**MIHOKO TESHIGAWARA**

**Department of Linguistics,**
**University of Victoria, Victoria, BC, Canada**
**E-mail: mteshi@uvic.ca**

# *Voices in Japanese Animation*

## 1. Introduction

Japanese Animation, known as *anime*, which depicts the world as inhabited by good and bad characters, has become a wildly popular medium in Japan and other Asian countries, as well as in North America. Scholarly studies on this medium have only recently started appearing and are still at the development stages (Lent, 2001). The present study sheds light on the voices of characters in this popular medium, focusing on the articulatory and acoustic characteristics of the voices of heroes/heroines and villains.

Vocal stereotyping plays an important role in animation, where 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 (Zuckerman and Miyake, 1993) reveal that people infer similar personality traits upon hearing a given voice. Yarmey (1993) investigated vocal as well as facial cues of good vs. bad characters, but he did not examine their auditory or acoustic properties. In effect, few psychological studies have investigated the acoustic correlates of personality in speech, and no study has investigated auditory correlates. This study will help identify auditory and acoustic correlates of vocal stereotypes of so-called good and evil in Japanese culture.

This paper presents the findings of a preliminary study which will serve as the basis for a larger phonetic investigation of the voices of heroes/heroines and villains in 30 Japanese animated cartoons (Teshigawara, forthcoming). Prior to this preliminary study, psychological studies on vocal cues to personality and emotion and vocal stereotype were reviewed as a means of formulating hypotheses about the auditory and acoustic characteristics of the voices of heroes/heroines. To test these hypotheses phonetically, the voices of heroes/heroines and villains from four Japanese animated cartoons were collected. These samples were first analyzed auditorily using Laver's (1994; 2000) descriptive framework for "voice quality"—quasi-permanent characteristics that are present more or less all the time while a person is talking (Abercrombie, 1967). Following the auditory analysis, wide and narrow band analyses of the voice samples were completed. Based on these analyses, the voices of the majority of villains are shown to be harsh with pharyngeal constriction (a voice quality which may involve activation of the aryepiglottic laryngeal sphincter; see Esling and Edmondson, 2002), while the voices of most heroes/heroines are lax and without pharyngeal constriction.

## 2. Hypothesis

Previous research on vocal attractiveness shows that the relationship between personality traits and vocal characteristics perceived by listeners is mediated by the vocal stereotypes they hold (Berry, 1990, 1992; Miyake and Zuckerman, 1993; Yarmey, 1993; Zuckerman and Miyake, 1993). Therefore, it may be hypothesized that heroes/heroines, who are thought to possess attractive personality traits, have attractive voices. Based on findings from studies in which listeners rated vocal characteristics and personality impressions from voices (Hecht and LaFrance, 1995; Yarmey, 1993) [1], it is hypothesized that the phonetic characteristics of the voices of heroes/heroines will include: high pitch, wide pitch and loudness ranges with temporal fluctuations, and a wide range of articulatory movements; in addition, for heroes only, low pitch, a wide loudness range, and, in Laver's (1980, 1994) terminology, "harsh voice " or "breathy voice"–voice qualities produced with high and low laryngeal tension respectively.

As a basis for hypotheses regarding the voices of villains, the author informally listened to the voices of villains in the materials used in this study. In contrast to the wide variety of positive and negative emotions expressed by heroes/heroines, villains primarily expressed negative emotions such as anger, disgust, frustration, etc. Therefore, it can be expected that vocal cues of the negative emotions mentioned above will be consistently found in villains' voices, in addition to those associated with unattractive personality traits. Based on Scherer's (1986) predictions for the four relevant emotions of displeasure/disgust, contempt/scorn, irritation/cold anger, and rage/hot anger, the articulatory correlates of villains' voices are hypothesized to be: faucal and pharyngeal constriction; tensing of vocal tract walls; vocal tract shortening with the larynx raised and the corners of the mouth retracted downward; overall tensing of the vocal apparatus; and chest register phonation (i.e., low pitch).

