Research Forum

A Preliminary Study of Voice Quality Differences between Japanese and American English: Some Pedagogical Suggestions[1]

Yuichi Todaka
Miyazaki Municipal University

Studies of voice quality, while limited, suggest there may be a normal voice quality difference across languages. This paper first reports on a study of measurable voice quality differences in bilingual English (L1)/Japanese (L2) speakers. Results suggest that a focus on voice quality, in addition to conventional phonological features, may aid in producing correct pronunciation. Activities for focusing on voice quality in the classroom are suggested.

声の質の研究は、その数は限られてはいるが、言語によって標準的な声の質が異なっているかもしれないということを示唆している。本稿は、英語を第一言語とし、日本語を第二言語とするバイリンガル話者の、測定しうる声の質の違いについて報告する。その結果は、伝統的な音韻的特徴だけでなく、声の質にも焦点をあてることが、正しい発音を身につける役に立つかもしれないということを示唆している。授業で使える声の質に焦点をあてた学習活動が提案される。

Though studies on voice quality have been conducted in the past (see Laver, 1980 for an extensive literature review), little research has been carried out to see if normal voice quality differences exist across languages. Hanley, Snidecor, & Ringel (1966) compared pitch and loudness among Spanish, American English, and Japanese speakers (eight male subjects each); they found that the Spanish and the Japanese groups had higher pitch and lower sound pressure levels than the American group. It is difficult, however, to assert conclusively that differences observed are due to cross-language factors. Individuals vary greatly in pitch and loudness of speech due to anatomical and speech-style differences. Use of either a much greater number of mono-

lingual subjects or carefully selected bilingual subjects would be necessary to eliminate these and other sources of variation.

The main objectives of this study were to find out if there are systematic voice quality differences between Japanese and English speakers in order to make appropriate pedagogical applications. The study attempted to answer the following questions: (1) Are there any speaking pitch and/or loudness differences between speakers of the two languages? If so, do the differences observed come from cultural or linguistic factors?; (2) Is there any difference in the overall supralaryngeal tension which is attributable to cultural or linguistic factors?; and (3) Is there any difference between the overall laryngeal tension which is related to cultural or linguistic factors?

Since it is not possible to observe directly any language-specific difference in supralaryngeal tension, the study focused mainly on the laryngeal and respiratory (i.e., speaking F0 and intensity) features. Thus, the investigation of language-specific supralaryngeal tension factors is based solely on Laver's (1980) classification of lax and tense settings.

Method

Two male and two female bilingual subjects were used to minimize anatomical effects on voice quality. The L1 of all speakers was English. Since one of the difficulties in conducting a cross-language analysis on voice quality is to find appropriate bilingual subjects, pre-screening and post-assessment procedures were conducted.[2] The selected subjects then participated in various aerodynamic and acoustic experiments. All aerodynamic and acoustic recordings were made in the UCLA Phonetics Lab.

In the aerodynamic experiments, each subject wore a mask containing a piece of gauze which exhibited a known amount of resistance through which the ongoing air pressure had to pass (Rothenberg, 1973). The flow rate was calculated from the pressure difference across the gauze. Oral pressure was recorded using a tube inserted through a hole in the mask designed for that purpose. Before the session the air pressure and flow devices were calibrated; the pressure by use of a manometer and the flow by introducing a known flow. In addition, EGG (Electroglottograph) data were collected through a loose collar with two surface electrodes placed around the neck while the subjects phonated, and the data were entered into a computer for subsequent analyses. A headphone microphone was placed approximately two inches from the mouth and to the side.

Regarding the acoustic experiments, five factors were measured using a computer: (1) formant frequencies; (2) harmonic amplitude differ-

ences; (3) average speaking fundamental frequency; (4) long-term spectral average, and (5) bandwidth differences.

