

음성합성을 이용한 병적 음성의 치료 결과에 대한 예측

연세대학교 의과대학 이비인후과학교실, 음성언어의학연구소
이주환 · 최홍식 · 김영호 · 김한수 · 최현승 · 김광문

= Abstract =

Prediction of Post-Treatment Outcome of Pathologic Voice Using Voice Synthesis

Joo Hwan Lee, M.D., Hong-Shik Choi, M.D., Young-Ho Kim, M.D.,
Han Soo Kim, M.D., Hyun Seung Choi, M.D., Kwang-Moon Kim, M.D.

*Department of Otorhinolaryngology, The Institute of Logopedics & Phoniatrics,
Yonsei University College of Medicine, Seoul, Korea*

Background and Objectives : Patients with pathologic voice often concern about recovery of voice after surgery. In our investigation, we give controlled values of three parameters of voice synthesis program of Dr. Speech Science. such as jitter, shimmer, and NNE(normalized noise energy) which characterize someone's voice from others and devised a method to synthesize the predicted voice after performing operation.

Subjects and Method : Values of vocal jitter, vocal shimmer, and glottal noise were measured with voices of 10 vocal cord paralysis and 10 vocal polyp patients 1 week prior to and 1 month after the surgery. With Dr. Speech science voice synthesis program we synthesized 'ae' vowel which is closely identical to preoperative and post-operative voice of the patients by controlling the values of jitter, shimmer, and glottal noise, then we analyzed the synthesized voices and compared with pre and post-operative voice.

Results : 1) After inputting the preoperative and corrected values of jitter, shimmer, and glottal noise into the voice synthesis program, voices identical to vocal polyp patients' pre- and postoperative voices within statistical significance were synthesized 2) After elimination of synergistic effects between three parameter, we were able to synthesize voice identical to vocal paralysis patients' preoperative voices. 3) After inputting only slightly increased jitter, shimmer into the synthesis program, we were able to synthesize voice identical to vocal cord paralysis patients' postoperative voices.

Conclusion : Voices synthesized with Dr. Speech science program were identical to patients' actual pre and postoperative voice, and clinicians will be able to give the patients more information and thus increased patients cooperability can be expected.

KEY WORDS : Voice synthesis · Pathological voice · Vocal cord paralysis · Vocal polyp.

논문접수일 : 2003년 6월 3일

심사완료일 : 2003년 6월 15일

책임저자 : 김광문, 서울 강남구 도곡동 146-92 연세대학교 의과대학 이비인후과학교실, 음성언어의학연구소
전화 : (02) 3497-3460 · 전송 : (02) 3463-4750 E-mail : kmkim97@yumc.yonsei

서 론

성대마비와 성대폴립환자에게 가장 유효한 치료방법은 지금까지 수술적 처치라고 알려져 있다. 병변에 대한 수술 전 음성검사와 후두내시경검사를 이용한 설명은 가능하나 음성장애로 수술을 받는 환자들의 경우, 음성장애의 원인이 되는 병변의 완전한 제거 또는 교정과 함께 수술 후에 어느 정도로 음성을 회복할 수 있는가에 대하여 여전히 많은 의문을 가지게 된다. 성형외과에서 코의 성형이나 턱 교정 수술 전에 환자의 사진을 컴퓨터로 합성하여 수술 후 예측되는 모습을 모니터로 보여 주면서 설명하는 것과 같이, 음성 장애로 수술을 받는 환자에게 수술 후에 예측되는 음성을 들려주는 것이 가장 효과적인 방법일 것이다.

음성의 특성을 결정하는 중요변수로는 지터(jitter), 쉬머(shimmer), 성대잡음(NNE(normalized noise energy)) 등이 있다. 성문과형 주기의 빠른 진폭 변화를 나타내는 지터는 성대의 진동주기성을 감소시키는 병변으로 인하여 발생하는 음성의 거칠음(vocal roughness)을 대변하는 수치이다. 연속되는 주기들 사이의 평균진폭의 차이라고 정의되는 쉬머 역시 성대진동 양식의 불규칙성을 유발하는 인자를 반영하는 수치로 여겨진다.¹⁾ 성대잡음은 부분적으로 개방된 성문에서 발생하는 음색과 거친 기류의 조합을 반영하는 수치이다.²⁾

연구자들은 성대 폴립과 성대마비환자에게 있어서 ① 수술 전과 수술 후에 음향학적 분석을 시행하고 ② 수술 전 음성분석 결과를 토대로 Dr. Speech science 음성합성 프로그램을 이용하여 수술 전 음성을 합성하고 그리고, ③ 합성된 수술 전의 병적 음성에 인위적으로

Voice Data (Start : 1.2s End : 2.4s)			
Habitual F0(Hz)	199.90	NNE(dB)	-0.02
Jitter(%)	2.51	HNR(dB)	14.06
Shimmer(%)	9.84	SNR(dB)	12.91
F0 Tremor(Hz)	10.96	Amp Tremor(Hz)	10.96
Mean F0(Hz)	197.21	MPT(s)	0.00
SD F0(Hz)	6.06	s/z ratio	0.00
Max F0(Hz)	217.24	Ratio(%)	39.00
Min F0(Hz)	180.74		

Fig. 1. Results of analysis of vocal paralysis patient's voice by Dr. Speech science voice analysis program 1 month prior to surgery. The values in rectangles are vocal jitter, shimmer, and glottal noise(NNE).

