Marzena MIĘSIKOWSKA, Leszek RADZISZEWSKI
KIELCE UNIVERSITY OF TECHNOLOGY,
Aleja Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Polska

Discriminant analysis of vowels of tracheoesophageal speakers

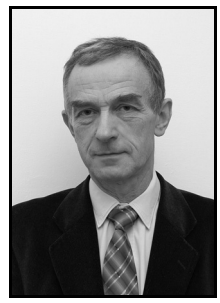
Dr inż. Marzena MIĘSIKOWSKA

Assistant Professor in the Department of Mechanics, Faculty of Mechatronics and Machine Design, Kielce University of Technology, Poland. Research interests: digital signal processing, acoustics.

e-mail: marzena@tu.kielce.pl

Dr hab. inż. Leszek RADZISZEWSKI

Professor in the Department of Mechanics, Faculty of Mechatronics and Machine Design, Kielce University of Technology, Poland. Research interests: acoustics.



e-mail: lradzisz@tu.kielce.pl

Abstract

The aim of this study was to compare normal (NL) and tracheoesophageal (TE) vowel speech signals in order to show differences between them. Cepstral features extracted from vowels of NL and TE speech were analyzed using discriminant analysis. The comparison was made on the basis of the classification function coefficients and the results of the classification for each speech. Vowels recordings were acquired from 10 NL speakers and 12 TE speakers. Discriminant analysis was based on cepstral features extracted from vowel recordings, and was performed separately for NL speech and TE speech. Then a comparison between the coefficients of classification functions of NL and TE vowels using the Euclidean distance was made. Based on the resulting classification matrix of NL and TE speech, the results of classification were compared. The discriminant analysis based on cepstral features showed 79% of the mean classification score for TE speech. The Euclidean distance showed low differences between vowel /a/ of NL speech and vowel /a/ of TE speech and between vowel /o/ of NL speech and vowel /o/ of TE speech.

Keywords: normal speech, tracheoesophageal speech, cepstral features, discriminant analysis, vowels recognition.

Analiza dyskryminacyjna samogłosek mówców mowy przetokowej

Streszczenie

Celem pracy było porównanie sygnału mowy przetokowej (TE) do mowy normalnej (NL) w zakresie samogłosek, aby wykazać różnice pomiędzy sygnałami. Współczynniki cepstralne uzyskane z samogłosek mowy NL i TE poddano analizie dyskryminacyjnej. Na podstawie uzyskanych współczynników funkcji klasyfikacyjnych oraz otrzymanych wyników klasyfikacji dokonano porównania sygnałów mowy NL i TE. Nagrania samogłosek pozyskane zostały od 10 mówców mowy NL i 12 mówców mowy TE. Analizę dyskryminacyjną przeprowadzono w oparciu o współczynniki cepstralne oddzielnie dla mowy NL i mowy TE. Następnie dokonano porównania uzyskanych współczynników funkcji klasyfikacyjnych samogłosek mowy NL i mowy TE, wykorzystując do tego celu odległość Euklidesa. Na podstawie macierzy klasyfikacji otrzymanej dla mowy NL i TE porównano rezultaty klasyfikacji. Analiza dyskryminacyjna w oparciu o współczynniki cepstralne wykazała 79% jako średni wynik klasyfikacji dla mowy TE. Odległość Euklidesa wskazuje na najmniejsze różnice w zakresie samogłoski /a/ i /o/ mowy NL i TE.

Słowa kluczowe: sygnał mowy normalnej i przetokowej, współczynniki cepstralne, analiza dyskryminacyjna, rozpoznawanie samogłosek.

1. Introduction

The focus of rehabilitation following total laryngectomy is voice restoration and speech production to restore communication. There are three major approaches used to restore oral communication: the artificial larynx, esophageal (ES) speech, and tracheoesophageal (TE) speech. TE speech is a surgical-prosthetic voice restoration method.

Analysis of formant frequency values of vowels produced by laryngectomees has been of interest to many researchers. Most of the analysis has been carried out during production of vowels and using first (F1) and second (F2) formant frequency values. Finnish alaryngeal speakers obtain higher formant values compared with normal (NL) speakers, with the exception of F1 of /u, o, e/ [1]. Among English alaryngeal speakers, the systematic changes in formant frequency values, F1 and F2, were found to be consistently higher than those of NL speakers [2]. Dutch alaryngeal speakers produce vowels with significantly higher F1 and F2 values compared to NL speakers [3]. Higher formant frequency values for alaryngeal speakers were also reported among native Spanish laryngectomees [4], and in a recent study of Mandarin ES speech [5] and Cantonese vowels [6]. Polish ES speakers obtain higher values of F1 and F2 compared with NL speakers, with the exception of F2 of /i/ and /y/ produced by ES speakers. Analysis of the third (F3) formant frequency of Polish ES vowels shows higher F3 values for vowels /e/, /u/, and /y/ of ES speakers than NL speakers. The analysis of variance showed significant differences between ES and NL speeches in F1, F2 but no significant main effects between speeches in F3.