Lastly, drawing on Yarmey (1993), who suggests that the schemata for good characters are more typical and likeable while those for bad characters are more unique and less enjoyable, it is hypothesized that the auditory and acoustic characteristics of heroes'/heroines' voices will be more salient and easier to generalize than those of villains, which are assumed to be more unique and to exhibit greater variety.

## 3. Materials

Four TV Japanese cartoon series containing both heroes/heroines and villains were chosen in order to investigate whether the voice quality features of characters in each category are similar, without limiting their personality and physical traits or the types of stories depicted. The official English titles of the four series, in alphabetical order, are:

[1] Hecht and LaFrance (1995) used both male and female speakers, whereas Yarmey (1993) used only male speakers; therefore, only the first half of the hypothesis, which was made based on Hecht and LaFrance (1995), includes heroines.

*Anpanman*; *Doraemon*; *Sailor Moon*; and *3X3 Eyes* (Three Eyes).[2] (See Appendix for reference information on each series.) The lengths of the chosen portions from the four series are: *Anpanman* – 35 min.; *Doraemon* – 40 min.; *Sailor Moon* – 90 min.; *3X3 Eyes* – 30 min. It was noted at which point each hero/heroine or villain appeared in the series and, for the purposes of acoustic analysis, which portions of their speech were free from sound effects or background music. Characters with noise-free speech samples longer than 5 sec were included in this study. The latter speech samples were digitized onto a personal computer at 22,050 samples per second, 16-bit, using Cool Edit Pro LE manufactured by Syntrillium Software Corporation. These digitized segments were stored for acoustic analysis. For characters whose digitized samples were less than 45 seconds, additional speech portions with sound effects and/or background noise were recorded to mini disc to ensure an adequate sample for auditory analysis; according to Laver (2000:43), repeated listening of 45-second speech samples is necessary to conduct auditory analysis using his vocal profile analysis protocol.

In the following analyses, for the sake of convenience, each character was assigned a combination of two letters and a number: the first letter represented the title initial of the series (A for *Anpanman*; D for *Doraemon*; S for *Sailor Moon* [two films together]; and T for *3X3 Eyes*), and the second (H or V) designated either a hero/heroine or a villain; these two letters were followed by a number to complete the character coding system. The age ranges of the characters were estimated by the author; two age ranges, i.e., children and adults, were treated separately in the analyses. In total, the voices of 17 heroes/heroines and villains were analyzed in this study, broken down as follows: four heroes (two children [AH1, DH1]; two adults [SH3, TH1]); five heroines (one child [SH2]; four adults [SV1, SV2, SV3, TV2]); seven male villains (three children [AV1, DV1, DV2]; four adults [SV1, SV2, SV3, TV2]); and one adult female villain (TV1).

## 4. Analysis

*Auditory Analysis*

Laver's vocal profile analysis protocol (see Laver, 1980, 1994, 2000) was used for this analysis. Laver distinguishes two types of factors that contribute to the characteristic sound of a speaker's voice, or "voice quality": "organic" and "phonetic". Of these two, only phonetic factors, which are under the speaker's volitional control, are the subject of description; organic factors, which derive from the speaker's anatomical features and cannot be controlled, are excluded from analysis. The phonetic quality of a voice is created by a combination of "settings". According to Laver (1994: 396), a phonetic setting can be defined as "any co-ordinatory tendency underlying the production of the chain of segments in speech towards maintaining a particular configuration or state of the vocal apparatus." Of the four groups of settings that are distinguished in Laver (1994), three were considered in this analysis: articulatory settings (supralaryngeal settings), phonatory settings (laryngeal settings), and settings of overall muscular tension. These three settings are sub-divided into smaller groups, which also consist of multiple settings, most of which represent the activity of individual articulators, such as the jaw or tongue body. Description of each setting is performed in reference to a neutral setting, from which deviation is measured. The neutral reference setting is the neutral disposition of the vocal tract: for articulatory settings, the neutral reference setting is one by which the central unrounded [ ] would be produced; for phonatory settings, the neutral reference setting is one where voicing shows modal phonation; for settings of overall muscular tension, the neutral requirement is a moderate degree of tension that characterizes the long-term articulatory adjustment of vocal apparatus (see Laver, [1994: 402–404] for more detail). Deviations from the neutral reference setting are accorded a value in terms of three scalar degrees: 1 represents a slight degree of deviation from neutral; 2 a moderate degree; and 3 an extreme degree. In order to identify the settings of a speaker's voice, one needs to listen to a fair amount of speech (45 seconds or longer), given that individual segments differ in their susceptibility to the effect of particular settings.

