

A study of sounds produced by Algerian esophageal speakers

*Ferrat K^{1,2}, Guerti M¹

1. National Polytechnic School (ENP), El-Harrach, Algiers, Algeria
2. Centre of Scientific and Technical Research for Development of Arabic Language (CRSTDLA), University of Bouzareah, Algeria

Abstract

Background: Total Laryngectomy is a mode of treatment of patients with advanced laryngeal cancer. It affects the voice and the speech communication.

Objective: To present an acoustic analysis of the new voice after total laryngectomy in Algerian hospital environment.

Methods: A corpus of sounds was collected from October 2008 to September 2009 and pronounced by eight male speakers who have undergone total laryngectomy. Minimum age of patients was 47 years and maximum age was 59 years with mean age 54.87 years. Recordings were made before the beginning of reeducation and after three, six, and eleven months using esophageal voice. The acoustic analysis includes the Pitch F_0 (Hz), Formants, intensity, Jitter (%), Shimmer (dB), harmonic to noise ratio HNR (dB), and degree of unvoiced frames DUF (%).

Results: We note a restriction in F_0 , increasing of Jitter and Shimmer, decreasing of HNR values, and reduced intensity compared to the voice of normal laryngeal speakers. In addition, we note a higher percentage of DUF during the pronunciation of sustained vowels.

Conclusion: Some deficiencies were reported in the taking care of patients. Therefore, the acoustic analysis may be used in evaluating the reliability of the technique of reeducation.

Key words: Total laryngectomy, pathological voice, acoustic analysis, Jitter, Shimmer.

African Health Sciences 2012; (4): 452 - 458 <http://dx.doi.org/10.4314/ahs.v12i4.9>

Introduction

Total Laryngectomy (TL) is a surgical procedure involving the complete removal of larynx after advanced laryngeal cancer¹. This surgical procedure includes vocal cords that normally produce speech sounds. As a result, the communication is seriously impaired and restoring this communication by speech has been a great challenge for speech pathologists^{2,3,4}. In order to communicate, an entire new type of speech will need to be learned. That is way; there are three primary approaches to speech rehabilitation after TL: Esophageal Speech (ES), TracheoEsophageal Speech (TES) and Artificial Larynx (AL)^{5,6,7,8}. In Algeria, the ES is the most privileged technique for voice rehabilitation after TL. Indeed, it does not require expensive devices such as prostheses. In addition, this method is generally convenient and allows a significant degree of intelligibility after a sufficient period of reeducation⁹.¹⁰. In this method, laryngectomized patients learn how

to swallow down air into the esophagus and create sounds by releasing the air which passes through a vibrating PharyngoEsophageal PE segment^{11,12}. The air is then shaped into intelligible speech by the organs of vocal tract cavities (tongue, lips and teeth).

Acoustic analysis of laryngectomee voice has received a great attention from researchers in other languages^{13,14,15,16}. Our study presents some observations on quality of voice rehabilitation with ES performed in Algerian hospitals.

Methods

Colloquial Algerian Arabic

Algerian colloquial Arabic is spoken daily by the vast majority of Algerians¹⁷. It's spoken as a native or as a second language by more than 83% of the population. However, in the media and on official occasions, the spoken language is Modern Standard Arabic. In Algeria, as elsewhere, spoken Arabic differs from written Arabic. The Algerian Arabic has a vocabulary mostly original Arabic, with significant Berber substrates, and new words borrowed from French language.

*Corresponding author:

Kamel Ferrat
National Polytechnic School (ENP)
El-Harrach, Algiers
Algeria
Email: kamelferrat@yahoo.fr

Participants

The sounds were pronounced by eight patients who have undergone TL, with age varying from 47 to 59 years, the mean age is 54.87 years. All the subjects have done speech reeducation with speech-language pathologist (sessions during 11 months). They have learned the method of esophageal speech with inhalation technique. All the subjects are male and they communicate in Algerian colloquial Arabic. Institutional review board approval was obtained for this study.

Speech database

We have used a corpus of sounds collected at the Speech Processing Laboratory of the CRSTDLA, Algiers (Algeria) between October 2008 and September 2009. The recordings were made before the beginning of reeducation, and after three, six, and eleven months of reeducation. The interest of recording speech after surgery and before speech therapy is to confirm the absence of voiced components, by measuring specifically the Degree of Unvoiced Frames DUF (%) during the pronunciation of sustained vowels. The same corpus was pronounced by five normal laryngeal speakers, researchers at the CRSTDLA, and without any pathology of voice.

