				0.02	0.03	0.05
M2: Fixed VoiceType	1	0.30	0.59			
M3: Random VoiceType	2	4.93	0.08			
Parameter 2: Speed						
M1: Random Intercept				0.02	0.03	0.06
M2: Fixed VoiceType	1	0.97	0.33			
M3: Random VoiceType	2	11.80	0.003	0.03	0.06	0.10
Parameter 4: Volume of sound with frequency band(20-320Hz)						
M1: Random Intercept				0.10	0.14	0.21
M2: Fixed VoiceType	1	2.00	0.16			
M3: Random VoiceType	2	0.24	0.89			
Parameter 5: Volume of sound with frequency band (320-1280Hz)						
M1: Random Intercept				0.05	0.08	0.13
M2: Fixed VoiceType	1	3.01	0.08			
M3: Random VoiceType	2	1.21	0.55			
Parameter 6: Volume of sound with frequency band (1280-5120Hz)						
M1: Random Intercept				0.09	0.14	0.20
M2: Fixed VoiceType	1	7.02	0.01	0.01	0.05	0.09
M3: Random VoiceType	2	3.33	0.19			
Parameter 7: Volume of sound with frequency band (5120-20480Hz)						
M1: Random Intercept				0.11	0.16	0.23
M2: Fixed VoiceType	1	0.52	0.47			
M3: Random VoiceType	2	0.28	0.87			

2.

3. **Table 2.** Multilevel analysis results of the parameter settings for Echo

Parameter 3: Echo			Lower 95%	mean	Upper 95%
M3: Random voice type	Fixed effects	Intercept	-2.99	-1.91	-0.84
		voice type	-2.45	-0.97	0.51
	Random Effects	Intercept	1.23	1.93	3.04
		voice type	1.56	2.48	3.93

The results of multilevel analyses showed that participants set the speed, echo and the volume sound for the frequency band 1280-5120Hz differently when considering inner voice or outer voice. This suggests that people's sound parameters setting is different when it comes to the type of voice. Furthermore, the finding of a significant fixed effect indicates that the difference in volume perception for the frequency band 1280-5120Hz was consistent across participants. Here, participants set the volume higher for outer voice than inner voice. While for speed and echo the finding of a significant random effect suggests deviation across participants for setting inner and outer voices. It also suggests for speed and echo consistency on an individual level, i.e. an individual using the same speed and echo settings across his or her own nine voice recordings. For example, some participants consistently raised the speed for their inner voice and lowered it for their outer voice, while others consistently did this the other way around.

7. Conclusion and discussion

Although the phenomenon of "inner voice" has been studied for decades, controversies concerning the nature and function of inner voice persist. In this study, we employed a parameters modification experiment to gain a better understanding of (1) the relationship between inner and outer voices; (2) the sound characteristic of the inner voice; and (3) simulated internal thoughts in virtual reality to further enable the creation of virtual cognitions. This study has some weaknesses. First, the sample size of our experiment is limited, and the participants are all university students or employees. Second, the study is an indirect perception study, asking participants to replicate the sound of their voice to the best of their abilities. It assumes that people can replicate their voice by modifying these parameters. Third, although the findings give some insight into differences between the inner and outer voices, they do not tell much about the accuracy of the replication. Future work might therefore examine this by asking people for example to rate sound recordings on an analogue scale from not very accurate to very accurate. Still, of course, such examination remains difficult because of the intensely private nature of the inner voice.

Despite these limitations, the study provides some insight into the phenomena of the inner voice needed to create virtual cognitions. Based on the results of this study, some conclusions can be drawn. First, these findings indicate that people perceive their inner voice to sound different from their outer voice. Second, individualization in the perception is observed for the difference between inner and outer voices. For developers who want to simulate inner voice in a virtual environment, these findings suggest that inner voice must be modulated separately from outer voice. The volume setting for the frequency band of 1280-5120Hz can be based on group perception, whereas for speed and echo settings it might require individualization. Interesting is also the absence of systematic differences for various bandwidths, except the 1280-5120Hz band. The 1280-5120Hz band roughly overlaps with the 1000-5000Hz band where humans have been found to be most sensitive [14], and therefore most capable to distinguish between inner and outer voices.

Recently, Craig et al. [15, 16] propose using avatar therapy to let individuals talk with a computerized representation of their inner voice hallucination, aiming at reducing the frequency and severity of auditory hallucinations. It might be interesting to examine whether consistency can be found in the sound parameters settings of these recreated voices, and how they relate to people's own inner and outer voice perception. Moreover, as this study found individual differences how people perceive inner and outer voices, future work might focus on individual factors that could predict these variations as a next step in understanding how inner voice is shaped.

To conclude, this study opens up research into inner and outer voices perception and ways to simulate these voices. It has the potential of exposing people to thoughts and ideas, with applications in entertainment, education and health domains.

Acknowledgements

This research was supported by the China Scholarship Council (CSC), grant number 201506090167.

References

[1] C. L. Heavey and R. T. Hurlburt, The phenomena of inner experience *Consciousness and cognition* **17** (2008) 798-810.

[2] L. S. Vygotskiĭ, E. Hanfmann, and G. Vakar, *Thought and language*. MIT press, 2012.

[3] A. Morin and B. Hamper. Self-reflection and the inner voice: activation of the left inferior frontal gyrus during perceptual and conceptual self-referential thinking, *The open neuroimaging journal*, **6**, 2012.

[4] W. James, *The principles of psychology*. Read Books Ltd, 2013.

[5] J. Joyce, *Ulysses*. Editora Companhia das Letras, 2012.

[6] M. Rohde, M. Di Luca, and M. O. Ernst, The rubber hand illusion: feeling of ownership and proprioceptive drift do not go hand in hand, *PloS one*, **6** (2011), e21659

[7] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, First person experience of body transfer in virtual reality, *PloS one*, **5** (2010) e10564

[8] J. B. Watson, Psychology as the behaviorist views it *Psychological review*, **20** (2013) 158

[9] G. Hickok, J. Houde, and F. Rong, Sensorimotor integration in speech processing: computational basis and neural organization, *Neuron*, **69** (2011) 407-422.

[10] M. Scott, Corollary discharge provides the sensory content of inner speech, *Psychological science* (2013) 0956797613478614

[11] R. Filik and E. Barber, Inner speech during silent reading reflects the reader's regional accent *PloS one* **10** (2011) e25782

[12] P. H. Brocklehurst and M. Corley Investigating the inner speech of people who stutter: Evidence for (and against) the Covert Repair Hypothesis *Journal of Communication Disorders* **44** (2011) 246-260

[13] T. Holmes, *Electronic and experimental music: technology, music, and culture*. Routledge, 2012.

[14] J. J. May Occupational hearing loss *American journal of industrial medicine* **37** (2000) 112-120

[15] J. Leff, G. Williams, M. A. Huckvale, M. Arbuthnot, and A. P. Leff Computer-assisted therapy for medication-resistant auditory hallucinations: proof-of-concept study *The British Journal of Psychiatry*, **202** (2013) 428-433

[16] T. K. Craig *et al.* AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial *The Lancet Psychiatry* **5** (2018) 31-40