이들 변수를 변화시켜 수술 후에 얻어지는 음성과 유사한 음성을 합성하여 음성합성 프로그램에 의한 합성음과 환자의 실제 수술 후 음성을 비교함으로써 합성음의 신뢰도와 임상에서의 유용성을 알아보자는 것이 이 연구의 목적이다. 합성된 수술 후 음성을 환자에게 제시함으로써 수술 후 얻어지는 음성에 대한 정보를 미리 제시하고, 수술 후의 치료성적을 판단할 수 있는 지표로 제시하려는 것이다.

연구대상 및 방법

1. 대 상

2001년 12월부터 2002년 12월까지 음성장애를 주소로 영동세브란스 이비인후과에 내원하여 후두내시경검사를 거침으로써 성대마비 진단을 받은 환자 10례와 음성검사 후 후두미세술을 시행 받은 환자 중 수술 후 조직검사에서 성대폴립으로 확진된 환자 10례를 대상으로 하였다. 성대마비환자 중 성대의 휘어짐이 있거나 일측성성대마비 외의 다른 병변을 가진 환자는 대상에서 제외하였다. 성대폴립환자군의 경우 6명의 여자, 4명의 남자로 구성되었으며 평균나이는 41.7세였다(범위, 28~63세). 성대마비환자군의 경우 7명의 여자, 3명의 남자로 구성되었으며 평균나이는 47.2세였다(범위, 25~67세).

2. 방 법

1) 수술 전 음성분석

두 환자 군을 대상으로 수술 1주 전 Dr. Speech Science 음성분석 프로그램(Tiger DRS, Inc. Version 4)을 이용하여 음성분석을 시행하였다. 주파수가 음성변수에 영향을 주므로 수술 전,후의 기본주파수는 비교적 같게 유지하도록 노력하였다.³⁾ 음성 차단이 된 방 안에서 마이크로폰에서 환자의 입까지 5cm의 간격을 둔 후 2초 동안 편안한 크기와 높이의 목소리로 '에' 모음을 발성한 후 첫 0.25초와 마지막 0.25초는 발성 개시시 발생하는 'burst'와 발성 종료 시 발생하는 'pitch break'를 방지하기 위하여 분석에서 제외하고⁴⁾ 음성분석을 실시하여 지터, 쉬머, 성대잡음 값을 측정하였다(Fig. 1, 2).

2) 수술 전 음성의 합성

Dr. Speech Science 음성합성 프로그램을 이용하여

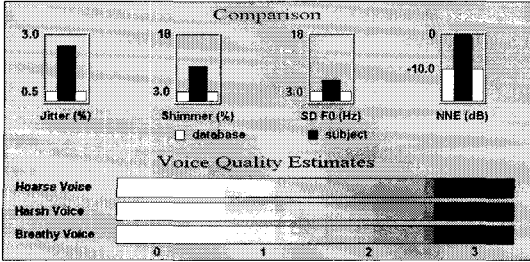


Fig. 2. Results of analysis of vocal paralysis patient's voice by Dr. Speech science voice analysis program 1month prior to surgery. Vocal jitter, shimmer, NNE are all out of normal range.

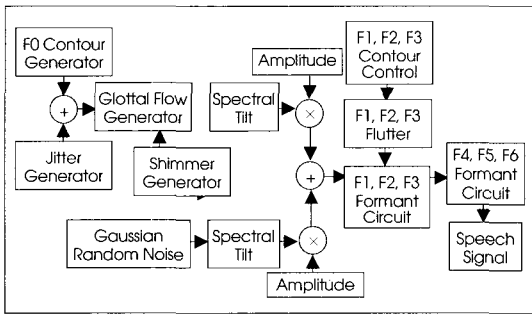


Fig. 3. Voice synthesis model propose by Klatt.

지터, 쉬머, 성대잡음 수치를 변화시켜 두 질병의 수술 전 병적인 음성과 유사한 '애' 모음을 합성한 후 음성분석을 실시하여 환자의 실제 수술 전 음성과 비교하였다.

3) Dr. Speech science 음성합성 프로그램(Voice synthesis and therapy)을 이용한 모음의 합성

수술 전, 후 음성의 합성은 Klatt이 제시한 '후두원 음모델(glottal source model)'에 기초된 음성합성 프로그램에 의하여 시행하였으며, 음성합성 단계는 다음 그림과 같다(Fig. 3).⁵⁾

모음을 합성하기 전 'Voice synthesis and therapy' 프로그램의 주된 구성요소를 차지하는 'Voice synthesis dialog box' 내의 조절변수인 모음변수들에 대한 지정은 아래와 같다(Fig. 4).⁶⁾

(1) 시간

(가) 표본 주파수(Sampling frequency(Hz))

합성된 음성의 출력표본율은 44100samples/second(Hz)로 제시되었다. 이는 44100Hz의 표본주파수에서 출력파의 주파수 성분은 22050Hz까지 이루어질 수 있

