

Voices in Japanese Animation: How People Perceive Voices of Good Guys and Bad Guys

Mihoko Teshigawara

Department of Linguistics, University of Victoria, P.O. Box 3045, Victoria, B.C. V8W 3P4 Canada

E-mail: mteshi@myrealbox.com

ABSTRACT

This paper attempts to correlate auditory characteristics and lay people's perceptions of the voices of heroes and villains in Japanese animated cartoons. Based on the auditory characteristics of voices of heroes and villains identified in a separate study, 27 characters were selected as target stimuli from Japanese cartoons. Thirty-two subjects listened to content-masked speech excerpts of the 27 target speakers and rated their impressions of physical and personality traits, emotional states and vocal characteristics on 7-point scales. Voices of heroes were generally rated more positively across all four categories, and voices with opposite auditory characteristics were rated more negatively. Gender also played an important role in the perception of the characters.

1. INTRODUCTION

Japanese *anime*, an animation medium which depicts the world as inhabited by good and bad characters, is wildly popular in Japan and other parts of the world. The few scholarly studies that have considered this medium are still at the development stage [1]. The present study examines the voices of characters in Japanese *anime*, focusing on the articulatory and perceptual characteristics of the voices of female and male heroes and villains.

Vocal stereotyping plays an important role in animation: voices need to reflect the physical attributes and personality traits of characters and the vocal stereotypes that consumers, filmmakers, and voice actors share. Previous studies on vocal stereotypes [2] reveal that people infer similar personality traits from voices. However, few studies have investigated the acoustic correlates of personality in speech [2, 3], and no study has investigated auditory correlates identified by phoneticians or the correspondence between auditory correlates and lay people's perceptions. This study will attempt to fill the gap between the two, using speech samples from Japanese *anime*.

Prior to the present study, the author identified auditory characteristics of the voices of 92 characters (46 heroes and 46 villains) from 20 animated cartoons, using Laver's framework for voice quality description [4]. The following summarizes the auditory characteristics identified across categories (see [5, 6] for more details):

1. Heroes' voices exhibited an absence of pharyngeal constriction and breathy voice.
2. The majority of villains' voices exhibited pharyngeal constriction and harsh voice caused by tense laryngeal tension settings; however, pharyngeal expansion accompanied by lowered larynx was observed in a majority of female and some male villains.

In order to investigate whether the identified auditory characteristics contribute to people's perception of good and bad characters, Japanese lay people's perceptions of selected speech samples were examined in an experimental setting. Speech samples were selected based on the auditory analysis results regarding the perceived pharyngeal states and phonatory settings from a separate study [5,6]. In addition to prototypical heroes and villains, heroes that exhibited pharyngeal constriction and harsh voice and villains without these characteristics were included in order to specifically examine the roles of these auditory characteristics in vocal stereotyping. Subjects were asked to rate their impressions of cartoon characters using trait items in the following four categories: physical traits, personality traits, emotional states, and vocal characteristics. It was hypothesized that subjects would attribute less favorable physical traits, personality traits, emotional states, and vocal characteristics to speakers who exhibited pharyngeal constriction/expansion no matter which roles they played in the original cartoons. Statistical analyses were performed in order to examine the relationship between the auditory characteristics of the voices and subjects' trait ratings of the speakers.

2. METHOD

2.1 Stimuli

In light of the auditory characteristics summarized above, the 92 character voices were divided into two groups, representative and non-representative: representative meaning characters exhibited auditory characteristics appropriate to their role, and non-representative meaning that characters exhibited auditory characteristics opposite or simply atypical of their role. Within these two groups, characters were examined according to role, gender and age (adult versus child). For example, villains showing either pharyngeal constriction or expansion were categorized into the representative villain group, while those showing neither trait fell into the non-representative villain group. There were 16 possible groups: hero or villain (2) □ gender (2) □ age (2) □ representativeness

(2). However, since there was only one child villain (male) in the corpus, this classification system yielded only 13 groups. Two speakers were chosen for each of 12 groups, with the exception of the child male villain group, which had only one speaker. In addition, two sidekicks (one child male and one adult female) exhibiting the auditory characteristics of villains' voices, i.e., pharyngeal constriction and harsh voice, were added in order to see whether they would be rated similarly to heroes or villains. Therefore, the voices of 27 speakers in total were chosen as the basis for experimental stimuli.