The corpus consists of sustained vowels, CV utterances, isolated words, and read texts. For the extraction of acoustic parameters, we have exploited the sustained vowels [a, i, u]. The CV utterances and the isolated words were exploited to detect eventual confusions between consonants having same place of articulation but differ by the index voiced/unvoiced. For that reason, we have chosen the consonants corresponding to voiced/unvoiced pairs: pharyngeals [E]/[H]; uvulars [G]/[X], velars [g]/[k] and dentals [d]/[t]. The read texts allowed us to see the fluency and intelligibility of speech.

Acoustic analysis

The acoustic analysis was carried out using the Matlab® and Praat® software, including the Pitch F_0 (Hz) with the standard deviation and range (max and min F_0), formants (F_1, F_2, F_3), Jitter (%), Shimmer (dB), Harmonic to Noise Ratio HNR (dB)¹⁸, Degree of Unvoiced Frames (DUF) and intensity of voice. The Praat software was used in order to extract following parameters: F_0, F_1, F_2, F_3 , DUF, Intensity and HNR.

This paper deals with analyzing and assessing of pathological voices. Hence, we give particular

attention to Jitter and Shimmer parameters that we have developed using Matlab® 7.5. Jitter and shimmer are defined as period-to-period fluctuations in vocal fold frequency and amplitude and they are largely used for the description of the quality of pathological voice^{19, 20, 21}. Also, Jitter and Shimmer are commonly measured for sustained vowels, and their values above a certain threshold are considered being related to pathological voices, usually perceived by humans as breathy, rough or hoarse voices²².

The algorithm designed for extracting Jitter and Shimmer is based on cepstral analysis technique which permits an easy and distinct separation of glottal pitch and filter function of the vocal tract²³. The cepstral analysis is a robust technique of representing short-time variation of speech spectral parameters and provides fine detail pitch information²⁴. That is why, in general, cepstral analysis applied to extract the glottal peaks is more accurate than other techniques such as autocorrelation. The prominence of the high-level cepstral peaks in relation to extraneous vocal frequencies provides a more efficient method of quantification for the voice disorders^{25, 26}. The cepstral peaks differ significantly with the different degrees of voice disorders. It's known that any change in the anatomical structure of the vocal tract, because of pathology, affects the cepstral coefficients and its derivatives. This cepstral technique allows us to calculate Jitter and Shimmer as well as to draw the melodic contour of the speech signal. The global melodic contour corresponds to all the obtained local F_0 in each frame of the signal. The frame and step durations are fixed respectively to 30 and 10 ms.

We note that the Cepstrum of a signal is defined as the Inverse Fast Fourier Transform (IFFT) from the logarithm of the magnitude of the data sequence, Fast Fourier Transform (FFT).

$$C_n = IFT (\log |FT(x([n]))|)$$

Where "x (n)" is data and "n" is a number of samples. The real part of the cepstrum is sufficient to detect the pitch which is located in the part of cepstrum corresponding to the excitation signal (higher "quefrequency"). The real Cepstrum is given by:

$$Cx(n) = \frac{1}{2\pi} \int \log |x(e^{j\omega})| e^{j\omega} d\omega$$

To obtain an estimation of F_0 from the cepstrum, we look for a peak in the "quefrequency" region corresponding to typical speech fundamental frequencies.

After pitch detection and peak-to-peak value extraction over time, we exploit the local periods and amplitudes of glottal peaks to extract Jitter and Shimmer values, as shown in figure 1.