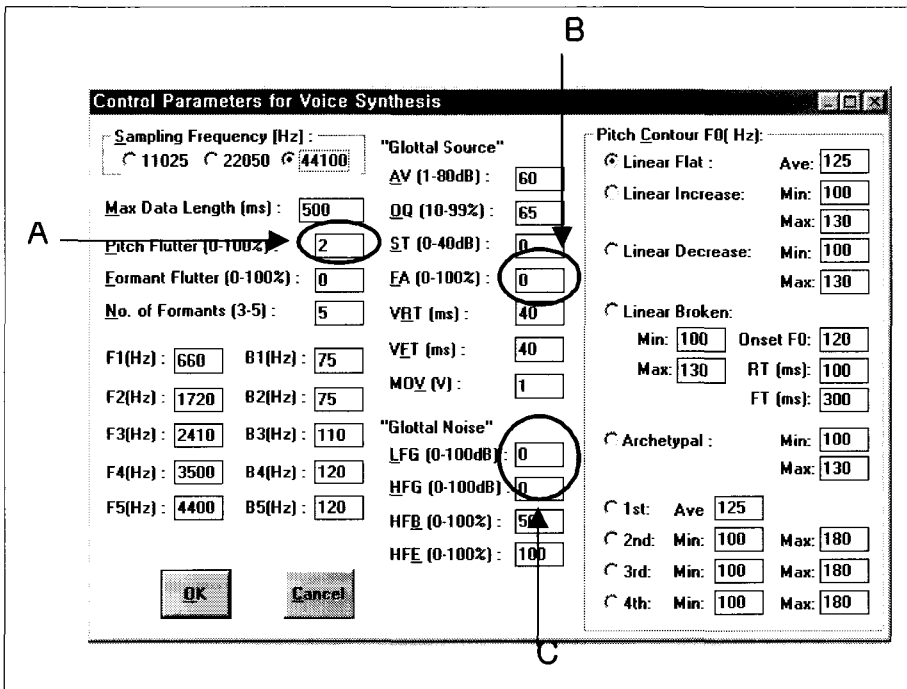


Fig. 4. The control parameters of Dr. Speech science voice synthesis program. A : Pitch flutter (representing jitter), B : FA (representing shimmer), C : HFG, LFG (representing glottal noise).

으며 이는 표본이론에 의한 고음질의 음성신호재현율을 보증 할 수 있는 상부 경계치이다.

(나) 최대 자료기간(Maximum data length)

합성된 발성의 최대 음절기간이며 msec로 표시된다. 이 실험에서 사용된 기간은 2,000msec(2sec)이다.

(다) VRT : Voice rising time(ms)

음성진폭상승시간을 의미하며 40msec로 주어졌다.

(라) VFT : Voice falling time(ms)

음성진폭하강시간을 의미하며 역시 40msec로 주어졌다.

(2) 포만트(Formants)

(가) 포만트 수

포만트 수는 5로 주어졌으며 이는 정상 성인남자의 발성관의 길이(17cm)와 부합되는 수이다.

(나) 포만트 주파수

환자의 음성을 Dr. Speech Science 'Real analysis' 프로그램의 'formant tracking program' 을 이용하여 분석하였다(Fig. 5).

(3) 후두원음(Glottal source)

(가) Pitch flutter(%) : pitch period flutter

기본주파수의 임의의 변동을 의미하며 이상적으로 일

정한 주기성을 가지는 파동의 합성을 방지하는 역할을 가지므로 지터율(jitter percent)은 사용자가 정의하는 피치변동(pitch flutter) 수치에 의해 영향을 받는다.

(나) AV : amplitude of voicing(dB)

60dB의 값이 모음합성에 있어서 최대 비과부화 음성 높이기므로 이 수치를 정하였다.

(다) OQ : Open quotient(%)

Open quotient는 성문주기의 전체기간분의 개방시간을 의미하며 첫 번째 harmonic의 진폭에 영향을 준다. 제시된 값은 65이다.

(라) FA : Flutter of amplitude(%)

성문근원 진폭의 임의의 변동을 의미하며 이상적으로 일정한 성문근원 진폭을 가지는 파동의 합성을 방지하는 역할을 가지므로 쉬머율(shimmer percent)은 사용자가 정의하는 강도변동(amplitude flutter) 수치에 의해 영향을 받는다.

(4) 후두잡음 근원(Glottal noise source)

(가) MFG : 저주파수 이득(low frequency gain) (dB)

저주파수영역의 잡음근원 이득의 값을 의미하며 이 수치가 높아질수록 저주파수의 상대잡음수치가 증가된다.

(나) HFG : 고주파수 이득(high frequency gain) (dB)

고주파수영역의 잡음근원 이득의 값을 의미하며 이 수

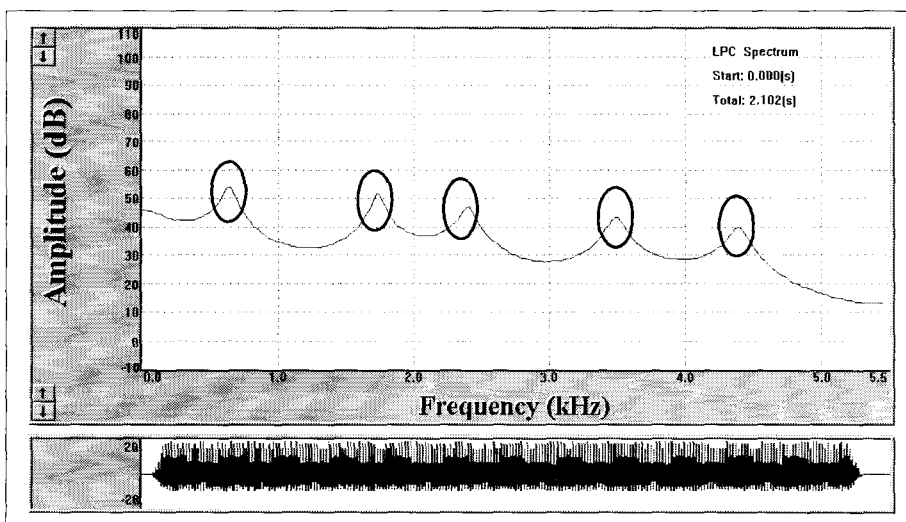


Fig. 5. The formant frequencies of the patients which were measured with Dr. Speech Science 'Formant tracking program'. These formant frequencies were inputted into the voice synthesis program.