After listening repeatedly to the speech samples of each character, the author reflected each articulator's movement and deviation from its neutral setting, and developed a vocal profile for each character using Laver's protocol. In Teshigawara (forthcoming), the auditory characteristics of voices of heroes/heroines and villains are discussed separately according to sex and age; however, in the following description, only general tendencies across categories are discussed.

As a general impression, the voices of heroes/heroines sound attractive, while those of villains sound unattractive.[3] This observation is consistent with Yarmey's (1993: 427) claim that "subjects are more likely to process stereotypes of criminals, in contrast to non-criminals, as a function of general images of 'badness'." In other words, voice actors, who share vocal stereotypes of good vs. bad characters with lay Japanese speakers, use those vocal stereotypes to depict characters. It was also observed that the voices of heroes/heroines are higher pitched than those of villains, which coincides with one of the hypotheses proposed in Section 2.

Auditorily, heroes'/heroines' voices are characterized by an absence of pharyngeal constriction (slight pharyngeal expansion in the case of heroes) and lax laryngeal tension settings. The other expected auditory correlate of heroes'/heroines' voices was a wide range of articulatory movements, or, in Laver's framework, a tense supralaryngeal setting. This quality was found in only one speaker (SH3). However, the auditory characteristics of villains' voices are consistent with some of the hypothesized correlates based on Scherer's (1986) negative emotional cues. With the exception of the female and effeminate characters, raised larynx, pharyngeal constriction and harsh voice were prominent characteristics among the villains. Although the female villain (TV1) and the

[2] Of the four series, *Anpanman* and *Doraemon* are for young children and feature child characters as principal roles, while the others are for older children and feature principal characters as old as or older than junior high school ages. Of the four, three, except for *Doraemon*, have obvious heroes and villains, each of whom represents good or evil; in *Doraemon*, two bullies that are elementary school students were regarded as villains, and the principal character (*Doraemon*) who helps the bullied was regarded as a hero.

[3] Exceptions are good-looking villains in *Sailor Moon*.

effeminate villains (SV2 and SV3) show pharyngeal expansion, in the case of TV1 and SV2, the degree of expansion seems to exceed a comfortable level and sounds forced, which may be distinct from the pharyngeal expansion that accompanies positive emotions. As shown in Esling et al. (1994) and Esling (1999), raised larynx and pharyngeal constriction (pharyngealization) go together, as in TV2 and the child villains, while lowered larynx and pharyngeal expansion go together, as in the female and effeminate villains. As illustrated in the foregoing discussion, villains' voices seem to exhibit greater variety than those of heroes/heroines—an observation that is consistent with one of the hypotheses in Section 2. In the next subsection, we will examine whether the acoustic correlates of these auditory characteristics are found in the speech samples.