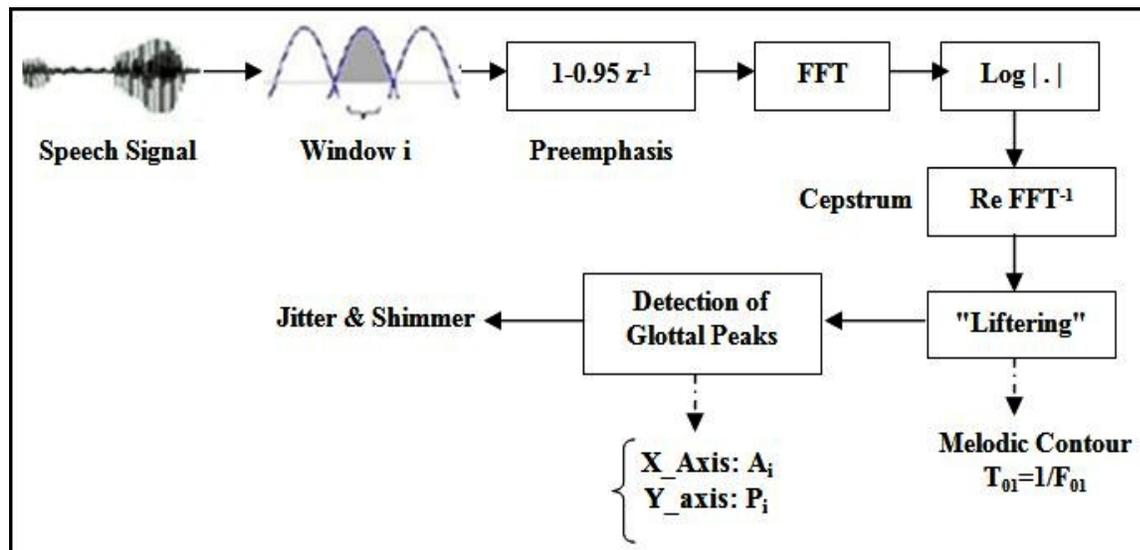


Figure 1: Block diagram of the calculation of Jitter and Shimmer

We have supplemented the acoustic analysis by a global perceptual judgment of speech in order to measure the voice acceptability after eleven months of reeducation, based on the subjective perception of the listener in regard to fluency and intelligibility of speech. For this, we have used an expert jury consisting of three Algerian native speakers, without any previous knowledge of the words and sentences of the recorded corpus. Finally, a look at the spectrographic representations of sounds pronounced by the patients helps to confirm some

speech disorders, such as confusions in the pronunciation of some consonants.

Results

An example of extracted parameters using the sustained vowel [a] is given in tables 1 and 2. All values of acoustic parameters are extracted from sounds pronounced by the patients before reeducation and after 3, 6 and 11 months of reeducation.

Table 1: Jitter, Shimmer and HNR values of the vowel [a]

[a]	Normal	Before reed.	3 months	6 months	11 months
Jitter (%)	0.247	-	12.271	6.013	1.745
Shimmer (%)	3.410	-	7.891	7.910	10.588
(dB)	0.297	-	1.119	1.114	0.681
HNR (dB)	20.748	-	1.710	2.154	3.414
DUF (%)	0	88.57	46.93	41.37	13.115

After reeducation, a voicing begins to appear ($F_0 > 60$ Hz). We note nevertheless a general restriction of pitch (unvoiced frame $>40\%$ after six months of reeducation). Thus, esophageal speech is characterized by a very low average Pitch.

The Jitter range tends to decrease during the reeducation and we note a certainly slow but

perceptible evolution of the Shimmer after six months of reeducation.

The HNR remains relatively low, compared to the HNR of normal voice. These values are still far from the normal threshold. Also, we note an increase of the values of the formants F_1 , F_2 and F_3 after ablation of larynx (figure 2), and less important values of

intensity. The timbre of voice which presents a significant reduction in harmonic components begins to become rich after a period of reeducation.

Nonetheless, the richness of harmonics is always far from the threshold of normality.

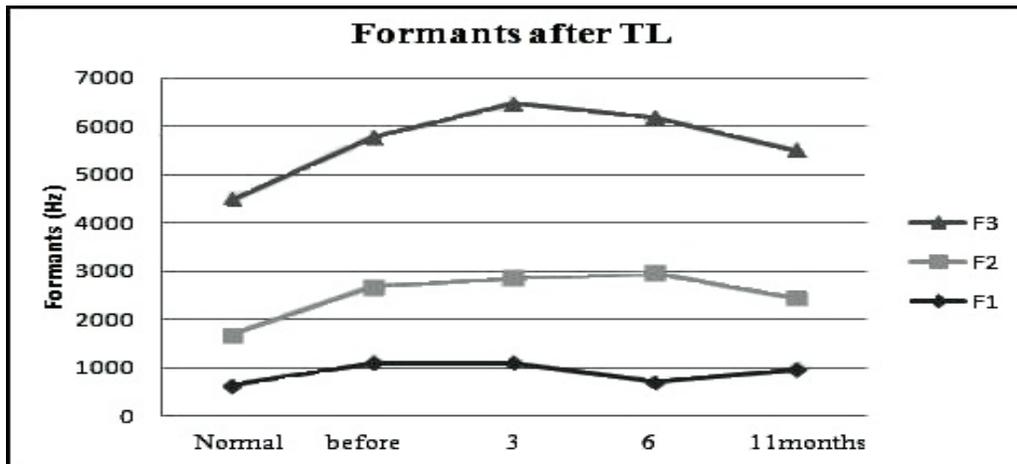


Figure 2: Increase of formants after total Laryngectomy (vowel [a])