치가 높아질수록 고주파수의 성대잡음수치가 증가된다.

(다) HFB and HFE(High frequency gain begin and end(%))

50과 100으로 주어졌으며 이는 고주파수영역의 시작점이 11025Hz이며 종말점이 22050Hz임을 의미한다.

4) 수 술

성대 마비환자의 경우 전신마취 하에 Type I thyroplasty와 arytenoid adduction을 시행하였으며 성대폴립환자의 경우 전신마취 하에 Kleinsasser 후두경와 Zeiss 현미경을 이용하여 후두미세수술을 시행하였다.

5) 수술 후 음성분석

두 환자 군을 대상으로 수술 한 달 후에 Dr. Speech Science 음성분석 프로그램(Tiger DRS, Inc. Version 4)을 이용하여 음성분석을 시행하였다. 수술 한 달 후 음성검사를 시행한 이유는 수술 후 발생하는 부종의 흡수를 기다리기 위하여서였다.⁷⁾ 수술 전 음성검사와 동일한 방법으로 음성분석을 실시하여 지터, 쉬머, 성대잡음 값을 측정하였다.

6) 수술 후 음성의 합성

Dr. Speech Science 음성합성 프로그램을 이용하여

두 질병의 합성된 수술 전 병적인 음성의 지터, 쉬머, 성대잡음 수치를 변화, 교정하여 두 질병의 수술 후의 음성과 유사한 음성을 합성한 후 합성된 수술 후 음성을 분석하여 환자의 실제 수술 후 음성과 비교하였다(Fig. 6).

7) 통계처리

모든 결과는 평균으로 표시하였다. 통계적 분석은 Mann-Whitney U test를 이용하였으며, p 값이 0.05 이상인 경우를 유의하다고 판정하였다.

결 과

1. 성대폴립환자의 수술 전 음성의 합성

Dr. Speech Science 음성합성 프로그램을 이용하여 성대폴립환자의 수술 전 음성과 유사하도록 합성한 /애/모음의 평균 지터값은 0.56%, 쉬머값은 4.42%, 성대잡음값은 -4.07dB이었고, 이는 수술 전 환자의 /애/음성에서 측정된 세 변수의 평균수치(지터 : 0.64%, 쉬머 : 4.69%, 성대잡음 : -4.47dB)와 통계학적으로 유의하였다(p value>0.05)(Table 1).

2. 성대폴립환자의 수술 후 음성의 합성

수술 후 음성과 유사하도록 합성한 /애/ 모음의 평

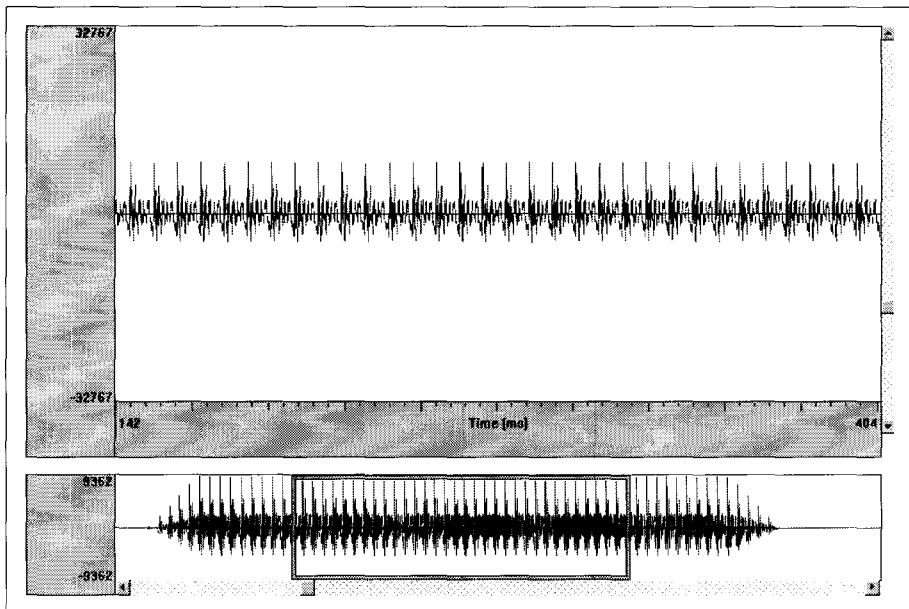


Fig 6. Synthesized postoperative voice of patient with vocal polyp by using Dr. Speech Science voice synthesis program.

균 지터값은 0.31%, 쉬머 값은 1.77%, 성대잡음값은 -12.96dB이었고, 이는 수술 후 환자의 /애/ 음성에서 측정된 세 변수의 평균수치(지터 : 0.33%, 쉬머 : 1.91%, 성대잡음 : -13.08dB)와 통계학적으로 역시 유의하였다(p value>0.05) (Table 2).

3. 성대마비환자의 수술 전 음성의 합성

Dr. Speech Science 음성합성 프로그램을 이용하여 성대마비환자의 수술 전 음성과 유사하도록 합성한 /애/ 모음의 평균 지터값은 3.99%, 쉬머 값은 10.12%, 성대잡음 값은 -1.31dB이었고 이는 수술 전 환자의 /애/ 음성에서 측정된 세 변수의 평균수치(지터 : 4.16%, 쉬머 : 10.27%, 성대잡음 : -1.36dB)와 통계학적으로 유의하였다(p value>0.05) (Table 3).

