EFFECTS OF RANDOM SPLICING ON LISTENERS' PERCEPTIONS

Mihoko Teshigawara[1, a], Noam Amir[2], Ofer Amir[2], Edna Milano Wlosko[2], Meital Avivi[2, b]

[1]The University of Tokushima, Japan; [2]Tel Aviv University, Israel
[a]mteshi@cicee.tokushima-u.ac.jp, [b]me_avv@yahoo.com

ABSTRACT

Twenty-one Hebrew speakers listened to speech excerpts of 27 Japanese cartoon voices in random-spliced and non-manipulated conditions and rated their impressions of physical and personality traits, emotional states, and vocal characteristics on 7-point scales. The correspondence of ratings between the two manipulation conditions was examined by calculating Pearson's correlations for individual participants, and for the mean ratings across participants. Cronbach's alpha was also calculated to assess inter-rater reliability. Possibilities of systematic biases introduced by the random-splicing technique are discussed.

Keywords: cartoon (*anime*) voice, Hebrew, Japanese, personality impression, random splicing

1. INTRODUCTION

People can infer personality traits, physical traits, vocal traits and emotional states of speakers by listening to their voices. Studies on voice and personality can be divided into three paradigms: accuracy studies, externalization studies, and attribution (or inference) studies [1]. The present study's focus is attribution. Unlike accuracy studies, which compare subjective judgments of personality from voice with standardized personality measures, attribution studies involve lay judges' personality attributions from voice without reference to accuracy. This type of research often asks lay judges to listen to voices and rate the vocal and personality traits of the speakers, and shows the statistical correlations between the two, e.g., [3].

Thus obtained listeners' attributions would generally reflect listeners' responses to both the voices and the verbal content if the verbal content were not controlled. Therefore, to elicit listeners' responses to the voices independent of verbal content, it is necessary to control the contents of the speech samples by using standardized speech materials or by masking the contents. The latter would be the only possible solution in cases where researchers are to use pre-recorded speech materials, which is also the case in this study.

Content-masking techniques that have been proposed and studied, along with their effects on expert ratings and/or laypersons' perceptions, include low-pass filtering, random splicing, backward speech, pitch inversion, tone-silence sequences, and reiterant speech [2, 4, 8]. Of these techniques, Teshigawara [5, 6] used random-splicing based on the findings that random splicing is the only one that retains voice quality information, which was the focus of the study [4, 8]. The present study also adopted random splicing to replicate Teshigawara's study with listeners of Hebrew, a different language than that of the spliced and non-manipulated stimuli (Japanese).

In random splicing, speech samples are divided into small segments (250 ms is conventional) and rearranged in an order different from the original. Van Bezooijen and Boves, investigating the effects of random splicing and low-pass filtering on expert ratings of voice quality and prosodic settings, found that random splicing retained information pertaining not only to voice quality features (e.g., harshness, denasality, pharyngeal constriction) but also to prosodic features (e.g., pitch level, loudness) [8].

However, note that it has also been suggested that the random-splicing technique may introduce systematic biases to perception. The vocal and personality traits whose ratings may have been affected by this technique include pitch level and variation [8], *relaxed* [4], and *anger* and *excitement* [2]. In these studies, the ratings obtained in the random-spliced condition were compared with those obtained in the non-manipulated and other content-masking conditions. In two of them, judges rated stimuli in all conditions [2, 8], whereas in Scherer et al.'s study judges were assigned to one of the conditions investigated [4]. Teshigawara [6] also speculates about possibilities of systematic biases introduced by random splicing to the listeners' impressions of *positive emotion* and (vocal) *relaxation*, however, without comparing the results with those from any other condition. Therefore, in order to examine the validity of the use of the random splicing technique in perceptual experiments, as suggested in Teshigawara [6], the experiment should be replicated in at least two ways: by using other content-masking techniques such as those mentioned above [2, 4, 8]; and by using listeners who do not understand the language of non-manipulated stimuli. Although the former has already been done in

previous studies [2, 4, 8], the latter has not been pursued in any of the abovementioned studies that used random splicing (and any other content-masking techniques). In addition, replicating the experiment in both conditions (i.e., random-spliced and non-manipulated conditions) would allow the researcher to compare results from the random-spliced condition in a new language with those from the original study in order to address issues of perceptual universality and cultural specificity. (See [7] for a preliminary report of this type of study.)