Noise-free speech samples of these 27 speakers had been stored on a personal computer for auditory and acoustic analyses for a separate study [5, 6]. They had been recorded from VHS tapes of the cartoons. In order to elicit listeners' responses to the voices independent of verbal content, it was necessary to mask the contents of the speech samples. Of the five content-masking techniques investigated by Scherer et al. [7], random splicing was chosen since it retains voice quality information (the focus of this study), while the other techniques mask it. First, 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. Intensities were standardized across speakers so that the maximum intensity was between 70 and 72 dB. Following previous research using the random-splicing technique [7, 8, 9], after removing pauses, the digitized speech samples were divided into 250-msec segments. The first and last 3 msec of each segment were linearly attenuated to zero amplitude in order to avoid the introduction of transients [9]. In order to create a 5-sec stimulus for each speaker, 20 250-msec segments were prepared and rearranged so that segments could not occur in the same relative order in the spliced stimulus as in the original.

In order to counterbalance the effects of ordering, two stimulus orders (A and B) were used: in A, the 27 speakers were randomly ordered disregarding the speaker groups, while B was the reverse of A. For each speaker, the speaker number was announced followed by the 5-sec stimulus; after one second of silence, the same segment was repeated, followed by 1 minute of silence. This gave subjects a total of 71 seconds to rate each speaker, which according to previous studies [10, 11] is considered sufficient to rate the 21 trait items selected in this experiment. Subjects were given a practice session in which they rated additional three speakers before rating the 27 target speakers.

2.2 Procedure and Subjects

Twenty-one trait items were initially chosen to be used in the questionnaire for the rating session. English translations are given for the items as follows: gender (female or male); age group (-10; 11-18; 19-35; 36-60; 61-); physical characteristics (big, good-looking); personality traits, 11 in total, of which three were chosen for their pertinence to heroes of Japanese *anime* in

particular (selfless, loyal, devoted) [12], three were thought to be universal characteristics of heroes (brave, intelligent, strong), and five represented each of the five factors in the NEO Personality Inventory (sociable, calm, curious, conscientious, sympathetic) [13]; emotional states (positive); and vocal characteristics (high-pitched, loud, relaxed, pleasant, attractive). The 19 adjectives were rated on 7-point scales, from 1 (not at all true) to 7 (extremely true). In order to counterbalance ordering effects, the order of the six trait categories and, where applicable, the items within trait categories were systematically varied, yielding four questionnaire types (I to IV).

Thirty-two subjects (17 females, 15 males; average age 22.8 years old) were recruited from Nagoya University, Japan and the vicinity. In total, eight experimental conditions were yielded, combining the two stimulus orders (A and B) and the four questionnaire types (I to IV). Subjects were randomly assigned to one of the eight conditions. Four subjects were assigned to each experimental condition, with the exception of the two groups that used Questionnaire I, in which five subjects listened to stimulus order A and three listened to B.

Experimental sessions were run in groups of up to five in a soundproof room in the School of Letters building at Nagoya University. Using a CD player, the experimenter played a CD containing instructions recorded by the author, a practice session and the 27 target stimuli. The same instructions were given in the questionnaire booklet as well. Subjects were told that they would hear two 5-sec content-masked cartoon speech excerpts for each speaker, and they were asked to rate impressions of the speakers' traits on 7-point scales and choose appropriate groups for gender and age. After the experiment, subjects completed a questionnaire including demographic information about themselves and their exposure to *anime*. Each session lasted less than one hour.

3. RESULTS AND DISCUSSION

Due to the asymmetry of the stimuli in child villain groups, the following analyses will be confined to the adult hero and villain groups, which have full contrasts in gender and representativeness—16 speakers, two from each of the eight groups, i.e., hero or villain (2) \square gender (2) \square representativeness (2). (Sidekicks will not be discussed in this paper due to space limitations.) In addition, among the 21 trait items, five items (gender, age, "big" from physical traits, and "high-pitched" and "loud" from vocal characteristics) were removed from the present analysis because it appeared that factors other than pharyngeal states played a decisive role in influencing subjects' impressions of these factors. (See a separate study [6] for an analysis of the omitted speakers and items.)