Discriminant analysis used as a classifier, based on formant frequency values or cepstral features, has been presented in studies on laryngectomees' speech signal. Such analysis based on formant frequency values of vowels in Castellano dialect applied by the authors of study [7] affirmed that TE voice did not come closer to the NL voice that the ES voice did. Among Polish-speaking laryngectomees, discriminant analysis based on formant frequency values showed 60% of mean classification score for ES speakers and 91% for NL speakers. Discriminant analysis based on cepstral features showed mean classification score equal to 90% for NL vowels, and 76% for ES vowels. Vowel /y/ obtained the lowest classification score for ES speech (57%), as well as for NL speech (80%). The vowel /u/ obtained the highest classification score for ES speech (92%) and for NL speech (98%) [8]. Cepstral features improved classification score for ES vowels. The mean classification score using cepstral features and dynamic time warping (DTW) classifier was 91% for the NL speech, 53% for the TE speech, and 33% for the ES speech. The recognition accuracy based on cepstral features with dynamic programming as a classifier instead of discriminant analysis seemed to be less effective. Analysis of vowels based on Artificial Neural Networks, Support Vector Machines and Naive Bayes, showed the highest recognition rate when using Support Vector Machines. Laryngectomees with different quality of speech achieved 75% acoustic recognition performances [9].

The aim of this study was to compare NL and TE vowel speech signals in order to show differences between them. Cepstral features extracted from vowels of NL and TE speech were analyzed using the discriminant analysis. The comparison was made on the basis of the classification function coefficients and the results of the classification for each speech.

2. Speech recordings, feature extraction and methods

Voice samples were recorded of 12 male TE speakers from Holy Cross Cancer Center, Department of Head and Neck Surgery in Kielce, Poland. Patients ranged in age from 50-73, with an average age of 65 years old. Provox2 prosthesis was used by TE speakers. Speech recordings were made in an audiometric room under regular conditions with an OLYMPUS LS-11 digital recorder. Patients were in a sitting position; the mouth-to-microphone distance ranged from 0.35 to 0.40 m. The voice samples of NL speech were collected under regular conditions in the same settings as laryngectomized patients from 10 speakers (mean age: 56). All speakers were native Polish speakers. The speech sound was recorded with a 22 kHz sampling rate and 16 bit signal resolution.

The recordings consist of six Polish isolated vowels, IPA notation presented in Tab. 1. Every vowel was uttered ten times by each speaker.

Tab. 1. The IPA notation
Tab. 1. Notacja IPA

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

Cepstral features were extracted from speech recordings using Matlab. They were calculated for a speech signal frame in the time domain using the following formula:

$$c(n) = \frac{1}{N}\sum_{k=0}^{N-1} \ln\left|\sum_{m=0}^{N-1} w(m)x(m)e^{-j2\pi km/N}\right| e^{+j2\pi kn/N} \qquad (1)$$

where: $w(m)$ - Hamming Window, $x(m)$ – signal, N-the length of the signal ($n = 1 \ldots N$). Twelve cepstral features (12CC) were calculated for each frame. The cepstral features were analyzed with discriminant analysis using STATISTICA Software [8, 10]. The discriminant analysis consists of a discrimination stage and a classification stage. In this study, the discriminant analysis was based on 12CC features as independent variables and vowels of TE speech as a grouping variable. After determining the variables that discriminate vowels occurring groups, the classification stage was applied to the analysis. Due to six vowel groups, six classification functions were created according to the following formula:

$$K_i(v) = c_{i0} + w_{i1}CC_1 + w_{i2}CC_2 + \ldots + w_{i12}CC_{12} \qquad (2)$$

where: the subscript i denotes the respective group; c_{i0} is a constant for the i'th group, w_{ij} is the weight for the j'th variable in the computation of the classification score for the i'th group; CC_j is the observed cepstral value for the respective case.

The Euclidean distance for each vowel, between the coefficients of NL classification functions and the coefficients of TE classification functions, was calculated with the following formula:

$$K(v) = \sqrt{(c_{i0}(v)_{NL} - c_{i0}(v)_{TE})^2 + (w_{i1}(v)_{NL} - w_{i1}(v)_{TE})^2 + \ldots + (w_{i12}(v)_{NL} - w_{i12}(v)_{TE})^2} \qquad (3)$$

where: v – vowel /a/, /e/, /i/, /o/, /u/, and /y/.