This study replicated Teshigawara's Japanese cartoon voice experiment [5, 6] with Hebrew listeners using both random-sliced and non-manipulated stimuli in order to examine the effects of random splicing on the listeners' perceptions of vocal stereotypes. Twenty-seven *anime* characters' voices, the same set of stimuli as in [5, 6] were presented to 21 Hebrew listeners in two conditions, i.e., (i) random spliced, and (ii) non-manipulated. The listeners rated their impressions of the speakers using trait items in the following four categories: physical traits, personality traits, emotional states, and vocal characteristics.

2. METHOD

The original Japanese experiment was conducted to investigate whether the auditorily identified characteristics of voices contribute to people's perceptions of the characters as good or bad [5, 6]. It was revealed that supraglottic states correlated well with participants' ratings of favorable trait items, suggestive of the importance of these articulatory characteristics in listeners' perception of voices. Since the present study aimed to replicate as closely as possible the experiment conducted with Japanese participants, the same methods and stimuli described in these studies were used.

2.1. Stimuli

In Teshigawara's studies, 27 character voices were selected for the perceptual experiment out of the 88 voices auditorily analyzed [5, 6]. The selected characters differed in degrees of laryngeal constriction/larynx lowering (i.e., supraglottic states) determined by the auditory analysis. Noise-free speech samples of the selected speakers were used to create stimuli. In order to create stimuli representative of each speaker, speech portions produced with a voice quality setting deviating from the speaker's normal setting were removed, with the exception of characters who were consistently angry or shouting. To make the stimuli in the two manipulation conditions as close as possible, the lengths of the stimuli were made equal in the two conditions, i.e., 5 s.

2.1.1. Non-manipulated stimuli

Speech portions meeting the criteria above were compiled to make a stimulus of 5 s for each speaker. No pause was inserted when connecting unconnected speech portions so that the portions used for the stimuli became almost identical in the two conditions.

2.1.2. Random-spliced stimuli

After removing pauses from the same candidate speech portions as in Section 2.1.1, speech samples were divided into 250-ms segments. Twenty 250-ms segments with the first and last 3 ms linearly attenuated to zero amplitude were prepared and rearranged so that segments could not occur in the same relative order in the spliced stimulus as in the original. (See [5, 6] for more details.)

2.2. Procedures and participants

The original Japanese questionnaire was translated into Hebrew. Twenty-one trait items were used in the questionnaire for the rating session, of which 19 are discussed here. The listeners were given adjectival labels and were asked to rate the characters on 7-point scales, from 1 (*not at all true*) to 7 (*extremely true*). English translations are given for the items as follows: two physical characteristics (*big*, *good-looking*); one emotional state (*positive emotion*); five vocal characteristics (*high-pitched*, *loud*, *relaxed*, *pleasant*, *attractive*); and 11 personality traits (*selfless*, *loyal*, *devoted*, *brave*, *intelligent*, *strong*, *sociable*, *calm*, *curious*, *conscientious*, *sympathetic*).

Twenty-one female undergraduate students (average age 23.02 years old) were recruited from the Communication Disorders Department at Tel Aviv University as participants. All were native Hebrew speakers with no knowledge of Japanese.

The experiment was run in small groups of four to six people in a sound-attenuated room. The participants had the two manipulation conditions (i.e., random spliced and non-manipulated) at least one month apart. The experiment was designed to counterbalance the effects of stimulus ordering (two orders), condition ordering (two orders), and questionnaire item ordering (two orders), yielding eight conditions in total (i.e., $2 \times 2 \times 2 = 8$). Thus, the participants received either of the two manipulations first according to the experimental design. However, due to space limitations, the remaining discussions will be confined to comparing the two manipulation conditions, disregarding the receiving order of the two manipulations and the other control conditions.

3. RESULTS

3.1. Correlations of individuals' ratings between the two conditions

Pearson's correlation coefficient was calculated between the two conditions for each trait item within participant. Table 1 summarizes the numbers of the items with moderate ($.40 < r \leq .70$) and strong ($r > .70$) correlations between the two conditions, and the mean strength of correlation for each participant.