The spectrographic analysis of the CV utterances and isolated words pronounced by the patients shows a real confusion between some consonants having the same point of articulation but differ by the index voiced/unvoiced. As a result, the voiced pharyngeal [E] is pronounced as unvoiced pharyngeal [H], and the voiced uvular [G] is pronounced as

unvoiced uvular [G] (figures 3, 4). For velar and dental consonants, five of eight patients are able to distinguish voiced /unvoiced consonants properly after eleven months of rehabilitation. However, the results can not be generalized because of the limited number of subjects.

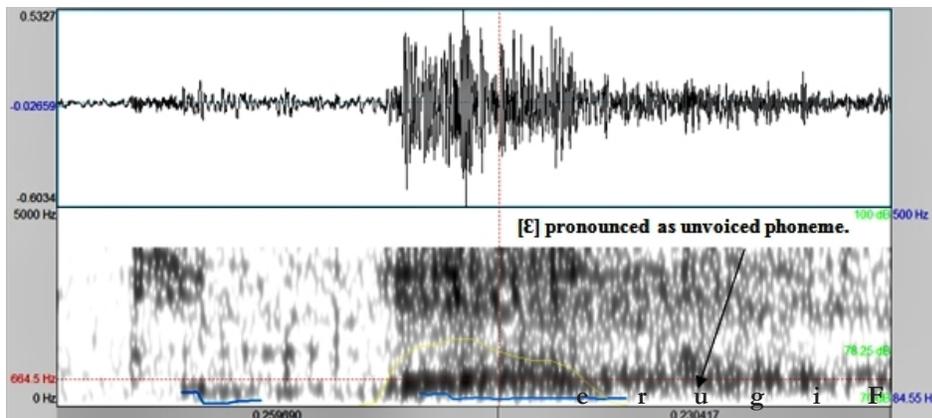


Figure 3: Pronunciation of [E] in the word [tbî•] (She sells)

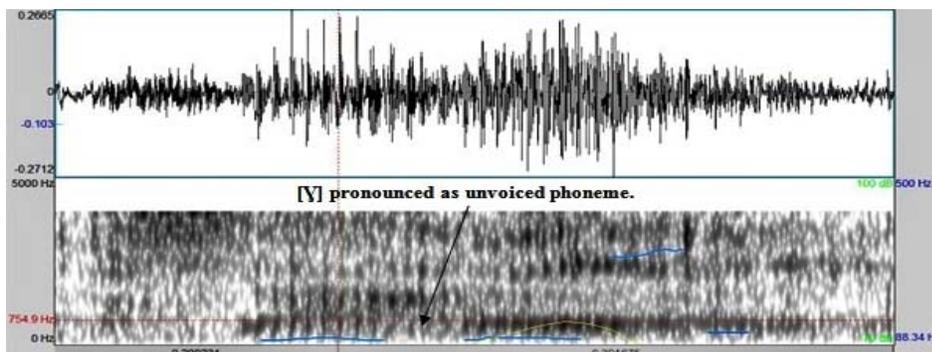


Figure 4: Pronunciation of [G] in the word [sacîR] (small)

At the beginning of the reeducation, the esophageal speech is unintelligible, and is characterized by several breaks within long sentences (as pauses necessary for air injection). Nevertheless, the speech becomes more fluid and less tense after rehabilitation. Also, the speech becomes more intelligible after eleven months of reeducation.

Discussion

Pitch F_0

Presumably, the poor ability of voicing can be explained by the shape, volume and elasticity of the PE segment which differ totally from those of vocal folds. The PE segment pattern appears rather unstable and not always periodic, because of the anatomical characteristics of the vibrating structures. We note that these results are in agreement with previous studies reported in other languages¹⁵.

Jitter and Shimmer

Tend to normal values, probably because the patient is able to better control of his efforts by a better knowledge of the various organs of phonation in his new mechanism of speech.

Harmonics/Noise Ratio (HNR)

The pathological voice is so much noised at the beginning, so the patient will probably need more time to acquire HNR values close to the threshold of normality. An effort should be done in this direction to improve the quality of speech. It is also true and natural that today, the neoglottis can never give a timbre equal to that of laryngeal voice.