성대마비환자의 수술 전 음성을 합성하는 경우, 수술 전 음성분석에서 측정된 지터, 쉬머와 성대잡음 값을 음성합성 프로그램에 입력하여 합성된 음성을 다시 음성분석하자, 지터, 쉬머, 성대잡음 값 모두 실제 환자의 수술 전 음성의 측정치보다 증가되어 있는 수술 전 음성 과 전혀 다른 음성이 합성되었다. 이는 정상범위를 과도로 초과한 세 수치들이 서로 상호작용을 유발하여 증가된 결과로 판단된다. 세 변수 모두를 입력하여 합성한 음성과 두 변수만을 입력하여 합성한 음성과의 차이점을 찾아, 두 변수가 나머지 한 변수에 미치는 영향을 알아보았다. 그리고, 이같은 상호작용을 교정한 후 교정된 순수한 세 변수를 다시 음성합성 프로그램에 입력 시 수술 전 음성과 유사한 음성을 합성해 낼 수 있었다.

Table 1. Comparison between acoustical parameters of measured voice and synthesized voice of patients with vocal polyp 1 month prior to surgery

	Measured	Synthesized	p
Jitter(% : normal<0.5) ¹	0.64	0.56	>0.05*
Shimmer(% : normal<3) ¹	4.69	4.42	>0.05*
NNE(dB : normal<10) ¹	-4.47	-4.07	>0.05*

1.Normal value of jitter, shimmer, and NNE were given by Dr. Speech science manual.

* : Significant Mann-Whitney test (p>0.05)

Table 2. Comparison between acoustical parameters of measured voice and synthesized voice of patients with vocal polyp 1 month after surgery

	Measured	Synthesized	p
Jitter(% : normal<0.5) ¹	0.33	0.31	>0.05*
Shimmer(% : normal<3) ¹	1.91	1.77	>0.05*
NNE(dB : normal<-10) ¹	-13.08	-12.96	>0.05*

1.Normal value of jitter, shimmer, and NNE were given by Dr. Speech science manual.

* : Significant Mann-Whitney test (p>0.05)

Table 3. Comparison between acoustical parameters of measured voice and synthesized voice of patients with vocal paralysis 1 month prior to surgery

	Measured	Synthesized	p
Jitter(% : normal<0.5) ¹	4.16	3.99	>0.05*
Shimmer(% : normal<3) ¹	10.27	10.12	>0.05*
NNE(dB : normal<-10) ¹	-1.36	-1.31	>0.05*

1.Normal value of jitter, shimmer, and NNE were given by Dr. Speech science manual.

* : Significant Mann-Whitney test (p>0.05)

Table 4. Comparison between acoustical parameters of measured voice and synthesized voice of patients with vocal paralysis 1 month after surgery

	Measured	Synthesized	p
Jitter(% : normal<0.5) ¹	1.03	0.98	>0.05*
Shimmer(% : normal<3) ¹	3.30	3.18	>0.05*
NNE(dB : normal<-10) ¹	-8.88	-8.70	>0.05*

1.Normal value of jitter, shimmer, and NNE were given by Dr. Speech science manual.

* : Significant Mann-Whitney test (p>0.05)

4. 성대마비환자의 수술 후 음성의 합성

수술 후 음성과 유사하도록 합성한 /에/ 모음의 평균 지터값은 0.98%, 쉬머값은 3.18%, 성대잡음값은 -8.70dB 이었고 이는 수술 후 환자의 /에/ 음성에서 측정된 세 변수의 평균수치(지터 : 1.03%, 쉬머 : 3.30%, 성대잡음 : -8.88dB)와 역시 통계학적으로 유의하였다(p value > 0.05) (Table 4).

성대마비환자의 경우 합성된 수술 전 음성의 지터, 쉬머, 성대잡음 값을 정상범위로 교정 후 이 수치를 음성 합성 프로그램에 입력하여 수술 후 음성을 합성한 결과 세 변수에서 모두 환자의 실제 수술 후 음성의 측정치보다 감소된 결과를 보였다. 다시 음성합성 프로그램에 정도의 지터, 쉬머 값과 정상 범위로 교정된 성대잡음 값을 입력하여 합성된 음성은 성대마비환자의 수술 후 음성과 유사하였다.

고 찰

음성인식, 음성합성, 음성입력, 음성명령, 음성보안 등 음성과 관련된 정보통신기술은 음성게임, 교육, 반도체 칩, 보안 솔루션, 사이버 캐릭터, 음성번역 전자사전, e 북, 홈오토메이션 등의 사업에서 빠르게 발전하고 있다. 이 가운데 음성합성은 여러 분야, 특히 정보통신 분야에서 많이 응용되는데, 텍스트를 음성으로 바꿔주는 TTS (Text-to-Speech)와 UMS 서비스에서 문자메일을 핸드폰을 통해 들려주는 서비스에 이용되며 통화음질의 향상, 음성인식 프로그램의 개발 등 공학적인 측면에서 그 연구가 활발히 진행되고 있다.⁸⁾

임상에서 음성신호처리가 이용되는 분야로는 음성신호 분석, 음성 합성, 음성 합성으로 얻어진 음성신호의 분석으로 구분할 수 있다. 이비인후과에서는 음성신호 분석을 통하여 병적음성에 대한 객관적인 자료를 얻어왔고, 임상적으로도 일반적인 음성 클리닉이나 음성 검사실에서 일상적으로 검사하는 항목 중의 하나이다. 그러나 음성의 합성분야나 음성합성에 의하여 얻어진 음성신호 분석분야는 상대적으로 이비인후과나 음성치료 분야에서 관심을 받지 못하였는데, 그 이유는 아마 임상적인 효용성을 찾기 힘든 데서 비롯한 것으로 사료된다.