A formant is "a group of overtones corresponding to a resonating frequency of the air in the vocal tract" (Ladefoged, 1993, p. 293). According to Laver's (1980) classification, formant ranges are narrower in lax voice than in tense. Less extensive radial movements of the center of the mass of the tongue away from the neutral configuration were found in lax voice. Therefore, formant frequencies of five vowels (i.e., /a/, /i/, /u/, /e/, /o/) were compared to test for supralaryngeal setting differences. It has been reported that the Japanese vowel space is much smaller than its English counterpart (cf. Keating and Huffman, 1984).

A harmonic is "a whole-number multiple of the fundamental frequency of a wave form" (Ladefoged, 1962, p. 112). Various researchers (cf. Henton and Bladon, 1985) inferred the differences between various phonation types by comparing the amplitude differences between the first harmonic and the second harmonic. Therefore, the harmonic amplitude differences were computed here to infer laryngeal setting differences.

Bandwidths are the range of frequencies to which a resonator responds effectively (Ladefoged, 1962). It has been reported that the bandwidths of the first formant is mainly affected by wall loss, whereas the bandwidths of the higher formants are influenced by radiation loss (Rabiner and Schafer, 1978). In other words, an examination of the bandwidth of the first formant will enable us to determine the overall supralaryngeal tension.

The long-term spectral average analytical method has been used by pathologists to establish criteria to quantify pathological voices (Kitzing, 1986; Hammerberg et al., 1986). A breathy voice is associated with a high noise level (cf. Laver, 1980), and high levels of energy at frequencies between 5 K and 8 KHz are said to be associated with noise component of a breathy-voiced source (cf. Yanagihara, 1967). A recent study (Shoji et al., 1993) reports that breathy voices can be clearly differentiated from normal voices by means of long-term spectral average techniques. Though such a technique is most often used by pathologists to quantify pathological voices, it can be used equally well to infer noise level differences in a high frequency band between normal voices in two languages, as in the present study. Therefore, this technique was used to infer laryngeal setting differences between the two languages.[3]

Results and Discussion

1. Laryngeal setting: The two female subjects in the present study showed a consistent pattern of employing a relatively lax laryngeal setting in Japa-

nese (i.e., their L2) in comparison with their setting in English ($p < 0.05$). This lax setting involves less complete closure of the vibrating vocal folds and also a constant aperture. On the other hand, no consistent differences were observed between the two languages in the two male subjects.

A great deal of research has revealed gender differences within a language (Henton and Bladon, 1985; Klatt and Klatt, 1990). It is said that there is a tendency for English female speakers to employ a breathy setting, though some inter- and intra-speaker variations are found (cf. Holmberg et al, 1988). Therefore, it is possible to interpret the male-female difference found here from the perspective of gender differences. Thus, the observed differences may be due to a sociocultural factor. This interpretation is, however, still speculative and needs to be tested in a subsequent study.

2. Speaking F0 and Intensity: All the subjects (both males and females) used higher speaking F0s in Japanese than in English ($p < 0.05$), but no consistent difference in SPL (Sound Pressure Level) was found. The observed F0 (i.e., acoustic correlate of pitch) difference may be the result of a sociocultural factor since the opposite was expected based upon linguistic factors. In other words, lower F0s in Japanese were expected because of a higher ratio of low to high vowels—three to five times higher in Japanese than in English monologue data. In addition, the SPL results may have been due to inter- and intra-subject variations of speaking style at the time of recording, since no consistent patterns were observed among the participants.

3. Supralaryngeal setting: All of the subjects used much more vowel space in English than in Japanese. In other words, the high tense vowel in English - /i/ - the low back vowel -/a/ - and the high back vowel /u/ describe a greater range of articulatory settings than the common range of Japanese vowels (i.e., a tenser supralaryngeal setting in English according to Laver's 1980 classification). However, no consistent bandwidth differences were observed across languages (i.e., no obvious tension difference according to Laver's 1980 classification).

To summarize these findings:

1. Female speakers employed a breathier laryngeal setting in Japanese than in English.
2. All speakers used a higher pitch in Japanese than in English.
3. All speakers used a wider vowel space in English than in Japanese.

Based upon the findings of the previous and the present studies, several teaching suggestions are given in the next section.