Frequency formants F_1 , F_2 and F

The increase of formant values can be probably explained by the fact that the distance between the neoglottis and the first cavity of the vocal tract is modified. These results are in agreement with those previously reported for other languages^{13, 14, 15, 27, 28}.

The intensity

It's less important after laryngectomy because the quantity of air obtained by eructation is insufficient (less than 70 ml) compared with that resulting from lungs in laryngeal speech (around 5000ml).

The confusion between some consonants is in agreement with previous studies reported in the literature of the laryngectomy in general^{29,30}, unlike to other studies which report a negligible rate of confusion³¹. In our study, we note a perceptible

confusion of the back consonants unlike the front consonants. In contrast to vocal folds which have to be adducted to begin vibration, the PE segment will be relaxed to assume vibration. Thus, with the new configuration of the vocal tract, and the presence of a tracheostomy, patients seem to have difficulty in properly pronounce pharyngeal and uvular consonants with relaxing of the contiguous PE segment, and they make great efforts to properly pronounce these consonants. It appears that to pronounce these consonants, patients seem to privilege the articulation to the detriment of consonantal voicing.

The patients are unanimous to note that the vowel [i] is very difficult to pronounce, in particular in continuous speech. The high front vowel [i] is articulated with the tongue body raised and fronted towards the alveolar region. As a result of this forward thrust of the mass of the tongue, the lower pharynx is widened during pronunciation of [i]. With the changes undergone by this part of the vocal tract after surgery, it is possible that it's the articulatory constraints that make [i] more difficult to be pronounced than the other vowels.

In Conclusion, the acoustic analysis of patient's voice with TL shows that the majority of acoustic parameters tend to come close to the threshold of normality after a sufficient period of reeducation. Nevertheless, deficiencies were reported in the taking care of patients:

Absence of formation of the speech therapists in manipulation of acoustic analysis software; The exclusive use of the ear (hearing) to evaluate the effect of voice rehabilitation in the Algerian hospitals. The evaluation of pathological voice is mainly based on the subjective perception of clinicians without any acoustic or articulator analysis.

Little attention is given to acoustic analysis in the university curriculum of the student, future speech-language pathologist at the hospital.

Certainly, acoustic analysis of pathological voices cannot replace the traditional work of the speech therapist but it may be a support of objective data that can help considerably in the reeducation of patients. Indeed, this acoustic analysis reinforces the hearing message and makes objective what sometimes escapes to hearing of the speech therapist (perceptual judgment).

According to the results we found, it's important to reinforce skills and knowledge of the speech therapist, allowing him to adapt his

rehabilitation technique by giving more importance to the voicing phenomenon and the back consonants. The employment of the cepstral parameters used in this study is of special interest, as they can provide an objective diagnosis of pathological voices, and may be used as complementary tool in the monitoring of the esophageal speech rehabilitation. Indeed, the use of these acoustic parameters can be considered as a sensitive and objective outcomes measure, even with extremely perturbed voice samples that would be difficult to analyze using traditional acoustic measures, such as F_0 , duration, and formant frequencies.

This study presents some observations on quality of voice rehabilitation with esophageal speech performed in Algerian hospitals. It is noteworthy that there is a blatant lack of cooperation between the speech therapist in the Algerian hospital environment, the engineer, researcher phonetician, and acoustician in the research laboratory, and finally the professor teaching at the university. A close collaboration between these institutions will allow a better care of the laryngectomized patients.

