Voice Quality in Native and Foreign Languages Investigated by Inverse Filtering and Perceptual Analyses

*Kati Järvinen, *Anne-Maria Laukkanen, and †Ahmed Geneid, *Tampere and †Helsinki, Finland

Summary: Objectives. Language shift from native (L1) to foreign language (L2) may affect speaker's voice production and induce vocal fatigue. This study investigates the effects of language shift on voice source and perceptual voice quality.

Study Design. This is a comparative experimental study.

Subjects and Methods. Twenty-four subjects were recorded in L1 and L2. Twelve of the subjects were native Finnish speakers and 12 were native English speakers, and the foreign languages were English and Finnish. Two groups were created based on reports of fatigability. Group 1 had the subjects who did not report more vocal fatigue in L2 than in L1, and in group 2 those who reported more vocal fatigue in L2 than in L1. Acoustic analyses by inverse filtering were conducted in L1 and L2. Also, the subjects' voices were perceptually evaluated in both languages.

Results. Results show that language shift from L1 to L2 increased perceived pressedness of voice. Acoustic analyses correlated with the perceptual evaluations. Also, the subjects who reported more vocal loading had poorer voice quality, more strenuous voice production, more pressed phonation, and a higher pitch.

Conclusions. Voice production was less optimal in L2 than in L1. Speech training given in L2 could be beneficial for people who need to use L2 extensively.

Key Words: L1–L2–voice source–vocal fatigue.

INTRODUCTION

Speaking a foreign language (L2) can impose a number of challenges for the non-experienced speakers. Such challenges are not only limited to the mental stress associated with speaking a foreign language but also include possible effects on voice that are not encountered while speaking the native language (L1).

According to earlier studies[1,2] and clinical observations, people often report experiencing more symptoms of vocal fatigue when they speak a foreign language than when they speak their native language. It has been hypothesized that the experience may be related to the fact that more mental stress is associated with speaking a foreign language.[3] Therefore, either all sensations of fatigue may be intensified, or mental stress itself results in a more taxing speech production (increased phonatory and articulatory effort), which causes overloading of the vocal organ.

Whether or not mental stress is involved, a shift from one language to another is prone to cause changes in voice and speech parameters. In earlier studies, for instance, a change in mean fundamental frequency and pitch range has been reported.[4–8] Changes in voice spectrum have also been observed, as measured by the shape and slope of long-term average spectrum (LTAS), and the level difference between the first harmonics and noise-to-harmonics ratio.[4,9,10]

Mechanical loading posed on the vocal folds increases with fundamental frequency (F0), intensity, and degree of adduction of the vocal folds.[11] The degree of adduction corresponds to phonation type (breathy having a low adduction and pressed showing a high adduction). The amount of vocal loading has been quantified by calculating vocal doses.[12,13] Vocal doses are measures based on fundamental frequency, sound pressure level, and voicing time. A previous study[1] applied vocal doses on the study of vocal loading in L1 and L2. Contrary to the expectations, the vocal doses decreased in L2. This was due to the fact that the total duration of voiced speech decreased significantly when speaking a foreign language. The time of voicing may be related to speech tempo and also to the ratio between voiced and voiceless sounds in a particular language. Therefore, calculation of doses as such does not seem to be sufficient to reflect vocal loading in speaking a foreign language. However, it was found in Järvinen et al[1] that the mean exposure *per second* tended to be somewhat higher in L2 than in L1, which suggests that speaking a foreign language may be more loading than speaking L1. LTAS results also seemed to point toward the same direction. A higher level difference between the peak in the first formant region between 300 and 1200 Hz and the peak in the F0 region between 0 and 300 Hz in L2 was found,[1] possibly indicating a more pressed phonation. That, in turn, is considered to be one factor in vocal loading, in addition to high pitch and intensity, and excessively prolonged voice use.[14,15]

The present study continues the previous one. Here the phonation type is studied by applying inverse filtering, which reveals voice source, that is, the airflow pulses that are generated by the vibration of the vocal folds.[16] Inverse filtering was developed in the late 1950s by Miller.[17] The source-filter approach is based on the ideas that the source and the filter are independent of each other, and speech consists of three separate and independent processes: glottal excitation, vocal tract filter, and lip radiation.[18] Rothenberg[19] introduced an inverse filtering method that estimates the airflow out of the mouth through a mask that avoids

Accepted for publication May 5, 2016.

From the *Speech and Voice Laboratory, School of Education, University of Tampere, Tampere, Finland; and the †Department of Otolaryngology and Phoniatrics, Head and Neck Surgery, Helsinki University Hospital and University of Helsinki, Helsinki, Finland.

Address correspondence and reprint requests to Kati Järvinen, Speech and Voice Laboratory, School of Education, University of Tampere, 33014 Tampere, Finland. E-mail: jarvinen.kati.h@student.uta.fi

Journal of Voice, Vol. 31, No. 2, pp. 261.e25–261.e31
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.05.003

the lip radiation effect. The iterative adaptive inverse filtering (IAIF), developed by Alku,[20] is one of the methods where the glottal flow signal is estimated from a corresponding acoustic speech pressure signal.

