

Dr. Speech Science의 음성합성프로그램을 이용하여 합성한 정상음성과 병적음성(Pathologic Voice)의 음향학적 분석

연세대학교 의과대학 이비인후과학교실, 음성언어의학 연구소

김 성 수 · 최 홍 식

= Abstract =

Acoustic Analysis of Normal and Pathologic Voice Synthesized with Voice Synthesis Program of Dr. Speech Science

Seong Soo Kim, M.D., Hong-Shik Choi, M.D.

Department of Otorhinolaryngology-Head and Neck Surgery, The Institute of Logopedics & Phoniatrics,
Yonsei University College of Medicine, Seoul, Korea

In this paper, we synthesized vowel /ae/ with voice synthesis program of Dr. Speech Science, and we also synthesized pathologic vowel /ae/ by some parameters such as high frequency gain (HFG), low frequency gain(LFG), pitch flutter(PF) which represents jitter value and flutter of amplitude(FA) which represents shimmer value, and grade ranked as mild, moderate and severe respectively. And then we analysed all pathologic voice by analysis program of Dr. Speech Science. We expect that this synthesized pathologic voices are useful for understanding the parameter such as noise, jitter and shimmer and feedback effect to patient with voice disorder.

KEY WORDS : Voice synthesis · Pathologic voice.

서 론

음성파는 연속적인 성문의 개폐에 의한 주기적인 공기진동과 성도를 지나면서 생긴 잡음성의 공기진동이 있으면 생성된다. 이를 신호변환시스템을 이용하여 다른 형

태의 신호로 만들어 내는 것을 신호처리라고 하며, 음향학적 검사가 신호변환의 대표적인 예라 할 수 있다. 또한 이와는 역으로 이런 음향학적 이론을 배경으로 기계적인 떨림을 음성파와 흡사한 음을 합성할 수 있게 되었고, 이를 바탕으로 현재는 훼손된 음성의 강조와 확장 및 압축에도 이용할 수 있게 되었다.¹⁻³⁾

이렇게 합성된 정상 음성에 인위적으로 여러 변수를 변화시켰을 때 어떠한 음성이 나올 것인지를 살펴보는 것은 매우 흥미로운 뿐만 아니라, 각 변수의 이해를 돕는데도 교육적인 효과가 클 것으로 기대된다. 또한 지금까지 음성을 분석한 연구들은 많았지만 음성합성

논문접수일 : 2001년 9월 25일
심사완료일 : 2001년 10월 17일
책임저자 : 최홍식, 135-720 서울 강남구 도곡동 146-92
연세대학교 의과대학 이비인후과학교실, 음성언어의학 연구소
전화 : (02) 3497-3461 · 전송 : (02) 3463-4750
E-mail : hschoi@yumc.yonsei.ac.kr

Table 1. Voice /ae/ and its pathologic voice synthesis (parameter file)

		(ae) F1=660 F2=1720
높은 잡음 첨가 High frequency gain (HFG)	경도 (mild)	HFG=30(ae1)
	중등도 (moderate)	HFG=45(ae2)
	고도 (severe)	HFG=60(ae3)
낮은 잡음 첨가 Low frequency gain (LFG)	경도	LFG=30(ae4)
	중등도	LFG=45(ae5)
	고도	LFG=60(ae6)
높은 잡음과 낮은 잡음의 혼합 첨가 High, low frequency gain	경도	HFG, LFG=30(ae7)
	중등도	HFG, LFG=45(ae8)
	고도	HFG, LFG=60(ae9)
jitter값 설정 Pitch flutter (PF)	경도	PF=30(ae10)
	중등도	PF=45(ae11)
	고도	PF=60(ae12)
Shimmer값 설정 Flutter of amplitude (FA)	경도	FA=30(ae13)
	중등도	FA=45(ae14)
	고도	FA=60(ae15)
Jitter와 shimmer 혼합	PF+FA	PF=15, FA=15(ae16)
	HFG, LFG, Jitter, shimmer 혼합	HFG+LFG+PF+FA HFG, LFG, PF, FA=15(ae17) HFG+LFG+PF+FA HFG, LFG, PF, FA=30(ae18) HFG+LFG+PF+FA HFG, LFG, PF, FA=45(ae19)

이나 합성한 음성을 분석한 연구자료는 부족한 형편이다.