3. Results

Before start of proceeding the discriminant analysis, there was investigated the significance of discriminant analysis. The discriminant stage informs about the significance of discriminant analysis on the basis of Wilks'-Lambda statistic.

The results of discriminant function analysis performed for vowels of NL speech are presented in [8].

Discriminant analysis performed for the TE vowels showed significant main effects for all cepstral variables (12CC) used in the model (Wilks'-Lambda: 0.035, approximation $F(60,3305) = 57.66$, $p<0.0001$). Five discriminant functions (Root1, Root2, Root3, Root4, and Root5), based on 12 CC entry variables, were created. Chi-Square tests with successive roots removed performed in canonical stage are listed in Tab. 2.

Tab. 2. Chi-Square Tests with Successive Roots Removed – TE speech
Tab. 2. Testy Chi-kwadrat dla mowy TE

Roots Removed	Canonical *R*	Wilks' - Lambda	*p*-value
0	0,913	0,035	0,000001
1	0,824	0,209	0,000001
2	0,530	0,652	0,000001
3	0,268	0,907	0,000001
4	0,151	0,977	0,037868

As presented in Tab.2, the chi-square tests of canonical stage showed significance of all created discriminant functions used in the model (R=0.913, Wilks'-Lambda=0.035, p<0. 000001). Removal of the first discriminant function showed high canonical value R between groups and discriminant functions (R=0.824, Wilks'=Lambda=0.209). In general, the more removed functions, the less discrimination between groups. Removal of 4 roots showed low significance of single discriminant function in discrimination (p=0.037868).

Tab. 3. The coefficients of classification functions obtained for TE vowels
Tab. 3. Współczynniki funkcji klasyfikacyjnych samogłosek mowy TE

c_i	$K_1(a)$	$K_2(e)$	$K_3(i)$	$K_4(o)$	$K_5(u)$	$K_6(y)$
Speech TE						
c_{i0}	-61,810	-48,661	-45,845	-57,051	-53,617	-43,555
w_{i1}	75,714	74,853	70,183	84,821	94,008	71,375
w_{i2}	-34,911	-33,884	-35,446	-34,371	-34,664	-33,103
w_{i3}	12,088	7,876	8,512	13,823	16,648	7,442
w_{i4}	5,689	7,568	4,701	8,530	10,163	5,614
w_{i5}	-21,691	-21,854	-19,413	-17,548	-16,704	-19,525
w_{i6}	9,573	11,845	17,089	13,250	17,644	12,661
w_{i7}	-12,784	-11,556	-9,016	-10,149	-6,345	-9,198
w_{i8}	-0,576	0,719	3,158	-0,927	-2,095	1,258
w_{i9}	-1,727	10,547	15,475	4,651	8,251	15,508
w_{i10}	-23,519	-19,417	-21,270	-21,106	-12,651	-14,453
w_{i11}	-21,765	-7,092	4,300	-20,320	-14,384	-1,297
w_{i12}	-125,134	-93,464	-94,545	-125,204	-96,508	-85,289

The issue of classification is another major purpose to which the discriminant analysis is applied. After deriving discriminant functions and determining variables, 12CC features, that

discriminate most between vowel groups, there was proceeded the classification stage. The coefficients of classification functions obtained for NL vowels are presented in [8]. For TE speech, the coefficients of classification functions are given in Tab. 3.

The c_{i0} coefficients obtained for vowels of TE speech were close to the same coefficients of NL speech. Tab. 4 presents the Euclidean distance between the coefficients of appropriate classification functions obtained for the NL and TE vowels.

Tab. 4. Euclidean distance between coefficients of appropriate classification functions of NL/TE vowels
Tab. 4. Odległość Euklidesa pomiędzy współczynnikami odpowiednich funkcji klasyfikacyjnych uzyskanych dla samogłosek mowy NL/TE

K(a)	*K*(e)	*K*(i)	*K*(o)	*K*(u)	*K*(y)
39	64	101	52	76	97

Both speeches differ in the obtained models of classification functions. The smallest distance value was observed for the vowels /a/ (39) and /o/ (52). The largest distance value was noted for the vowel /i/ (101). The Euclidean distance presented in Tab. 4 for NL and TE speeches was smaller than the Euclidean distance obtained for NL and ES speeches [8].

The results of classification using classification functions $K_i(v)$ for NL and TE vowels are listed in Tab. 5.