*Acoustic Analysis*

*Pitch Analysis*

Mean F0 was computed across all speech samples for each of the 17 characters using the Multi-Speech model 3700 analysis package manufactured by Kay Elemetrics. In this analysis, the observation made in the auditory analysis, i.e., that the voices of heroes/heroines are higher pitched than those of villains, was examined. In addition, the relationship between pitch and perceived larynx height, which varies with pharyngeal expansion/constriction (Esling et al., 1994) was investigated, comparing mean F0 and perceived larynx height for characters showing pharyngeal expansion/constriction specifically. For this purpose, the pitch analysis program in Multi-Speech was used, with a frame length of 25 msec. The analysis ranges were determined according to the sex of the voice actors rather than that of the characters in order to match the ranges exploited by voice actors[4]; the ranges were 70 to 450 Hz for male speakers and 100 to 600 Hz for female speakers. Table 1 shows the results of the pitch analysis and auditory impressions of larynx height and pharyngeal constriction/expansion for the 17 characters from 4.1.

Table 1 Pitch Analysis Results, Perceived Larynx Heights and Pharyngeal States. For comparison, characters are grouped by sex. Under "Larynx Height", parentheses indicate slight raising or lowering. Under "Pharynx", parentheses indicate slight or intermittent constriction or expansion.

| | Heroes/Heroines | | | | Villains | | | |
|---|---|---|---|---|---|---|---|---|
| | No. | Mean F0 (Hz) | Larynx Height | Pharynx | No. | Mean F0 (Hz) | Larynx Height | Pharynx |
| Anpanman | AH1 | 255.9 | Neutral | Neutral | AV1 | 148.7 | Raised | Constriction |
| Doraemon | DH1 | 249.6 | Neutral | Neutral | DV1 | 237.3 | Raised | Constriction |
| | | | | | DV2 | 270.6 | (Raised) | Constriction |
| Sailor Moon | SH3 | 150.9 | Neutral | (Expansion) | SV1 | 175.2 | (Raised) | (Constriction) |
| | | | | | SV2 | 182.4 | Neutral | (Expansion) |
| | | | | | SV3 | 196.5 | Neutral | (Expansion) |
| | SH1 | 243.3 | Raised | Neutral | | | | |
| | SH2 | 247.3 | Raised | Neutral | | | | |
| | SH4 | 241.9 | (Raised) | Neutral | | | | |
| | SH5 | 245.5 | Neutral | Neutral | | | | |
| | SH6 | 264.9 | Neutral | Neutral | | | | |
| 3X3 Eyes | TH1 | 198.1 | Neutral | (Expansion) | TV2 | 177.3 | Raised | Constriction |
| | | | | | TV1 | 175.0 | (Lowered) | Expansion |

In *Anpanman* and *3X3 Eyes*, the mean F0s of heroes are higher than those of villains. However, in *Doraemon*, only one villain's (DV1) F0 is lower than the hero's, and in *Sailor Moon*, none of the villains' F0s are lower than the hero's. However, if sex is disregarded in *Sailor Moon*, the villains' F0s are much lower than those of the heroines. Note that the three *Sailor Moon* villains are effeminate, and higher F0s in those characters are thought to contribute to their femininity. (These confounding characters are eliminated from the remaining discussion in this subsection.) For TV1, who initially disguises as an innocent landlady and then later appears as her true villainous character, only the latter value is shown in Table 1. In disguise, TV1's mean F0 is 233.5Hz—considerably higher than the value shown in Table 1.

With respect to the relationship between pharyngeal states and mean F0s, it is not possible to directly compare males with pharyngeal expansion vs. constriction because of the lack of male characters with pharyngeal expansion (TV1 is a female). However, AV1, DV1, DV2, and TV2 have higher F0s than the male speakers analyzed in Oguchi and Kikuchi (1997), who have a mean F0 of 119.850 Hz. Based on this comparison, it may be said that pharyngeal constriction accompanied by raised larynx is associated with a high F0. In contrast, the only character that shows obvious pharyngeal expansion (TV1) has fairly low pitch for a female speaker (175.0Hz) compared to the mean F0 for female speakers in Oguchi and Kikuchi (1997), 228.125Hz. Therefore, it may be said that pharyngeal expansion and low F0 go together.