이에 저자들은 Dr. Speech Science 프로그램을 이용하여 합성한 정상 음성에 여러 변수를 첨가하여 병적 음성을 합성한 후 이에 대한 음향학적 분석을 함으로써, 합성한 병적음성에 대한 객관적인 자료를 제시하고자 하였다.

연구방법

Dr. Speech Science사의 음성합성프로그램을 이용하여 /애/ 모음을 합성한 후 각각의 모음에 임의로 정한 경도, 중등도, 고도의 잡음을 첨가한 후 이에 대한 음성분석을 시도하였다.⁴⁾ 또한 같은 방법으로 각각의 모음에 jitter와 shimmer값을 경도, 중등도, 고도로 분류하여 임의로 정한 값을 주어 음성을 합성한 후 (Table 1) 이에 대한 음성분석을 하였다(Fig. 1). /애/ 모음은 Dr. Speech Science에서 제시한 /애/모음의 참고치에 따라 제 1 포만트는 660, 제 2 포만트는 1720, 제 3 포만트는 2410, 제 4 포만트는 3500, 제 5 포만트는 4400 frequency를 주었고 음성발성시간은 2초로 하여 모음 /애/를 합성하였다(Fig. 2). 음성분석은 Dr. Speech Science사의 음성분석프로그램(Fig. 3)을 이용하였으며 기본 주파수, 신호대 잡음비, jitter, shimmer등의 값을 구하였다.

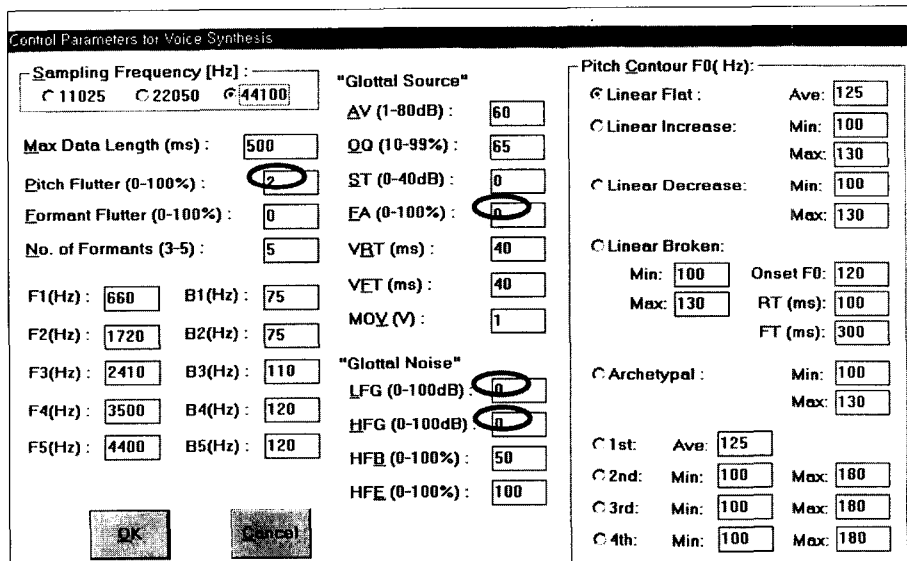


Fig. 1. Pathologic /ae/ vowel synthesis with parameter value changing.