연구자는 음성의 합성으로 병적음성을 합성한 후, 지터, 쉬머, 성대잡음의 세 변수를 각각 변화시켜 새로운 음성을 실험적으로 합성해 보았으며 이를 통하여 어떤

변수의 변화를 주었을 때 수술 후 음성과 유사한 음성이 합성되는지를 알아보았다. 본 연구에서는 진단목적이 아닌, 후두질환으로 인한 음성장애 환자에서 수술을 시행한 후 어떤 음성을 낼 수 있을지를 예측하는 방법을 연구하였다.

음성의 음향학적 특성을 결정하는 변수로는 발성법, 음질, 음의 강도와 주파수의 변화, 성대잡음, 포만트 등이 있으며⁹⁾ 이중 음성 질환과 관련하여 많은 연구자들의 관심 대상이 된 변수는 지터, 쉬머와 성대잡음이다. 병적 음성의 낮은 신호 대 잡음비와 높은 지터, 쉬머 수치는 발성주기의 폐쇄시기 도중 성문의 불완전한 폐쇄에 의한 기류유출로 유발되는 거친 잡음과 성대의 무주기적인 파동을 유발하는 병변(성대폴립, 성대결절, 종양, 성대마비)들에 의한 것으로 알려져 있다.¹⁰⁾¹¹⁾

실험적으로 음성을 합성하는데 있어서, 지터 수치를 변화시키는 경우 청취자 지각은 거칠음(roughness), 성대잡음 수치를 변화시키는 경우 지각은 기식음(breathiness)을 나타내었다.¹⁰⁾¹²⁾ Childers 등은 정상음성에서 거친 잡음근원의 진폭을 조절함으로써 자연발생적인 기식적 음성을 합성하였다.²⁾¹³⁾

연구자는 Dr. Speech Science 프로그램을 이용하여 세 가지 변수(지터, 쉬머, 성대잡음)를 변화하여 음성을 합성해낼 수 있었다.

성대폴립환자의 경우 수술 전 음성분석에서 측정된 지터, 쉬머와 성대잡음 값을 음성합성프로그램에 입력하여 합성된 수술 전 음성을 분석한 결과 환자의 실제 수술 전 음성과 유사한 음성을 합성해 낼 수 있었다. 그리고 합성된 수술 전 음성의 지터, 쉬머, 성대잡음 값을 정상 범위로 교정한 후, 이 교정된 수치를 다시 합성 프로그램에 입력하여 수술 후 음성을 합성하였다. 합성된 수술 후 음성을 분석한 결과 합성된 음성은 환자의 실제 수술 후 음성과 유사하였다.

그러나 성대마비 환자의 경우 성대폴립환자에서처럼 수술 전 시행한 음성검사에서 측정된 지터, 쉬머, 성대잡음 값을 음성합성 프로그램에 입력하여 합성된 음성을 분석한 결과 환자의 실제 수술 전 음성검사에서 측정된 쉬머, 성대잡음 수치보다 증가된 값을 가지는 전혀 다른 음성이 합성되었음을 알 수 있었다.

Hillenbrand는 음성합성연구에서 합성음성의 세 개의 변수(지터, 쉬머, 성대잡음) 가운데 어느 한 변수에만 변화가 주어져도 나머지 두 변수의 변화가 발생함을 밝

혀내었다. 예를 들면 음성 신호에 성대잡음 값을 증가시키는 경우 신호 대 잡음비의 감소뿐만이 아니라 지터, 쉬머의 증가까지도 발생하였다.¹⁰⁾ 이는 예전의 개념인 지터, 쉬머, 성대잡음이 각각 서로 다른 성대진동 유형의 무주기성의 근원을 반영하는 수치라는 것에 반대되는 개념이다. Hillenbrand 등이 밝힌 바와 같이 측정되는 쉬머 수치는 잡음성분 수준과 주파수 동요의 임의의 변동에 따라 변화가 발생된다. 즉 합성되는 음성에 더해지는 성대잡음의 양이 크면 클수록 이 같은 변동은 피치-펄스(pitch-pulse) 진폭의 변화에 큰 영향을 주어 예측되는 쉬머 값보다 높은 수치를 나타내게 한다.¹⁰⁾ 이 외에도 주파수 동요의 변동 증가에 따라 측정되는 쉬머 수치도 증가됨을 볼 수 있는데, 이는 첫째로 출력과의 피치-펄스의 강도는 성문근원파동의 강도뿐만 아니라 성문근원의 harmonics와 발성관 내의 공명 위치와의 관계에 의해서도 결정이 된다.¹⁰⁾ 둘째로, 더 중요한 원인으로 인접 피치-펄스의 중복이 생각되며, 이로 인하여 일정한 피치-펄스의 후반부의 에너지가 다음 피치-펄스의 전반부의 에너지와 중복되어 진폭의 변화가 발생한다. 이 같은 중복현상은 주파수 동요의 변화가 심한 경우 그 정도가 증가되어 진폭의 주기간 변동, 즉 쉬머 값을 증가 시키게 한다.¹⁰⁾¹¹⁾ 예측된 수치보다 높은 지터 값도 역시 성문근원의 주파수의 주기간 변동의 큰 변화와 큰 강도의 성대잡음의 조합에 기인한다.¹⁰⁾

성대폴립환자의 경우 수술 전 지터, 쉬머, 성대잡음 수치가 정상범위에서 크게 변어나지 않을 정도로 낮기 때문에 변수간의 상호작용이 미약하여 환자의 수술 전 음성검사에서 측정된 세 변수 값을 음성합성 프로그램에 입력하여 음성을 합성할 때 환자의 실제 수술 전과 유사한 음성 즉 예측되는 변수 값을 가지는 음성을 합성할 수 있었다. 이는 김 등의 연구와 부합되는 결과이다.¹⁴⁾ 그리고 세 변수 값을 정상범위로 교정한 후 음성을 합성하였을 때 환자의 실제 수술 후 음성과 유사한 음성을 합성할 수 있었다.