*Spectrographic Analysis*

In order to illustrate the acoustic correlates of selected phonatory settings common to villains, spectrographic images were obtained for two voices that are harsh with intermittent aryepiglottic fold vibration, using the WaveSurfer program version 1.4.2 (Sjölander and Beskow, 2002). A window length of 172 Hz was used.

Figures 1 and 2 are examples of harsh voices with intermittent aryepiglottic fold vibration (AV1 and TV2 respectively). Both Figures 1 and 2 have relatively high energy in the high frequency range, which is an acoustic characteristic of

[4] The only characters played by the opposite-sex actors were two child heroes, i.e., AH1 and DV1, who were played by adult females using middle to high pitches; child villains were played by high-pitched adult male actors.

tense voice. In Figure 1, the secondary pulse of aryepiglottic fold vibration occurring every other glottal period can be seen most clearly between 4-5 kHz from 0.35-0.45 sec and around 3 kHz from 0.45-0.55 sec; this is similar to what Esling and Edmondson (2002) describe. Although an auditory impression of this voice is higher pitched than some other characters (including TV2, whose spectrogram is shown in Figure 2), the pitch analysis result shows that this voice has the lowest F0 (148.7 Hz). In Figure 1, at the bottom frequency, the same length of presumably aryepiglottic fold vibration can be observed (0.45-0.55 sec); however, these pulses double around 1 kHz, which may have given an impression of higher pitch than would be suggested by the acoustic analysis program. Possibly, the aryepiglottic fold vibration is so strong that it is interpreted as the primary source by the acoustic analysis program.

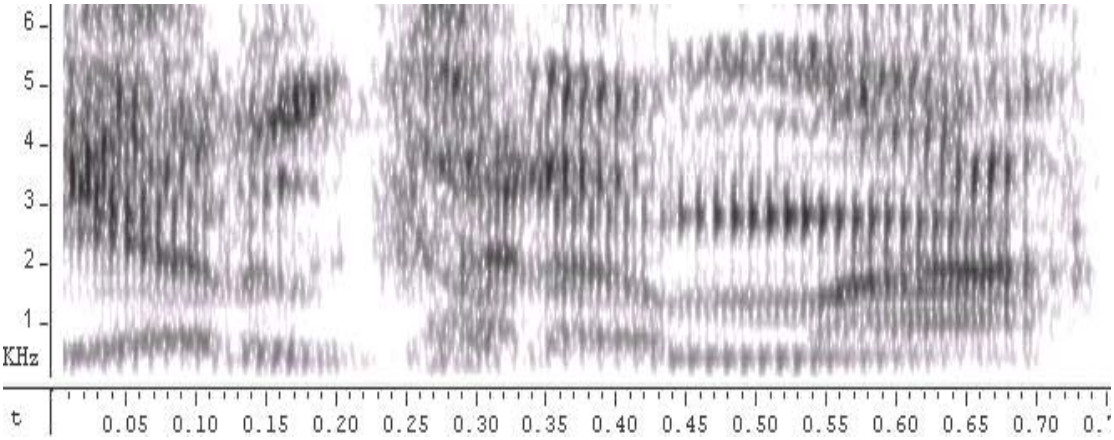


Figure 1 Spectrogram of harsh voice with aryepiglottic fold vibration (AV1: Child male villain) uttering the phrase /yarukarana/ “[I] will make you [feel miserable]”

Figure 2 is also an example of harsh voice with aryepiglottic fold vibration; however, in this example, the aryepiglottic fold vibration seems to be at frequencies lower than half the vocal fold vibration. Between approximately 0.3 and 0.4 sec, seven or so secondary pulses can be observed at around 5 kHz and above, and much finer crepe-like pulses are observed at lower frequencies up to 3 kHz. According to the pitch analysis results, the primary pulses are around 200 Hz, while the secondary pulses seem to be around 50-70 Hz by estimation (one cycle is 14 to 17 msec long).