Teaching Suggestions

When Japanese learners of English practice the pronunciation of the target language, they tend to focus upon segmental features by listening to and repeating model pronunciation. However, such practice has failed to produce satisfactory results (Celce-Murcia, 1987). The suggestions in the present study are, therefore, based on the assumption that the general aspects of the voice quality setting of the target language should be taught in addition to conventional lower-level features (i.e., segmentals and suprasegmentals). This assumption is in line with Esling and Wong (1983), who advocate the importance of teaching the higher-level setting features (i.e., voice quality settings) in the target language. We therefore suggest that the higher-level features be assimilated into the lower-level features. Regarding English as the target language, in particular:

1. When producing the sounds, the speaker should apply more subglottal pressure (i.e., speak louder). Holmberg et al. (1988) found that in changing the vocal effort from soft to normal to loud, the intra-speaker variation of voice quality showed a rather consistent result. A soft voice was often breathier and a loud voice creakier than a normal voice.[4] Though only the females employed breathier laryngeal setting in Japanese than in English, this exercise would also help the learner be aware of the aspiration and frication noises of the consonants of English. It is also suggested that the teacher have the learner pay attention to those noises, and have the learner find the proper settings in order to produce the noises effectively. Lack of aspiration and frication noises produced by the Japanese learner of English has been reported elsewhere (cf. Vance, 1987).

2. When producing the vowels, Japanese learners should use as much vowel space as possible. To widen the vowel space, the speaker can expand or constrict the pharynx. However, an easier way is to spread one's lips as far as possible and open one's mouth widely when producing the high front vowel /i/ and the low back vowel /a/, respectively.

3. When speaking, Japanese learners (especially females) should understand that it is not necessary to raise pitch to express politeness, as is common in Japanese. Though this is not a linguistic but a sociocultural factor, it is valuable for the learner to understand the degree to which this cultural difference could cause some misunderstanding, if not confusion of offense, in communication.

4. When speaking, Japanese learners should use relatively wider pitch ranges to learn the various intonational patterns of English. In addition, learners may come to understand the appropriate rhythm of the target language by first imitating how many native English speakers tend to speak Japanese. They often lengthen stressed vowels when compared to non-stressed vowels due to L1 interference (i.e., stress-timed language, Todaka, 1990). Figure 1 illustrates the above suggestions schematically.

Figure 1: A Holistic Approach to Teaching Pronunication[5]

Voice Quality Setting		
Sociolinguistic Properties	*Linguistic Properties*	Higher Level
Segmentals	Suprasegmentals	
Vowels	*Stress/Duration*	Lower-Level
Consonants	*Rhythm/Intonation*	

The training method described above might seem unrealistic; however, once learners understand these important higher-level differences, they may be able to find more natural ways to produce the L2 sounds effectively. I have used these techniques to teach English pronunciation, and the results are encouraging.

Conclusion

The findings in the present study are still preliminary due to the limited number of subjects. Therefore, the suggestions made here may have to be modified in accordance with further research. However, it is clear that many EFL/ESL professionals are now considering the aspect of pronunciation teaching to be an essential component of communicative competence (Morley, 1991), and that a systematic approach to teaching pronunciation should be considered from various aspects. It is hoped that the present study can serve as a guide for future cross-language studies of voice quality. Subsequent studies should reveal language-specific factors which can then be used for language instruction purposes.

Yuichi Todaka has a Ph.D in Applied Linguistics from UCLA. He is currently an assistant professor at Miyazaki Municipal University and is interested in language education and phonetics.

