ON PREDICTING PATIENT'S VOICE AFTER SURGICAL OPERATION

Cheolwoo Jo*, Daehyun Kim*, Moojin Baek**, Sugeon Wang***
*Changwon National University, **Inje University,
***Busan National University
KOREA

ABSTRACT: This paper describes a procedure to predict a patient's voice after surgical operation. To do this, the voice before and after surgical operation is collected from the same patient. Collected voice is analyzed to obtain differences affected by surgical operation. To measure the change of acoustical characteristics of voice, jitter, shimmer and other spectral domain and time domain parameters are computed and compared. According to the result, it is shown that the factors that change are caused not only by vocal fold components but also by vocal tract. One method to implement the predictive synthesis of voice after surgery, residual excited PSOLA method is applied. The resulting voice is compared to the voice after surgery in terms of spectral and perceptual similarity.

1. INTRODUCTION

Recently interest on the human health is increasing. And speech is a basic mean to human communication. In many cases, diseases at the vocal folds cause the change of voice quality. This is caused by tumor, folip, swelling, hardening of vocal folds and its vicinity etc. In some cases, patients are asked to take surgery. Patients, who have distorted voice, generally want to know how much their voice can be improved after surgery. Also doctors want to let them know the possible improvements and make them have peace in mind. This process very important.

The relationship at the quality change of the voice before and after surgery is not well known. Some doctors can predict the status of patient's neck by hearing speech only. Based on this fact, there are researches to diagnose patient's voice only by acoustic signal. (Jo & Kim,1998) (Koike, Takahashi , Calcatera,1977) (Issiki, Okamura, Tanabe , Morimoto,1969) (Murry, Singh , Sargent,1977) (Hammarberg, Fritzell , Gaffin,1967) Some of them are based on the perceptual distinctions and subjective tests, but there are some which are based on the objective or numerical methods. Predicting voice after surgery is not reported much yet on the literature.

In this paper changes of various parameters from voices before and after surgery. Based on the observed phenomena, Predictive synthesis of voice after surgery is conducted. And the procedures are introduced.

2. PATHOLOGICAL VOICE DATA

Voice data is collected from the patients at ENT department of hospital. Range of patients' age is 50-60. Names of diseseas are mostly polyp with two cases of pholyposis. All the patients are diagnosed to be benign by several tests and taken surgery afterwards. Collection is performed in silent room. 5 vowel sounds (/a/, /e/, /i/, /o/, /u/) and one sentence sound are collected. Patients are asked to pronounce the vowels consecutively for 3 seconds with proper silence between utterances. Microphone is positioned at 15 cm front of the mouth. Spoken voice is recorded using Sony's DAT(DTC-59ESJ) recorder. Collected materials are stored in wav file format. Pathological data is typically noisy and husky. It is easily recognizable voice after surgery is less noisy and clearer.

3. ANALYSIS OF VOICE

Speech materials are analyzed by Kay's MDVP software package. Total 14 different parameters are computed. Those are as follows. In terms of short-term and long-term frequency disturbance, Jita, Jitt, RAP, PPQ, sPPQ, vF0. In terms of short-term and long-term amplitude perturbation, ShdB, Shim, APQ, sAPQ, vAm. In terms of noisy components, NHR, VTI, SPI.

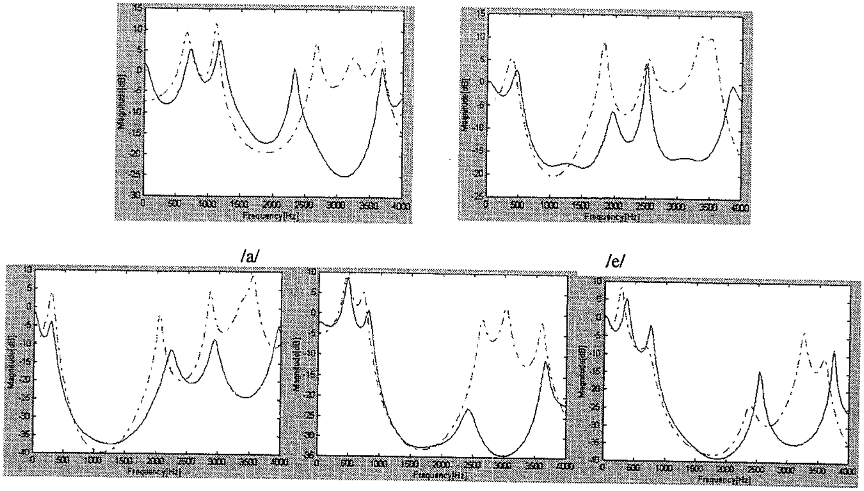
Table.1 shows relative changes of each parameters. It is shown that voice after surgery has less jitter and shimmer in general. It is a quite predictable result based on perceptual difference.