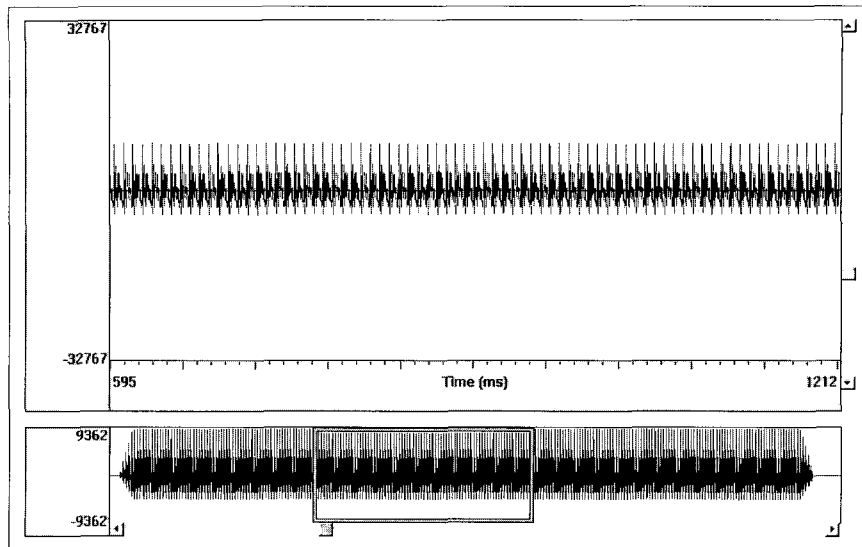


Fig. 2. Synthesized /ae/ voice with voice synthesis program.

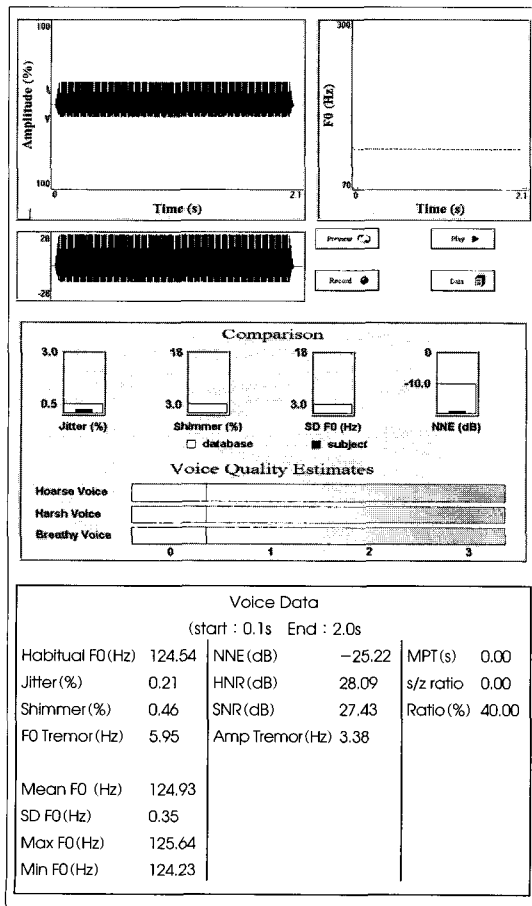


Fig. 3. Analysis of synthesized /ae/ vowel with voice analysis program.

Table 2. Result of voice analysis after the high and low frequency noise was added

	Mean FO	Jitter (%)	Shimmer (%)	S/N ratio
ae0(normal)	124.93	0.21	0.46	-25.22
reference value		0.5이하	3이하	-10이하
ae1 (HFG=30)	125.25	0.23	0.48	-23.90
ae2 (HFG=45)	124.93	0.21	0.59	-23.79
ae3 (HFG=60)	124.74	0.50	8.41	-3.83
ae4 (LFG=30)	124.93	0.23	0.43	-24.59
ae5 (LFG=45)	124.93	0.23	0.56	-23.60
ae6 (LFG=60)	124.93	0.33	6.67	-3.47
ae7 (HFG, LFG=30)	125.15	0.20	0.42	-25.11
ae8 (HFG, LFG=45)	125.43	0.23	0.70	-22.49
ae9 (HFG, LFG=60)	125.38	0.31	5.97	-3.12

결 과

Dr. Speech Science사의 음성합성 프로그램을 이용하여 합성한 /ae/ 모음의 평균 기본주파수는 124.93이었고, jitter값은 0.22, shimmer값은 0.45, 신호대 잡음비는 -24.44였고 이는 정상성인의 /ae/ 목소리와 유사하였다(Fig. 1, 2). 높은 잡음을 첨가하였을 때 경도와 중등도의 잡음에는 jitter나 shimmer, 신호대 잡음비에 큰 변화를 보이지 않았으나 고도잡음을 주었을 때는 jitter값은 변화하여 한계치의 경계선정도의 값을 보였으나 shimmer나 신호대 잡음비는 매우 큰 변화를 보여, 한계