Notes

1. The author presented an earlier version of this paper at the Twenty-Eighth Annual TESOL Convention, Baltimore, March 8-12, 1994.
2. The pre-screening test was conducted by four native speakers of Japanese (all of them are Japanese teaching assistants at UCLA), with 10 possible candidates for subjects selected based on their Japanese proficiency in terms of fluency and pronunciation. All of the raters agreed that the four subjects (two males and two females) selected did not have any English accent in Japanese. However, the two male subjects were rated as having a slight accent in Japanese when they were rated by four monolingual Japanese raters who had never left Japan. The above inconsistency regarding the raters' decisions on the subjects' nativeness in speaking Japanese may be due to differences in tolerance of accent. Therefore, the results for the two male subjects do not necessarily reflect full bilingual competence though they seem to have acquired Japanese effectively.
3. A full discussion of methods, results and analytical procedures is neither appropriate nor desirable in this article. For a detailed discussion of the experimental techniques, see Todaka, 1993.
4. Regarding the effects of vocal effort on laryngeal quality, subglottal pressure ranges were measured since these are said to be the primary factor in raising voice intensity (Fant, 1982). It was found that the difference observed in laryngeal settings between the two languages in the present study was not due to a change in voice intensity, but rather to language-specific factors.
5. Suprasegmental features are sometimes placed at a lower level than segmental features (cf. Gilbert, 1986). However, both features are placed at the same level under the voice quality features here to show that both aspects should receive the same amount of attention.

References

Celce-Murcia, M. (1987). Teaching pronunciation as communication. In J. Morley (Ed.), *Current perspectives on pronunciation* (p. 1-12). Washington DC: TESOL.

Esling, J. H., & Wong, R. F. (1983). Voice quality settings and the teaching of pronunciation. *TESOL Quarterly, 17,* 89-95.

Fant, G. (1982). Preliminaries to analysis of the human voice source. *Speech Transmission Laboratory, Royal Institute of Technology, QPSR/4,* 1-27.

Gilbert, J. (1986). Pronunciation and listening comprehension. In J. Morley (Ed.), *Current perspectives on pronunciation* (p. 29-40). Washington DC: TESOL.

Hammerberg, B., Fritzell, B., Gauffin, J., & Sundberg, J. (1986). Acoustic and perceptual analysis of vocal dysfunction. *Journal of Phonetics, 14,* 533-547.

Hanley, T. D., Snidecor, J. C., & Ringel, R. L. (1966). Some acoustic differences among languages. *Phonetica, 14,* 97-107.

Henton, C. G. & Bladon, R. A. W. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language Communication, 5,* 221-227.

Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988). Glottal air flow and pressure measurements for soft, normal and loud voice by male and female

speakers. *Journal of Acoustical Society of America, 84,* 511-529.
Keating, P. A. & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica, 41,* 191-207.
Kitzing, P. (1986). LTAS criteria pertinent to the measurement of voice quality. *Journal of Phonetics, 14,* 477-482.
Klatt, D. H. & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of Acoustical Society of America, 87,* 820-857.
Ladefoged, P. (1962). *Elements of acoustic phonetics.* Chicago: The University of Chicago Press.
Ladefoged, P. (1993). *A course in phonetics* (3rd ed.). Fort Worth, TX: Harcourt Brace Jovanovich College Publishers
Laver, J. (1980). *The phonetic description of voice quality.* Cambridge: Cambridge University Press.
Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly, 25,* 481-517.
Rabiner, L. R. & Schafer, R. W. (1978). *Digital processing of speech signals.* Englewood Cliffs, NJ: Prentice-Hall.
Rothenberg, M. (1973). An inverse filtering technique for deriving the glottal airflow waveform during voicing. *Journal of Acoustical Society of America, 53,* 1632-1645.
Shojii, K., Rogenbogen, E., Yu, D., & Blaugrund, M. S. (1992). High-frequency power ratio of breathy voice. *Laryngoscope, 102,* 267-271.
Todaka, Y. (1990). *An error analysis of Japanese students' intonation and its pedagogical applications.* Unpublished master's thesis, UCLA, Los Angeles.
Todaka, Y. (1993). *A cross-language study of voice quality: Bilingual Japanese and American English speakers.* Unpublished doctoral dissertation, UCLA, Los Angeles.
Vance, T. (1987). *An Introduction to Japanese phonology.* Albany: State University of New York Press.
Yanagihara, N. (1967). Significance of harmonic changes ad noise components in hoarseness. *Journal of Speech and Hearing Research, 10,* 531-541.

(Received October 10, 1994; revised February 24, 1995)