그러나 성대마비환자의 수술 전 음성을 합성하는 경우 수술 전 음성분석에서 측정된 지터, 쉬머와 성대잡음 값을 음성합성 프로그램에 입력 하여 합성된 음성을 분석한 결과 지터, 쉬머와 성대잡음 값이 실제 수술 전 음성의 측정치보다 증가된, 수술 전 음성과 전혀 다른 음성이 합성되었음을 알 수 있었다. 이는 정상범위를 과도로 초과한 세 수치들이 서로 상호작용을 유발하여 증가

된 결과로 판단된다. 이와 같은 상호작용효과를 교정하기 위하여 세 변수 모두를 프로그램에 입력하여 합성된 음성과 두 변수만을 입력하여 합성된 음성과의 차이점을 찾아, 두 변수가 나머지 한 변수에 미치는 영향을 알아보았다. 예를 들어 지터와 쉬머 두 변수만을 입력하여 합성된 음성(성대잡음 값 ; 0)에서 측정된 성대잡음 값은 지터와 쉬머로 인하여 발생한 이차적 성대잡음 수치로 정의할 수 있으며 세 변수(지터, 쉬머, 성대잡음) 모두를 입력하여 합성된 음성의 성대잡음 수치에서 성대잡음을 제외하고 두 가지 변수(지터, 쉬머)만을 입력하여 합성된 음성의 성대잡음 수치를 감한 경우 순수한 성대잡음 수치를 구할 수 있었다. 쉬머 변수의 경우도 위와 같은 방법으로 음성합성시 입력하는 변수를 변화하여 순수한 쉬머 값을 구하였다. 교정된 순수한 세 변수 수치를 다시 음성합성 프로그램에 입력하였을 때, 성대마비환자의 실제 수술 전 음성과 유사한 음성을 합성해낼 수 있었다.

성대마비환자의 수술 후 음성의 합성에서도 합성된 수술 전 음성의 지터, 쉬머, 성대잡음 값을 정상범위로 교정한 후 이를 음성합성 프로그램에 입력하여 음성을 합성하여 분석한 결과 세 변수에서 모두 환자의 실제 수술 후 음성의 측정치보다 감소된 수치를 보였다. 이는 성대마비환자의 경우 성대폴립환자와는 다르게 수술 후에 발생 시 성대 내전의 완전한 복귀로 기류의 유출이 방지되어 성대잡음은 발생하지 않으나 'body and cover theory'의 cover의 변화가 미약하게나마 발생하여 발생시 성대의 비정상적인 파동으로 인하여 지터, 쉬머 수치가 정상범위로 교정되지 않는 결과로 유추할 수 있다.⁷⁾ 다시 음성합성 프로그램에 경도의 지터, 쉬머 값과 정상범위로 교정된 성대잡음 값을 입력하여 합성된 음성은 성대마비환자의 수술 후 음성과 유사하였다.

본 연구에서는 음성합성 프로그램을 이용하여 합성된 음성의 음향학적 분석을 실시하여 환자의 수술 전,후의 음성과 비교한 후 성대마비와 폴립환자의 수술 전,후의 음성과 유사한 음성을 합성할 수 있다는 만족스런 결과를 얻었다. 그러나 궁극적인 음성합성 프로그램의 개발을 위하여는 비언어적인 요소(감정상태, 다중언어, 개인의 발언형태, 나이, 성별)등 표준화된 방법에 대한 연구도 함께 이루어져야 할 것으로 생각된다.¹⁵⁻¹⁷⁾

음성합성 프로그램은 이외에도 통합된 임상과 연구의 도구로서 이상적인 음향 목표를 창조할 수 있게 한다.

수술 후 예측되는 음성의 합성 외에도 환자의 현재의 상태와 정상 상태 사이의 음향학적 특성을 가지는 중간적 음성을 합성하여 임상가가 음성치료를 받는 환자에게 중간의 목표를 제시하여 점차 정상음성의 방향으로 치유될 수 있게끔 한다.

결 론

이 연구는 Dr. Speech Science 음성합성 프로그램을 이용하여 수술 후 예측되는 음성을 합성하는 방법을 고안하기 위한 것이다. 10명의 성대마비환자와 10명의 성대폴립환자를 대상으로 Dr. Speech Science 음성분석 프로그램과 음성합성프로그램(Tiger DRS, Inc. Version 4)을 이용하여 수술 전, 후 음성분석을 시행하였다. 지터, 쉬머, 성대잡음 값을 측정하고 이를 토대로 수술 전, 후의 음성을 합성하여 다음과 같은 결과를 얻었다.

1) 성대폴립환자의 수술 전 음성의 지터, 쉬머, 성대잡음 값을 음성합성프로그램에 입력하여 음성을 합성하고 이를 음성분석한 결과 환자의 실제 수술 전 음성과 합성된 음성은 유사하였다.