Means and standard deviations were calculated for each of the 16 items for each speaker and then for each of the eight groups across subjects. Table 1 presents mean scores by item for each of the eight groups. Standard deviations ranged from 0.60 to 1.38.

Items	Hero				Villain			
	Male		Female		Male		Female	
	Rep	Non-rep	Rep	Non-rep	Rep	Non-rep	Rep	Non-rep
Physical characteristic								
Good-looking	5.44	3.95	5.19	4.98	3.25	4.27	2.77	4.77
Personality traits								
Brave	5.08	4.72	4.08	4.44	5.19	4.48	3.72	3.66
Selfless	4.56	3.84	4.84	5.07	4.17	3.92	3.34	4.59
Loyal	4.97	4.19	4.92	4.97	4.41	4.33	3.36	4.31
Devoted	4.70	4.59	5.36	5.33	4.84	4.09	3.94	5.13
Intelligent	4.78	3.88	4.67	4.69	4.30	5.02	3.80	3.94
Strong	5.03	4.77	3.72	4.34	5.72	4.44	4.72	3.69
Sociable	4.31	4.22	4.05	4.63	3.52	3.44	3.27	4.67
Calm	4.47	3.47	3.84	3.59	3.28	4.41	2.72	3.81
Curious	4.09	4.44	4.09	3.88	3.19	2.98	3.45	4.25
Conscientious	4.89	3.94	4.98	5.20	4.91	4.97	3.69	4.64
Sympathetic	4.88	4.06	4.84	5.06	3.70	3.66	3.16	4.52
Emotional State								
Positive	4.09	3.83	3.41	3.03	2.94	3.75	2.66	3.44
Vocal characteristics								
Relaxed	4.48	3.81	3.11	2.84	2.91	3.97	2.81	3.16
Pleasant	4.75	3.81	4.03	3.73	3.16	3.72	2.58	3.69
Attractive	4.80	3.97	4.55	4.20	3.55	3.86	2.80	4.05

Table 1: Means of the selected 16 items by speaker group
Rep stands for representative groups and Non-rep for non-representative groups.

In order to examine whether subjects responded to stimuli according to the auditory characteristics of the voices, a series of three-factor repeated measure ANOVAs with three between-subject factors for control purposes was carried out for each item. The three factors were role (hero or villain), gender (male or female), and representativeness (representative or non-representative). The three between-subject factors were the two stimulus orders and the two category and item orders in the questionnaire. For the purpose of Type I error protection, the Bonferroni correction was used; therefore, an alpha level .003 (i.e., .05 divided by 16) was adopted instead of the standard .05 level. Table 2 summarizes the significant main and interaction effects that emerged in the analyses. Partial eta squared is shown as an estimated effect size for each significant effect.

For 11 of the 16 items, the factor role had significant main effects. Therefore, it can be said that the voices of heroes were generally perceived as having more favorable physical and personality traits, emotional states, and vocal characteristics than those of villains. The main effects of the factors of gender and representativeness did not emerge in as many items as for role (six main effects for gender, and three for representativeness). However, there are a number of interactions between any combination of two of the three factors and among the three.

Except for “brave,” “strong,” and “curious,” the same patterns were observed in the directions of interactions across items. In the interaction between role and gender, males’ ratings were generally consistent across roles while

females were rated significantly higher when they were heroes than villains. The items that showed this tendency include “conscientious,” “intelligent,” and “loyal.” This tendency was not predicted; rather it was expected that both male and female villains would receive low scores. In the interaction between role and representativeness, heroes were rated significantly higher when they were representative than non-representative, while villains were rated significantly lower when they were representative than non-representative. This trend was observed in such items as “good-looking,” “calm,” (vocally) “attractive,” and “positive emotion.” This pattern reveals that the classification of auditory characteristics into representative and non-representative based on the auditory characteristics identified in a separate study [5, 6] was valid. In the interaction between gender and representativeness, males were rated significantly higher when they were representative than non-representative, while females were rated significantly lower when they were representative than non-representative. The items where this pattern emerged include “selfless,” “sociable,” and “sympathetic.” In three-way interactions, no consistent pattern was seen among the three items where the interactions were observed. There were also four interactions between one of the three factors and one or more control factors (i.e., stimulus order and orders regarding questionnaire), which were not expected to emerge (see the right-most column of Table 2). More careful controlling may be necessary in future research.