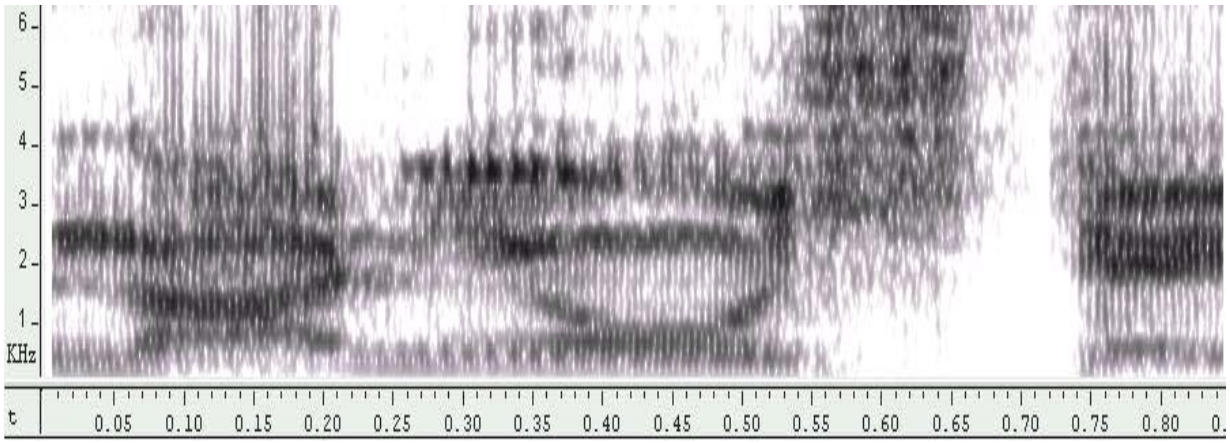


Figure 2 Spectrogram of harsh voice with aryepiglottic fold vibration (TV2: Adult male villain) uttering the phrase /naniosite/ “What [are you] doing?”

## 5. Conclusions

In this paper, hypotheses about the auditory and acoustic characteristics of voices of heroes/heroines and villains in Japanese animation were formulated, based on findings from the research literature on vocal cues of personality and emotion. A preliminary study using four TV animation series was conducted. An auditory analysis using Laver’s (1980, 1994, 2000) framework was performed on the voices of 17 characters (nine heroes/heroines and eight villains). The auditory characteristics of the voices of heroes/heroines voices were an absence of pharyngeal constriction (slight expansion in adult heroes) and lax laryngeal tension settings. In contrast, the main auditory characteristics of villains’ voices were pharyngeal constriction and harsh voice caused by tense laryngeal tension settings; however, in the only female villain and in the three effeminate villains in *Sailor Moon*, pharyngeal expansion was observed. Lax laryngeal tension settings in heroes/heroines and pharyngealization and tense laryngeal tension settings in villains were consistent with some of the hypotheses. The hypothesis that the voices of villains are more unique and exhibit wider variety was also demonstrated. Following the auditory analysis, a series of acoustic analyses, i.e., pitch and spectrographic analyses, were performed, and acoustic correlates of some auditory characteristics mentioned above were identified. In the pitch analysis, it was suggested that, if sex is disregarded, heroes/heroines have higher F0s than villains. It was also suggested that voices with pharyngeal constriction may be higher than those without. In the spectrographic analysis, examples of harsh voice with aryepiglottic fold vibration were shown.

Based on these findings, Teshigawara (forthcoming) is conducting a study on a larger set of samples from 30 Japanese animation series. Aspects not covered in this study, such as supralaryngeal settings, vowel formant analysis, and Japanese lay people's perceptions of the voices, are discussed in the study. Careful examination of both the auditory and acoustic characteristics of these voices will result in a better understanding of the vocal cues to personality and emotion and vocal stereotypes. In addition, auditory and acoustic analyses of the voices of villains may contribute to a fuller understanding of pharyngeal and related articulation, a subject which has been studied extensively by Esling and his colleagues (Esling, 1996, 1999; Esling and Edmondson, 2002; Esling et al, 1994).