2) 성대폴립환자의 합성된 수술 전 음성의 지터, 쉬머, 성대잡음 값을 정상범위로 교정한 후 이를 음성합성프로그램에 입력하여 음성을 합성하여 분석한 결과 환자의 실제 수술 후 음성과 합성된 음성은 유사하였다.

3) 성대마비환자의 수술 전 음성의 지터, 쉬머, 성대잡음 값을 음성합성프로그램에 입력한 후 음성을 합성한 결과 합성음성의 세 변수들 모두 실제 수술 전 음성의 수치보다 증가되었음을 알 수 있다. 이 같은 결과를 유발한 세 변수들 간의 상승효과를 제거한 후 교정된 지터, 쉬머, 성대잡음 값을 음성합성프로그램에 입력하여 합성된 음성은 성대마비환자의 수술 전 음성과 유사하였다.

4) 성대마비환자의 합성된 수술 전 음성의 지터, 쉬머, 성대잡음 값을 정상범위로 교정한 후 이를 음성합성프로그램에 입력하여 음성을 합성한 결과 세 변수에서 모두 환자의 실제 수술 전 음성의 측정치보다 감소된 수치를 보였다. 다시 음성합성 프로그램에 경도의 지터, 쉬머 값과 정상범위로 교정된 성대잡음 값을 입력하여 합성된 음성은 성대마비환자의 수술 후 음성과 유사하였다.

이상의 연구결과로 음성합성 프로그램을 이용하여 합

성한 음성의 음향학적 분석에서 합성음은 수술 전, 후의 음성과 유사하다는 만족스런 결과를 얻을 수 있었다. 이와 같은 방법으로 합성된 '수술 후 예측 음성'을 환자에게 제시함으로써 환자에게 더 많은 정보를 줄 수 있고 나아가서 치료에 대한 환자의 협조를 높일 수 있을 것으로 생각한다.

중심 단어 : 음성합성, 병적음성.

References

- 1) Berke GS, Gerratt BR, Hanson DG : *An acoustic analysis of the effects of surgical therapy on voice quality. Otolaryngol Head Neck Surg.* 1983 ; 91 : 502-508
- 2) Childers DG, Lee CK : *Vocal quality factors : Analysis, synthesis, and perception. J Acoust Soc Am.* 1991 ; 90 (5) : 2394-2410
- 3) Harries ML, Morrison M : *Short-term results of laryngeal framework surgery-thyroplasty type 1 : a pilot study. J Otolaryngol.* 1995 ; 24 (5) : 281-287
- 4) Rosen AR, Lori EL, Murry T : *Acoustic, aerodynamic, and videostroboscopic features of bilateral vocal fold lesions. Ann Otol Rhinol Laryngol.* 2000 ; 109 : 823-828
- 5) Klatt DH : *Analysis, synthesis and perception of voice quality variations among female and male talkers. J Acoust Soc Am.* 1990 ; 87 : 820-857
- 6) Huang DZ, Engr M, Lin S, Engr B : *Speech training user's manual ; Guide to voice synthesis and therapy, 1995 : 1-10*
- 7) Kraus DH, Orlikoff RF, Roiss SS, Rosenberg DB : *Arytenoid adduction as an adjunct to type I thyroplasty for unilateral vocal cord paralysis. Head and neck.* 1999 ; 54 : 52-59
- 8) 조미라. 음성인식 기술의 눈부신 성장. 311호 : 주간동아, 2001 ; 11
- 9) Minifie S, Hixon TJ, Williams F : *Normal aspects of speech, hearing, and language. Prentice-Hall Inc, 1973*
- 10) Hillenbrand J : *A methodological study of perturbation and additive noise in synthetically generated voice signals. J Speech and hearing research.* 1987 ; 30 : 448-461
- 11) Uloza V : *Effects on voice by endolaryngeal microsurgery. Eur Arch Otorhinolaryngol.* 1999 ; 256 : 312-315
- 12) Martin D, Fitch J, Wolfe V : *Pathologic voice type and the acoustic prediction of severity. J Speech Hear Res.* 1995 ; 38 : 765-771

- 13) Lablance GR, Maves MF : *Acoustic characteristics of post-thyoplasty patients. Otolaryngol Head Neck Surg.* 1992 ; 107 (4) : 558-563
- 14) Kim SS, Choi HS : *Acoustic analysis of normal and pathologic voice synthesized with voice synthesis program of Dr.Speech Science. J Korean Logo Phon.* 2001 ; 12 (2) : 115-120
- 15) Furui S : *Toward the ultimate synthesis/recognition system. Proc. Natl Acad Sci Am.* 1995 ; 92 : 10040-10045
- 16) Baek MJ, Hwang MH, Wang SG : *Prediction of postoperative voice by speech synthesis in benign laryngeal disease. Korean J Otolaryngol.* 2002 ; 45 : 279-284
- 17) Alwan AA, Bangayan P, Gerratt BR, Kreiman J, Long C : *Analysis by synthesis of pathological voices using the Klatt synthesizer. In : Kent RD, Ball MJ, editors. Voice quality measurements. 1st ed. San Diego : Singular Pub, 2000 : 307-335*
- 18) Liberman M : *Computer speech synthesis : Its status and prospects. Proc Natl Acad Sci Am.* 1995 ; 92 : 9928-9931
- 19) Hirano M, Hibi S, Yoshida T, Hirade Y, Kasuya H, Kikuchi Y : *Acoustical analysis of pathological voice. Acta Otolaryngol (Stockh).* 1988 ; 105 : 432-438