Table 1: Numbers of items with moderate to strong correlations (out of 19 items) and mean correlation strength for each participant.

participant ID	11	12	13	14	15
$.40 < r \leq .70$	13	8	9	10	3
$r > .70$	0	2	0	3	0
mean corr. strength	.49	.48	.31	.49	.16

16	21	22	23	24	25	26	31	32
13	10	8	12	10	10	9	4	9
2	5	4	1	0	2	2	0	1
.46	.54	.46	.43	.38	.46	.34	.23	.40

33	34	41	42	43	44	45	sum
5	7	7	12	5	9	11	184
0	1	0	1	1	1	2	28
.26	.33	.35	.43	.32	.39	.44	.39

As shown in Table 1, there are 212 moderate to strong correlations of 399 possible combinations (53.1% of possible combinations); that is, if individuals' ratings are averaged out, each participant has approximately 10 moderate to strong correlations, among which there is one strong correlation. Numbers of moderate to strong correlations vary greatly across participants, i.e., 3 to 15 correlations. In sum, individuals' ratings can vary considerably although they are often averaged after Cronbach's alpha is assessed in this type of study [3]. (See the next section for Cronbach's alpha.) When examined individually, correlations between ratings in the two manipulation conditions are only moderate.

A close examination of trait items revealed that the ratings for *high-pitched* were strongly correlated in 13 out of 21 participants, making up almost half of strong correlations (i.e., 13 out of 28 strong correlations). Other items strongly correlated in more than one participant were *big* (3 participants), *attractive* (3), and *pleasant* (2). There were two participants whose ratings for *high-pitched* were not correlated: Participants 24 ($r = .16$) and 31 ($r = -.41$). These participants' ratings in the random-spliced condition were not correlated well with the other participants'. In Participant 31, the rating range was reduced for the spliced stimuli, whereas that for the non-manipulated stimuli was used fully – a possible adverse effect of random splicing. No consistent trend was observed in Participant 24's ratings.

3.2. Inter-rater reliability

Inter-rater reliability was assessed using Cronbach's alpha as in previous studies on voice and personality [3]. Cronbach's alpha was calculated for each item in each condition (middle two columns in Table 2).

Table 2: Cronbach's alpha for each of the two manipulation conditions by trait item and Pearson's correlation between the two conditions averaged over participants by trait item. For Pearson's correlations, $N = 27$; * $p < .05$; ** $p < .01$

Items	Random-spliced	Non-manipulated	Pearson's coefficient
Big	.94	.94	.93**
Good-looking	.94	.91	.88**
Brave	.68	.81	.71**
Selfless	.76	.77	.85**
Loyal	.85	.86	.92**
Devoted	.51	.71	.42*
Intelligent	.78	.85	.89**
Strong	.75	.88	.83**
Sociable	.90	.92	.90**
Calm	.91	.94	.91**
Curious	.83	.85	.79**
Conscientious	.87	.87	.94**
Sympathetic	.91	.91	.90**
Positive Emotion	.92	.94	.87**
High-pitched	.96	.98	.98**
Loud	.83	.85	.83**
Relaxed	.92	.93	.89**
Pleasant	.94	.96	.93**
Attractive	.91	.90	.94**

Contrary to what might be expected based on the moderate correlations within individuals, Cronbach's alphas are generally high in both conditions except the alphas for the item *devoted*. Therefore, we can say that in general participants agreed well on ratings. The items whose alphas are smaller than .80 are five in the random-spliced condition, and two in the non-manipulated condition. There are four items that have a relatively large decrease in alpha for the random-spliced condition (i.e., *brave*, *devoted*, *intelligent*, *strong*), which may be attributable to systematic error introduced by this technique. Aside from these items, the differences between the two conditions are minimal or the trend is even reversed (e.g., *good-looking*). Considering the lowest alpha in both conditions for the item *devoted*, we may speculate that Hebrew listeners had difficulty imagining how devoted a character may be, hearing the voice.

3.3. Correlations of mean ratings between the two conditions

Next, Pearson's correlation coefficients were calculated between ratings in the two manipulation conditions averaged over participants (rightmost column in Table 2). Even more surprising than the

relatively high Cronbach's alphas is that all but one correlation (i.e., *devoted*) is strong, i.e., over .70. Therefore, we can say that the strength of correlation increased when ratings were averaged over participants. It can also be noted that in general, high Cronbach's alphas and strong correlations go together in this set of trait items (cf. [8]).