**Appendix: Materials analyzed in this study**

*Anpanman*: *Sore Ike! Anpanman* [Go for It! Anpanman], Vol. 2. TV series. TMS, Nippon TV, 1989.

*Doraemon*: *Doraemon*, TV series, Vol. 5. Studio Take, Studio Joke, NTV Animation, Shinei, Nippon TV, 1979.

*Sailor Moon*: *Bishojo Senshi Sailor Moon R* [Pretty Soldier Sailor Moon R], movie. Toei, Aoi, TV Asahi, 1994; *Bishojo Senshi Sailor Moon Super S* [Pretty Soldier Sailor Moon Super S], TV series, Vol. 1, 1995.

*3X3 Eyes*: *Sazan Eyes* [Three Eyes], video, Vol. 1, episode 1. Tabac, Toei, 1991.

**References**


Abercrombie, D. (1967) *Elements of General Phonetic*s, Edinburgh University Press.

Berry, D. S. (1990). Vocal attractiveness and vocal babyishness: Effects on stranger, self, and friend impressions. *Journal of Nonverbal Behavior, 14*, 3, 141–153.

Berry, D. S. (1992). Vocal types and stereotypes: Joint effects of vocal attractiveness and vocal maturity on person perception. *Journal of Nonverbal Behavior, 16*, 1, 41–54.

Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association, 26*, 65-88.

Esling, J. H. (1999). The IPA categories "pharyngeal" and "epiglottal": Laryngoscopic observations of pharyngeal articulations and larynx height. *Language & Speech, 42*, 349-372.

Esling, J. H., & Edmondson, J. A. (2002). The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai. In A. Braun & H. R. Masthoff (Eds.), *Phonetics and Its Applications: Festschrift for Jens-Peter Köster on the Occasion of His 60th Birthday*. Stuttgart: Franz Steiner Verlag.

Esling, J. H., Heap, L. M., Snell, R. C., & Dickson, B. C. (1994). Analysis of pitch dependence of pharyngeal, faucal, and larynx-height voice quality settings. *ICSLP 94* (pp. 1475–1478). Yokohama: Acoustical Society of Japan.

Hecht, M. A., & LaFrance, M. (1995). How (fast) can I help you? Tone of voice and telephone operator efficiency in interactions. *Journal of Applied Social Psychology, 25*, 23, 2086–2098.

Laver, J. (1980). *The Phonetic Description of Voice Quality.* Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of Phonetics.* Cambridge: Cambridge University Press.

Laver, J. (2000). Phonetic evaluation of voice quality. In R. D. Kent & M. J. Ball (Eds.), *Voice Quality Measurements* (pp. 37–48). San Diego, CA: Singular Publishing Group.

Lent, J. A. (2001). *Animation in Asia and Pacific*. Bloomington and Indianapolis: Indiana University Press.

Miyake, K., & Zuckerman, M. (1993). Beyond personality impressions: Effects of physical and vocal attractiveness on false consensus, social comparison, affiliation, and assumed and perceived similarity. *Journal of Personality, 63*, 3, 411–437.

Oguchi, T., & Kikuchi, H. (1997). Voice and interpersonal attraction. *Japanese Psychological Research, 39*, 1, 56–61.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*, 2, 143–165.

Sjölander, K., & Beskow, J. (2002). WaveSurfer, version 1.4.2. Centre for Speech Technology, KTH, Stockholm, Sweden.

Teshigawara, M. (forthcoming). Voices in Japanese Animation. PhD Dissertation, University of Victoria.

Yarmey, A. D. (1993). Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Applied Cognitive Psychology, 7*, 419–431.

Zuckerman, M., & Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior, 17*, 2, 119–135.