References

1. Ward E, Van As-Brooks C. Head and neck cancer: treatment, rehabilitation, and outcomes. Ed. San Diego: Plural Publications, 2007.
2. Singer MI, Blom ED, Hamaker RC. Voice Rehabilitation after Total Laryngectomy. *J Otolaryngol* 1983; 12:329-334.
3. Koike M, Kobayashi N, Hirose H, Hara Y. Speech Rehabilitation after Total Laryngectomy. *Acta Otolaryngol* 2002; 122:107-112.
4. Stajner-Katusic S, Horga D, Musura M, Globek D. Voice and speech after laryngectomy. *Clin Linguist Phon* 2006; 20(2-3):195-203.
5. Diedrich WM. The mechanism of esophageal speech. *Annals of the New York Academy of Sciences* 1968; 155:303-317.
6. Duguay MJ. Esophageal voice: an historical review. *J Voice* 1989; 3:264-268.
7. Hocevar-Boltezar I, Zargi M. Communication after laryngectomy. *Radiol Oncol* 2001; 35:249-254.
8. Elmiyeh B, Dwivedi RC, Jallali N, Chisholm EJ, Kazi R, Clarke PM, Rhys-Evans PH. Surgical voice restoration after total laryngectomy: An overview. *Indian Journal of Cancer* 2010; 47(3): 239-247.
9. Sato T. Oesophageal Speech and Rehabilitation of the Laryngectomized. Kanehara & Co., Ltd., Tokyo 1993.
10. Kearney A. Esophageal Speech. *Otolaryngol Clin N Am* 2004; 37:613-625.
11. Dworkin JP, Meleca RJ, Zormeier MM, Simpson ML, Garfield I, Jacobs JR, Mathog RH. Videostroboscopy of the pharyngoesophageal segment in total laryngectomees. *The Laryngoscope* 1998; 108(12):1773-1781.
12. Lundström E, Hammarberg B, Munck-Wikland E, Edsberg N. The pharyngoesophageal segment in laryngectomees -videoradiographic, acoustic, and voice quality perceptual data-. *Logoped Phoniatr Vocol.* 2008; 33(3):115-25.
13. Kyttae J. Finnish oesophageal speech after laryngectomy: sound spectrographic and cineradiographic studies. *Acta Otolaryngol suppl* 1964; 195:1-94.
14. Cervera T, Miralles JL, González-Alvarez J. Acoustical analysis of Spanish vowels produced by laryngectomized subjects. *J Speech Lang Hear Res* 2001; 44:988-96.
15. Liu H, Wan M, Wang S, Wang X, Lu C. Acoustic characteristics of Mandarin esophageal speech. *J Acoust Soc Am.* 2005; 118:1016-1025.
16. Gang Lv. Acoustical analysis of Chinese vowels produced by alaryngeal speaker. *Applied Mechanics and Materials* 2011; 48-49:135-138.
17. Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Algerian_Arabic
18. Yumoto E, Sasaki Y, Okamura H. Harmonics-to-noise ratio and psycho physical measurement of the degree of hoarseness. *J Speech Hear Res* 1984; 27:2-6.
19. Heiberger VL, Horii Y. Jitter and Shimmer in sustained phonation. *Speech and Language: Advances in Basic Research and Practice* 1982; 7:299-332.
20. Munoz J, Mendoza E, Fresneda MD, Carballo G, Lopez P. Acoustic and perceptual indicators of normal and pathological voice. *Folia Phoniatr Logop* 2003; 55:102-114.
21. Kreiman J, Gerratt BR. Perception of aperiodicity in pathological voice. *J Acoust Soc Am* 2005; 117:2201-2211.
22. Naufel de Felipe AC, Grillo MH, Grechi TH. Standardization of acoustic measures for normal voice patterns. *Rev Bras Otorrinolaringol* 2006; 72:659-664.

23. Oppenheim AV, Shafer RW. From frequency to Quefrency: A history of the cepstrum. *IEEE Signal Processing Magazine* 2004; 95-106.
24. Noll AM. Ceptsrum Pitch Determination. *J. Aconst. Soc. Amer.* 1967; 41(2): 293-309.
25. Titze IR, Liang H. Comparison of f0 extraction methods for high-precision voice perturbation measurements. *J Speech Hear Res* 1993; 36(6):1120–1133.
26. Murphy PJ, Akande OO. Noise Estimation in Voice Signals Using Short-term Cepstral. *Journal of the Aoustical Society of America* 2007; 121(3):1679-1690.
27. Sisty NL, Weinberg B. Formant frequency characteristics of esophageal speech. *J Speech Hear Res* 1972; 15:439-448.
28. Ng ML, Chu R. An Acoustical and Perceptual Study of Vowels Produced by Alaryngeal Speakers of Cantonese. *Folia Phoniatr Logop* 2009; 61:97-104.
29. Nichols AC. Confusions in recognizing phonemes spoken by esophageal speakers: II. Vowels and diphthongs. *J Commun Disord* 1976; 9:247-260.
30. Crevier-Buchman L, Vaissière J, Maeda S, Brasnu D. Etude de l'intelligibilité des consonnes du français après laryngectomie partielle supracricoidienne. *Rev Laryngol Otol Rhinol* 2002; 123:307-310.
31. Hirose H. Voicing Distinction in esophageal Speech. *Acta Otolaryngol suppl* 1996; 524:56-63.