Items	Main Effects				Interaction Effects				Others
	H/V	SG	R	H/V*SG	H/V*R	SG*R	H/V*SG*R		
Physical characteristic									
Good-looking	.76**		.37*		.79**	.69**			
Personality traits									
Brave		.74**				.38*			
Selfless	.58**			.43**	.42**	.54**			
Loyal	.55**			.47**	.48**	.45**			
Devoted	.51**			.38*		.61**	.60**		R*QO2 (.41**)
Intelligent				.58**	.46**		.43**		
Strong		.72**	.51**		.58**				
Sociable	.41**		.52**			.64**			
Calm		.42**			.78**				H/V*SO (.50**)
Curious	.54**			.41**			.43**		
Conscientious				.78**	.42**	.53**			
Sympathetic	.80**				.49**	.64**			
Emotional State									
Positive	.53**	.63**			.54**				
Vocal characteristics									
Relaxed	.36*	.71**			.61**				SG*SO*QO1*QO2 (.33*)
Pleasant	.71**	.52**			.72**				SG*SO*QO1*QO2 (.33*)
Attractive	.66**				.69**	.32*			

Table 2: Estimated effect sizes for the significant main and interaction effects

H/V, SG, and R stand for hero vs. villain, stimulus gender, and representativeness respectively. The interactions with between-subject factors (stimulus, category, and item orders; SO, QO1, QO2, respectively) are listed under “Others”.

* p < .003, ** p < .001

REFERENCES

- [1] J.A. Lent, *Animation in Asia and Pacific*. Bloomington and Indianapolis: Indiana University Press, 2001.
- [2] M. Zuckerman and K. Miyake, “The attractive voice: What makes it so?” *Journal of Nonverbal Behavior*, vol. 17, pp. 119–135, 1993.
- [3] C. D. Aronovitch, “The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker,” *The Journal of Social Psychology*, vol. 99, pp. 207–220, 1976.
- [4] J. Laver, *Principles of Phonetics*. Cambridge, UK: Cambridge University Press, 1994.
- [5] M. Teshigawara, “Voices in Japanese animation,” *Proceedings of the Western Conference on Linguistics*, University of British Columbia, Vancouver, November 1-3, 2002, forthcoming.
- [6] M. Teshigawara, “Voices in Japanese animation,” PhD dissertation, University of Victoria, BC, Canada, forthcoming.
- [7] K. R. Scherer, S. Feldstein, R. N. Bond, and R. Rosenthal, “Vocal cues to deception: A comparative channel approach,” *Journal of Psycholinguistic Research*, vol. 14, pp. 409–425, 1985.
- [8] R. van Bezooijen and L. Boves, “The effects of low-pass filtering and random splicing on the perception of speech,” *Journal of Psycholinguistic Research*, vol. 15, pp. 403–417, 1986.
- [9] M. Friend and M. J. Farrar, “A comparison of content-masking procedures for obtaining judgments of discrete affective states,” *Journal of the Acoustical Society of America*, vol. 96, pp. 1283–1290, 1994.
- [10] R. van Bezooijen, “Sociocultural aspects of pitch differences between Japanese and Dutch women,” *Language and Speech*, vol. 38, pp. 253–265, 1995.
- [11] H. Kido and H. Kasuya, “Everyday expressions associated with voice quality of normal utterance: Extraction by perceptual evaluation,” *Journal of the Acoustical Society of Japan*, vol. 57, pp. 337–344, 2001.
- [12] A. Levi, “The new American hero: Made in Japan,” in *The Soul of Popular Culture: Looking at Contemporary Heroes, Myths, and Monsters*, M. L. Kittelson, Ed., pp. 68–83. Chicago and La Salle IL: Open Court Publishing Company, 1998.
- [13] R. R. McCrae and P. T. Costa, Jr., “Validation of the Five-Factor Model of Personality Across Instruments and Observers,” *Journal of Personality and Social Psychology*, vol. 52, pp. 81–90, 1987.