Tab. 5. The classification scores obtained for NL and TE vowels
Tab. 5. Wyniki klasyfikacji uzyskane dla samogłosek mowy NL i TE

Vowel	Speech	a	e	i	o	u	y
a	NL	**90%**	3%	-	7%	-	-
	TE	**88%**	7%	-	5%	-	-
e	NL	5%	**85%**	-	-	-	10%
	TE	4%	**75%**	-	5%	-	16%
i	NL	-	-	**97%**	-	-	3%
	TE	-	2%	**82%**	1%	-	15%
o	NL	7%	1%	-	**88%**	4%	-
	TE	12%	3%	-	**74%**	11%	-
u	NL	-	-	-	2%	**98%**	-
	TE	-	1%	-	10%	**89%**	-
y	NL	-	10%	10%	-	-	**80%**
	TE	-	19%	16%	-	1%	**64%**
Mean:	NL	**90%**					
	TE	**79%**					

As presented in Tab. 5, a dash "-" in the column means that no case was classified to this vowel group. The mean value obtained for each speech was calculated as a mean value of the best results of classification bolded values in every column that represented the vowel group of the considered speech. For the TE vowels, the lowest result of classification was obtained for the vowel /y/ (64%). The highest score was obtained for the vowel /u/ (89%). The mean classification score was equal to 79% and was 11% lower than the classification score obtained for the NL speech.

4. Conclusions

The aim of this study was to compare NL and TE vowel speech signals in order to show differences between them. The comparison of signals can be useful to find quantitative parameters indicating significant differences between the signals.

The discriminant analysis based on cepstral features extracted from vowels of NL and TE speech was performed. The comparison was made on the basis of the classification function coefficients and the results of the classification for each speech.

The discriminant analysis showed significance of all discriminant functions used in the model for both speeches due to the canonical value R obtained in chi-square tests with successive roots removed. Cepstral features much better discriminate vowels of NL speech than of TE speech.

The discriminant analysis and cepstral features gave the mean classification score equal to 90% for NL vowels, and 79% for TE vowels. NL speakers obtained better classification scores than TE speakers. The discriminant analysis applied as a classifier improved classification scores of TE vowels more than the DTW classifier. The vowel /y/ obtained the lowest classification score for TE speech (64%), as well as for NL speech (80%). The vowel /u/ obtained the highest classification score for TE speech (89%) and for NL speech (98%).

The Euclidean distance calculated for appropriate classification functions of NL and TE vowels differs in values. The smallest differences were observed for the vowels /a/ and /o/. The highest value of the Euclidean distance was obtained for the vowels /i/, /y/, /u/, and /e/. It may suggest that TE speech differs from NL speech in the vowels /i/, /y/, /u/, and /e/, but it requires further study.

5. References

[1] Kytta J.: Finnish oesophageal speech after laryngectomy: sound spectrographic and cineradiographic studies. Acta Otolaryngol, 195: 1–94, 1964.

[2] Sisty N.L., Weinberg B.: Formant frequency characteristics of esophageal speech. Journal of Speech and Hearing Research, 15: 439–448, 1972.

[3] van As C.J., van Ravesteijn A.M.A., Koopmans-van Beinum F.J., Hilgers F.J.M., Pols L.C.W.: Formant frequencies of Dutch vowels in tracheoesophageal speech. Institute of Phonetic Sciences, University of Amsterdam, Proceedings,21: 143–153, 1997.

[4] Cervera T., Miralles J.L., Gonzáles-Àlvarez J.: Acoustical analysis of Spanish vowels produced by laryngectomized subjects. Journal of Speech, Language, and Hearing Research, 44: 988–996, 2001.

[5] Liu H., Manwa L. Ng: Formant characteristics of vowels produced by Mandarin esophageal speakers. Journal of Voice, 23(2):255-260, 2009.

[6] Manwa L. Ng, Rhoda Chu: An Acoustical and Perceptual Study of Vowels Produced by Alaryngeal Speakers of Cantonese. Folia Phoniatr Logop, 61:97–104, 2009.

[7] Rosique M., Ramón J. L., Canteras M., Rosique L.: Discriminant analysis applied to formants of vowels in Castellano dialect during the phonation with prosthesis and esophageal voice after total laryngectomy. Acta Otorrinolaringol Esp. 54(5) (2003) 361-366.

[8] Mięsikowska M., Radziszewski L.: Cepstral Analysis of Vowels of Esophageal Speakers, Measurements Automation and Monitoring (PAK), Vol.58, 11 (2012).

[9] Pietruch R., Grzanka A.D.: Vowel Recognition of Patients after Total Laryngectomy using Mel Frequency Cepstral Coefficients and Mouth Contour. Journal of Automatic Control, Vol.20(1): 33-38, 2010.

[10] Discriminant analysis – STATSoft Electronic Documentation: http://www.statsoft.com/textbook/discriminant-function-analysis/

otrzymano / received: 08.03.2013
przyjęto do druku / accepted: 01.05.2013 *artykuł recenzowany / revised paper*