As a preliminary analysis, correlations among trait items were calculated using the mean ratings for each condition, separately for characters played by male and female voice actors as in [5, 6, 7]. It was revealed that the items which have a lower coefficient (i.e., *devoted*, *brave*; see Table 2) formed different correlation patterns in the two conditions. In the ratings for female voice actors in the random-spliced condition, almost all items appeared to be moderately to strongly correlated with one another including *brave* (less so in *devoted*), whereas *brave* was not correlated very well with the other positive traits in the non-manipulated condition but *devoted* was correlated slightly better. In addition, the mean ratings for *positive emotion* and *relaxed* were examined, and it was confirmed that they were generally lower in both conditions (i.e., mean ratings being below 4 out of 7 points) than the other items (above 4 points); in other words, the lower ratings in these items were probably not systematic biases brought about by the random-splicing technique contrary to Teshigawara's speculation [6].

4. DISCUSSION AND CONCLUSIONS

The correspondence of ratings between the two manipulation conditions was examined individually by calculating Pearson's correlations, which revealed that correlations between ratings in the two conditions are only moderate. However, Cronbach's alphas by item calculated separately for each condition were relatively high, suggesting that participants generally agreed well on ratings. Pearson's correlations calculated between the two conditions using the mean ratings across participants were also high on the whole, indicating that the strength of correlation between the two conditions increased when ratings were averaged across participants. To sum up, although the ratings in the two conditions were only moderately correlated at the individual level, the correlations became stronger at the group level when calculated using the mean ratings across participants. Therefore, we may say that on the whole, participants rated voices comparably in the two manipulation conditions.

Possibilities of systematic biases introduced by the random-splicing technique have also been discussed. First, the range of the rating scale used by a participant appeared to have decreased for the random-spliced condition for one trait item. Second, Cronbach's alphas were lower for a few items in the random-spliced condition. However, Teshigawara's speculation about possible biases of this technique introduced to a couple of trait items [6] was not proved correct.

In future research, it would be necessary to examine how correlations between the two manipulation conditions improve or not as the number of participants increases. It would also be interesting to replicate this experiment in different cultures to see whether results are comparable in the two conditions. A cross-experiment using non-manipulated Hebrew and random-spliced stimuli would also further corroborate the present findings. Lastly, use of other content-masking techniques [2, 4, 8] would also be desired to examine the validity of the use of the random splicing technique in perceptual experiments.

5. ACKNOWLEDGMENT

This work was partly supported by KAKENHI No. 19320060 of the MEXT, Japan. We are grateful for their support.

6. REFERENCES

[1] Brown, B.L., Bradshaw, J.M. 1985. Toward a social psychology of voice variations. In: Giles, H., St. Clair, R.N. (eds.), *Recent Advances in Language Communication and Social Psychology*. London: Lawrence Erlbaum, 144–181.

[2] Friend, M., Farrar, M.J. 1994. A comparison of content-masking procedures for obtaining judgments of discrete affective states. *J. Acoust. Soc. Am.* 96, 1283–1290.

[3] Hecht, M.A., LaFrance, M. 1995. How (fast) can I help you? Tone of voice and telephone operator efficiency in interactions. *J. Applied Social Psychology* 25, 2086–2098.

[4] Scherer, K.R., Feldstein, S., Bond, R.N., Rosenthal, R. 1985. Vocal cues to deception: A comparative channel approach. *J. Psycholinguistic Research* 14, 409–425.

[5] Teshigawara, M. in press. Vocal expressions of emotions and personalities in Japanese *anime*. In: Izdebski, K. (ed.), *Emotions of the human voice, Vol. III Culture and Perception*. Plural Publishing: San Diego.

[6] Teshigawara, M. 2003. *Voices in Japanese Animation: A Phonetic Study of Vocal Stereotypes of Heroes and Villains in Japanese Culture*. PhD dissertation, University of Victoria, Canada.

[7] Teshigawara, M., Amir, N., Amir, O., Milano Wlosko, E., Avivi, M. in press. Perceptions of Japanese *anime* voices by Hebrew speakers. In: Izdebski, K. (ed.), *Emotions of the human voice, Vol. III Culture and Perception*. Plural Publishing: San Diego.

[8] Van Bezooijen, R., Boves, L. 1986. The effects of low-pass filtering and random splicing on the perception of speech. *J. Psycholinguistic Research* 15, 403–417.