The following are the research questions of the present study: (1) Is the perceived phonation quality more pressed when speaking a foreign language (L2) than when speaking the native one (L1). (2) Do the characteristics of the voice source (revealed through inverse filtering) differ between L1 and L2? (3) Do the subjects who report more vocal fatigue in L2 than in L1 have more pressed phonation in L2?

METHODS

Participants and recordings

Originally, 43 subjects in total participated in the study. Twenty of the subjects were native Finnish speakers and 23 were native English speakers, and the foreign languages used were either English or Finnish. A questionnaire about vocal fatigue experienced when speaking L1 and L2 was filled in by all the subjects. Based on the answers to the questionnaire, we picked out 12 subjects (six men and six women) who reported that language shift from L1 to L2 does not increase vocal fatigue (group 1, G1), and another 12 subjects (also 6 men and 6 women) who instead reported more vocal fatigue in L2 than in L1 (group 2, G2). In G1, three of the subjects were native English speakers and nine were native Finnish speakers, and in G2 nine of the subjects were native English speakers and three were native Finnish speakers. The mean age for G1 was 31.6 (standard deviation [SD]: 9.1), and for G2 it was 35.6 years (SD: 15.4). The subjects in G1 considered themselves as more experienced in speaking L2 than the subjects in G2: 50% of the subjects in G1 considered themselves as very experienced or native like, whereas 83% of the subjects in G2 considered themselves to have only some experience.

The recordings took place in a well-damped recording studio using a combined level meter and microphone (Bruel et Kjær Mediator, Type 2206, Copenhagen, Denmark). The microphone was placed 40 cm from the mouth. *Sound Forge* software (Sony Creative Software Inc, Middleton WI, USA) was used. The input frequency was 44,100 Hz, and the amplitude quantization was 16 bits. One-minute text reading and spontaneous speech samples in L1 and L2 were recorded. The texts were the same in content in both languages, and the spontaneous speech samples were recorded from a description of a cartoon using the same picture for both languages.

Perceptual analyses

From the speech samples, one sentence (from 5 to 8 seconds in duration) was extracted for the listening analysis. The text extracts for L1 and L2 were taken from the same place in the text. Three experienced vocologists listened to each subject's samples in L1 and L2. The samples were replayed in random order, but the same speaker's L1 and L2 samples were always presented in pairs. Half of the samples had L1 sample first, and half L2 sample first. Two persons' samples in text reading and in spontaneous speech were repeated in order to study the intra-rater reliability of the perceptual analysis. The listeners first listened to the text reading samples and then to the spontaneous speech samples. After every sample pair, the listeners set a mark on a scale from 0 to 10 reflecting four voice characteristics: general voice quality (0 = very poor – 10 = very good), strenuousity of voice production (0 = very strenuous – 10 = very easy), firmness of phonation (0 = very pressed – 10 = very breathy), and suitability of the pitch for the speaker (0 = too low – 10 = too high).

Acoustic analyses

Short stressed vowels (/a/ in Finnish and /ʌ/ in English) were extracted in three different phonetic contexts both in text reading and in spontaneous speech. The length of the extracted vowels was 0.045 seconds. The extracted vowels were inverse-filtered by *TKK Aparat* software,[21] which has been developed for IAIF inverse filtering of speech pressure signal.[20] The formant and lip radiation effects were manually set for every vowel in the inverse filtering. The low frequency noise cutoff was set at 80 Hz, and changed if necessary. In the acoustic analyses, the mean of each acoustic parameter of the three vowels per text type was calculated for every subject.

The following time and amplitude-based parameters were studied from the inverse-filtered signal (Figure 1): (1) open quotient (OQ), which measures the glottal open phase in comparison to cycle duration; (2) closing quotient (ClQ), which is the ratio between the duration of the glottal closing phase to the period length; (3) speed quotient (SQ), which is the ratio between the duration of the opening phase and the duration of the closing phase, and two amplitude-based parameters; (4) amplitude quotient (AQ), which is the ratio of the peak-to-peak amplitude of the flow (Aac) and the minimum peak of the pulse derivative (Admin); and (5) normalized amplitude quotient (NAQ), which is the AQ divided by period length.[16,20,22,23]

It is common to find that the opening phase has rather two opening instants (t_{o1} and t_{o2} in Figure 1) after the closed phase with a possible (knee) shaped waveform in the beginning of the opening phase. Accordingly, OQ1 is calculated from the primary opening instant, whereas OQ2 is calculated from the secondary opening instant.[24] Similarly, SQ1 is calculated from the primary opening instant, whereas SQ2 is calculated from the secondary opening instant. If a flow pulse did not show two opening instants, then the two opening instants were actually one rather than two separate ones. In such a case, OQ1 = OQ2 and SQ1 = SQ2.

According to earlier studies, phonation quality is signaled by various voice source parameters as follows. Breathy voice implies increased open quotient.[25–27] The properties of the closing phase most directly affect the voice quality as the abrupt closure of the glottis at the end of the closing phase involves the majority of the voicing energy.[17] When intensity and pressedness increase, ClQ decreases.[25] Amplitude-based AQ and NAQ have been found to correlate negatively with pressedness of voice.[20,28]

Statistical analyses

Median and interquartile range were calculated for each parameter because parameters were not normally distributed. Wilcoxon signed-rank test was used for testing the statistical significance of differences between L1 and L2 in perceptual analyses.