Next changes of spectrum is computed using lpc and fft method. This is to examine the change of spectral components and shift of formants etc. Figure2 shows change of spectrum from two patients. LPC spectrum of five vowels are shown. From the figures the amplitude of speaker1's higher formant components are much higher after surgery than before it. But in speaker2's case the raise is not dominant compared to speaker1's. That means there is no constant tendency of changes on vocal tract characteristics. Table2. shows the mean and standard deviations of formant frequency changes. Mean changes are bigger at higher formants. But their standard deviation is too big, so there was no general tendency in terms of formant shifts. But peak amplitude values are lowered in high frequency region and slight downward shifting of higher formant frequency is observed in many cases. One of the possible reason for this phenomena is the change of patient's speaking pattern after surgery.

Also it is observed that the amplitudes of higher formant components, i.e. frequency range over 1500Hz, is lowered in many cases. Lowering higher formant component is similar to de-emphasis effect. Perceptually it can be related to smoother voice.

Table.1 Relative changes of parameters

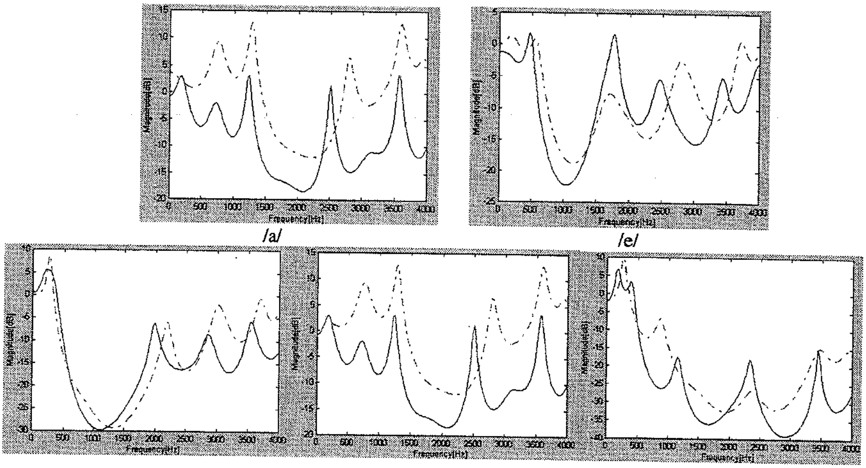
	Before		After	
	Average	Relative Ratio	Average	Relative Ratio
Jita	218.38	2.62	109.81	1.32
Jitt	3.02	2.91	1.38	1.33
RAP	1.69	2.49	0.72	1.07
PPQ	1.67	1.98	0.707	0.84
sPPQ	2.13	2.09	0.92	0.90
VF0	4.51	4.10	2.67	2.43
ShdB	0.92	2.61	0.437	1.25
Shim	7.42	1.95	3.79	0.99
APQ	5.07	1.65	3.09	1.00
sAPQ	5.83	1.38	4.09	0.97
vAm	12.64	1.54	9.79	1.19
NHR	0.17	0.18	0.14	0.15
VTI	0.07	1.22	0.05	0.86
SPI	28.31	2.00	27.64	1.96



/a/ /e/

/i/ /u/ /o/

(a) Speaker1's lpc spectrum comparison



/a/ /e/

/i/ /u/ /o/

(b) Speaker2's lpc spectrum comparison

(Broken line: before, Solid line: after)

Figure1. LPC spectrums for two speakers

Table2. Changes of formant frequences

		/a/	/e/	/i/	/o/	/u/
F1	mean	-28.8	-21.3	2.5	7.7	30.9
	SD	137.1	80.8	45.9	64.8	66.6
F2	mean	24.8	49.8	-34.6	21	-15.1
	SD	153.1	84.3	158.9	106.1	84.5
F3	mean	-5.8	-39.3	-76.1	-38.2	-71.8
	SD	205.1	168.8	175.46	176.7	212.6

4. RE-PSOLA SYNTHESIS SCHEME

To implement the changes of voice, RE-PSOLA(residual excited PSOLA) is used to re-synthesize voice. From the measured acoustic characteristics, it is shown that both source and vocal tract characteristics are changing after surgery. To be able to change source and vocal tract component effectively, generally source-filter model is required. But separating true source components are not easy and the quality of reproduced sound is not good as natural voice. PSOLA method gives natural voice quality if pitch variation is within about 20 percent range and easy to implement. But conventional PSOLA cannot effectively change the source characteristics. One modified version of PSOLA is RE-PSOLA. Its basic structure is the same as conventional PSOLA method with only difference that in this method decomposition is done at the residual signal level not at speech signal itself. It can be thought as a kind of pitch-synchronous linear prediction method but it doesn't require explicit pitch synchronization and maintains each period's characteristics of speaker. Although error signal is a little different from real source, it can be thought as a pseudo source.

