Analysis of Polish Vowels of Tracheoesophageal Speakers

Marzena Męsikowska, *Kielce, Poland*

Summary: Objectives/Hypothesis. The aim of this study was to determine the acoustical differences between normal and tracheoesophageal Polish speakers during Polish vowel production.

Methods. Formant frequencies, namely, the first (F1) and second (F2) formant frequencies for 6 Polish vowels produced by 11 normal and 11 tracheoesophageal speakers, were analyzed using statistical analysis of variance and discriminant analysis.

Results. Spectral analysis showed that the F1 and F2 values of Polish vowels produced by tracheoesophageal speakers were significantly higher than those produced by normal speakers, with the exception of the F2 value of /i/ produced by tracheoesophageal speakers. Analysis of variance showed significant differences between speeches based on the F1 and F2 formant frequencies. Discriminant analysis based on the formant frequencies for F1 and F2 exhibited 73.33% of the mean classification score for tracheoesophageal speakers and 96.36% for normal speakers.

Conclusions. Tracheoesophageal speakers exhibit higher F1 and F2 formant frequencies, with the exception of the F2 value for the vowel /i/ than normal speakers. Discriminant analysis showed that the classification process for TE speech exhibits lower accuracy due to the poorer classification of the vowels /i/, /u/, and /y/.

Key Words: Tracheoesophageal speech–Formant frequency–Discriminant analysis–Vowels–Classification.

INTRODUCTION

Analysis of vowels based on formant characteristics has been of interest to many researchers. Various methods have been used to study vowels. Some vowels are better understood than others due to the "limit" positions of articulatory mechanism representation.[1] Authors studying English vowels produced by laryngeal speakers have shown that children exhibit the highest values, women exhibit mid-level values, and men exhibit the lowest values of the first, the second, and the third formant frequencies (F1, F2, and F3).[1] The differences in formant frequencies among these three groups of normal, laryngeal (NL) speakers have been attributed to vocal tract length.[1,2]

The effect of shortening vocal tract has been attributed to laryngectomy.[3,4] F1 and F2 were found to be significantly higher in alaryngeal speech than in laryngeal speech. The explanation for the increased formant frequencies provided in some studies[4–9] is that a reduction in the effective length of the vocal tract may account for these changes in formant frequencies. Finnish vowels were characterized with higher formant frequencies F1 and F2 in all vowels produced by alaryngeal speakers compared with NL speakers, with the exception of F1 for /u, o, e/.[3] Among English esophageal (ES) speakers, formant frequencies F1 and F2 were found to be significantly higher in ES speakers than in NL speakers.[4] For Dutch tracheoesophageal (TE) speakers, higher formant frequencies F1 and F2 in TE speakers than in NL speakers were reported.[5] An additional explanation for changes in formant frequencies was that the back of the tongue might be slightly lowered due to the removal of the larynx.[5] It was also reported that the variation among TE speakers may be larger than among NL speakers because the anatomy of the voice source and the vocal tract both depend on the type and extent of the surgical intervention.[5,8] Regarding Spanish vowels, TE speakers achieved higher F1 and F2 values compared with NL speakers, with the exception of F2 for the vowel /o/.[6] According to other studies, it was suggested that TE and ES speakers articulate vowels with fronted and higher tongue positions relative to the tongue position in NL speakers.[6] Higher F1 and F2 values were also reported in studies of Cantonese vowels.[7] Formant frequency F3 values were also found to be significantly higher in English TE speakers[8] and Mandarin ES speakers[9] than in laryngeal speakers. Although, the changes in formant frequency values in laryngectomy population were attributed to chemoradiotherapy and postoperative complications.[8] Another possible explanation provided in literature when investigating Russian vowels for the changes in formant frequencies was in paralinguistic conditions and the strong psychological preoperative stress that may induce abnormally high formant frequencies.[10] Three laryngectomized patients produced vowels with the formant frequency values move closer to normal values 2 weeks after the operation, and two of them 2 years after the operation.[10] Unfortunately, the findings of the study[10] are supported by only a very few subjects observed.

Discriminant analysis is present in the studies of vowels of laryngeal[11] and alaryngeal[12] speakers. Rosique et al[12] analyzed the energy, bandwidth, and frequency of the four first formant frequencies F1, F2, F3, and F4 of the five Castilian vowels using an established phrase produced by TE, ES, and NL speakers. Discriminatory analysis affirmed that TE vocalization is not as similar to NL vocalization as the ES vocalization.[12]

The aim of the present study was to compare the formant frequencies F1 and F2 during the production of Polish vowels by Polish-speaking TE speakers versus NL speakers using statistical analysis of variance (ANOVA) and discriminant analysis. The analysis provided in this study will allow to compare Polish

Accepted for publication April 8, 2016.

From the Kielce University of Technology, Aleja Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland.