Figure2 shows the block diagram of used analysis and synthesis scheme.

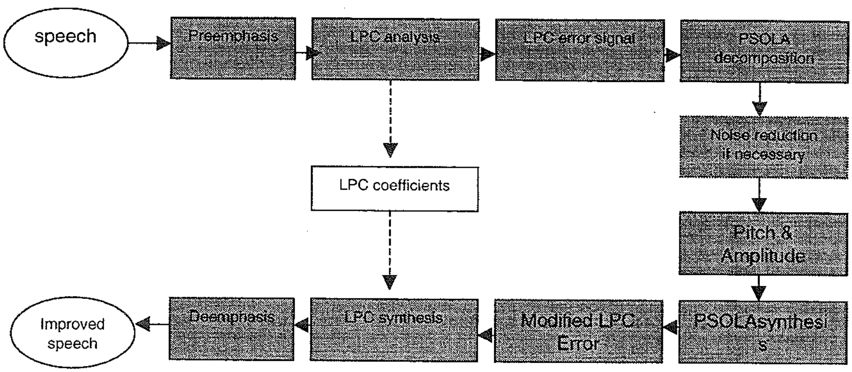


Figure2. residual excited PSOLA analysis and synthesis scheme

5. FEATURE ESTIMATION , SYNTHESIS AND DISCUSSIONS

Based on the suggested analysis and synthesis scheme, feature parameters are estimated based on acoustic measurements. Based on the statistics of parameters, pitch and spectral components are modifies in systematic way. 4 basic categories which is considered to be significant are jitter, shimmer, noise components and spectral change.

At first, jitter values can be modified by adjusting the consecutive pitch values. Let a and b the consecutive pitch values and c is moving average of pitch track. And if we call R change ratio, new pitch value is as eq (1). Computed jitter value become close to the desired one.

$$\begin{aligned} b' &= c+(b-c)*R \\ a' &= c+(c-a)*R \end{aligned} \quad \text{, if } b > c > a \qquad (1)$$

Reducing noise components are performed by choosing different window shapes during the PSOLA decomposition. In terms of spectral change, -6dB/oct low-pass filter is used to emphasize low frequency components. In this work other parameters, except jitter, is decided empirically based on the previous analysis. At this stage modifying source characteristics, such as open quotient values, is not considered.

Hypothetical voice was synthesized using modified parameters. Changed voice showed some parameter changes in desirable way. But their perceptual quality is still not comparable to the patient's voice after surgery. Mean opinion test over 10 people with 5 scale showed only average score of 3.3.

6. CONCLUSIONS

In this paper procedures of collecting, analyzing and re-synthesizing methods with voice before and after surgery on vocal tract was described.

First patients' data was collected before and after surgery. Then the changes of parameters are computed. Based on the parameters, voice before surgery is converted to generate cured voice.

From the comparisons between voice before and after surgery, it is observed that the change of acoustical characteristics is caused not only by source components but also by vocal tract components. And it is hard to predict the possible change of the voice due to the wide diversity of parameter change according to each patient. But there was definite improvements toward normal voice in most of the acoustic parameters. By relating the changes of the parameters to the ones of synthesizer, weak mapping of signal is possible.

Although the current result is not satisfactory, this experiment showed a possibility of predicting patient's voice after surgery. In future research, more precise mapping and quantitative control of synthesizer will be conducted. Also relations between physical change of vocal folds and acoustical signal have to be investigated further.

REFERENCES

(1967) Hammarberg,B , Fritzell,B , Gaffin,J , Sundberg,J , Weidin,L , "Perceptual and acoustic correlates of abnormal voice quality", Acta Otolaryngology, 90, pp.441-451

(1969) Issiki,N, Okamura,H , Tanabe,M , Morimoto,M , "Differential diagnosis of hoarseness", Folia Phoniatrics, 21, pp.9-19

(1998) Jo,C.W., Kim,D.H. , "Analysis of acoustic characteristics of pathological speech by source analysis", Proceedings of Summer Conference, pp.163-166, Acoustical Society of Korea

(1977) Koike,Y, Takahashi,H, Calcatera,TC, “Acoustic measurements for detecting laryngeal pathology”, Acta Otolaryngology, 85, pp.105-117

(1977) Murry,T , Singh,S , Sargent,M , “Multidimensional classification of abnormal voice qualities”, JASA, 61, pp.1630-1635

(1993) Operations Manual, “Multi-Dimensional Voice Program(MDVP)”, Model 4305, Kay Elemetrics Corp, 1993