Address correspondence and reprint requests to Marzena Męsikowska, Kielce University of Technology, Aleja Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland. E-mail: marzena@tu.kielce.pl

Journal of Voice, Vol. 31, No. 2, pp. 263.e5–263.e11
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.04.007

language with other languages with respect to formant frequencies and to consider the classification accuracy of TE vowels using discriminant analysis and formant frequencies F1 and F2. Alaryngeal speech is of interest both from the standpoint of the ongoing need to improve speech rehabilitation approaches for laryngectomy patients, and because the investigation modes of alaryngeal speech offers unique opportunities for examining the impact of altered voicing source parameters on speech production/acoustics.

METHODS

Participants

Eleven male TE speakers from Holy Cross Cancer Center, Department of Head and Neck Surgery in Kielce, Poland, participated in the study. TE speakers ranged in age from 50 to 73 years, with a mean age of 63 years. Postoperation time ranged from 6 months to 4 years. All TE speakers were using the Provox2 (Atos Medical AB, Kraftgatan 8, 242 35 Hörby, Box 183, SE-24222, Hörby, Sweden) prosthesis. Eleven male NL speakers participated in the study. NL speakers ranged in age from 47 to 64 years, with a mean age of 58 years. All speakers were native Polish.

Speech materials and recordings

Speech materials consisted of six isolated vowels in Polish, which are presented in Table 1 in International Phonetic Alphabet (IPA) notation. The vowels presented in Table 1 were uttered an average of 10 times by each speaker in the order /a/, /a/, . . ., /a/, /e/, /e/, . . ., /e/, /i/, /i/, . . ., /i/, /o/, /o/, . . ., /o/, /u/, /u/, ..., /u/, and /y/, /y/, . . ., /y/.

In the present study, isolated vowels were investigated to find out if in Polish-isolated vowels produced by TE speakers higher formant frequencies can be observed in comparison to NL speakers and how these changes affect vowel classification in TE and NL speakers.

Speech recordings were made in an audiometric room in regular conditions with a digital recorder. Speakers were in a sitting position; the mouth-to-microphone distance ranged from 0.35 to 0.40 m. The speech sound was transmitted via an electret condenser microphone (Olympus Corporation, Head office: Shinjuku Monolith, 3-1 Nishi-Shinjuku 2-chome, Shinjuku-ku, Tokyo 163-0914, Japan) with a 22-kHz sampling rate and a 16-bit signal resolution.

TABLE 1.
Six Polish Vowels—IPA Notation

Vowel	IPA
/a/	/a/
/e/	/ɛ/
/i/	/i/
/o/	/ɔ/
/u/	/u/
/y/	/ɨ/

Formant frequencies extraction method

First (F1) and second (F2) formant frequencies were extracted automatically at the midpoint of each vowel instantiation using script written in Praat software (Praat Software, the authors: Paul Boersma and David Weenink, Phonetic Sciences, University of Amsterdam, Spuistraat 210, 1012VT Amsterdam, The Netherlands). The Praat formant extraction algorithm works by resampling the speech signal to a frequency of twice the maximum formant (a user-defined parameter in the algorithm). After this, preemphasis is applied, the signal is windowed with a Gaussian-like window, and the linear predictive coding (LPC) coefficients with the algorithm by Burg are computed. In this study, formant frequencies were also visually inspected using Praat software.

Formant frequencies F1 and F2 were analyzed in the present study due to be clearly visible and strong present in spectrograms of vowels of TE speakers.

Procedures

To numerically evaluate the difference between TE and NL speakers in vowel production, the Euclidean distance (ED) between formant frequencies was calculated for each vowel with the following equation:

$$ED(v) = \sqrt{\left(F1_{NL}(v) - F1_{TE}(v)\right)^2 + \left(F2_{NL}(v) - F2_{TE}(v)\right)^2} \quad (1)$$

where v indicates vowels /a/, /e/, . . ., /y/.

The obtained values for formant frequencies were analyzed using STATISTICA software (StatSoft, Inc., 2300 East 14th Street, Tulsa, OK. 74104, USA), including descriptive statistical analysis, one-way ANOVA, and discriminant analysis. ANOVA was performed with F1 and F2 as dependent variables and the NL and TE speaker groups as factors. Tukey *post hoc* analysis was conducted to test differences among group factor levels across the dependent variables.

Discriminant analysis was performed with F1 and F2 as the independent (or entry) variables and the vowels produced by TE and NL speakers as the grouping variables. Discriminant analysis was applied to investigate the classification of vowels, especially to observe misclassifications. Two discriminant functions (namely, Root1 and Root2) were created due to the two entry variables used in the model. For the classification process of discriminant analysis, the classification functions as linear combination of entry variables were used. For every group, a separate linear combination function expressed by Equation (2) was introduced:

$$K_i = c_{io} + c_{i1}F1 + c_{i2}F2 \quad (2)$$

where c_{ij} is the coefficients of entrance variables. The sample was assigned to the group when it obtained the highest K_i value.

RESULTS

Formant frequencies F1 and F2

The mean and standard deviation values of F1 and F2 of Polish vowels produced by TE and NL speakers are presented in Table 2.

To provide a visual comparison between NL and TE vowels for F1 and F2, a plot of the mean values of F1 and F2 for vowels produced by TE and NL speakers is presented in Figure 1. For F1