치를 크게 넘어섰다. 낮은 잡음을 첨가하였을 때나 두 개의 잡음을 혼합하여 첨가하였을 때도 비슷한 결과를 보여 다른 음역대의 잡음을 혼합하여도 결과에는 크게 영향을 미치지 못하는 것을 알 수 있었다(Table 2). 한편, jitter값을 점차 높였을 때 음성분석상 jitter값은 그에 따라 점차 커짐을 볼 수 있었고, jitter의 강도가 세어질수록 shimmer값 또한 그에 따라 점차 커짐을 볼 수 있었다. 하지만 shimmer값을 점차 높였을 때에는 shimmer값은 점차 커졌으나, jitter값에는 크게 영향을 미치지 못하였다. jitter와 shimmer값을 혼합하여 주었을 때의 jitter결과는 2.51로 같은 값의 jitter만 주었을 때의 결과인 2.21에 비해 더 나쁜 값을 보임을 알 수 있었고, 또한 shimmer의 결과도 8.76으로 같은 값의 shimmer에만 변화를 준 결과인 7.19보다 더 나쁜 음성지표가 나왔으나 그 차이는 크지 않았다(Table 3). 4개의 변수를 모두 변화를 주었을 때는 정도의 잡음에는 jitter와 shimmer값만이 결과에 영향을 미쳤으나 중등도의 잡음에서는 각각의 변수에 의한 결과보다는 더 나쁜 결과를 보였고 신호대 잡음비는 참고치내에서 변화를 보였다.

고 찰

음성신호처리가 이용되는 분야로는 음성신호 분석분야, 음성의 합성분야, 음성의 합성에 의해 얻어진 음성신호의 분석분야로 구분할 수 있다. 그 중 이비인후과에서는 음성신호 분석을 통해 병적음성에 대한 객관적인 자료를 얻어왔고 임상적으로도 일반적인 음성클리닉이나 음성검사실에서 일상적으로 검사되어지는 항목중의

Table 3. Result of voice analysis after jitter (PF) and shimmer (FA) parameter was changed

	Mean FO	Jitter (%)	Shimmer (%)	NNE ratio
ae0(normal)	124.93	0.21	0.46	-25.22
reference value		0.5이하	3이하	-10이하
ae10(PF=30)	125.15	2.21	4.85	-12.28
ae11(PF=45)	125.43	3.60	8.37	-9.30
ae12(PF=60)	125.38	4.58	13.44	-10.64
ae13(FA=30)	124.93	0.21	7.19	-19.67
ae14(FA=45)	124.93	0.17	8.46	-17.51
ae15(FA=60)	124.91	0.24	12.27	-15.80
ae16(PF, FA=30)	125.03	2.51	8.76	-11.79

하나이다. 하지만 음성의 합성분야나 음성합성에 의해 얻어진 음성신호의 분석분야는 상대적으로 이비인후과나 음성치료분야에서 관심을 받지 못하였는데 그 이유는 아마도 임상적인 효용성을 찾기 힘든 데서 찾을 수 있을 것이다. 본 저자들은 음성의 합성으로 병적 음성을 만들어 낼 수 있을 거라는 기대를 갖고, 각 각의 변수를 주어 새로운 음성을 실험적으로 합성해 보았으며 이를 통해 어떤 변수를 주었을 때 어떠한 음성이 나오는지를 실제로 확인 할 수 있었다. 성문과형 주기의 빠른 변화로 거친 상대음(vocal roughness)과 관련되어지는 음향학적 특성인 vocal jitter와 유성음신호의 빠른 진폭변화를 나타내며, 연속되는 주기들간의 평균진폭차이로 정의되어지는 vocal shimmer¹⁾ 등은 이론적으로는 알고 있으나 실제로 이들 jitter나 shimmer값이 변화하면 어떤 음성이 나오는지, 높은 음의 잡음이 들어가거나 낮은 음의 잡음이 들어갔을 때 어떤 음성이 나오는지들을 들 수 있는 기회는 드물다. 저자들은 본 연구를 통하여 각 변수를 각각 변화시켜 실제로 들어봄으로써 각 변수의 특성을 미루어 짐작할 수 있게 되었다. 더 나아가서는 병적음성의 정도를 달리하여 합성하면 청각적 음성인지검사(perceptual analysis)를 할 때 음성 샘플로도 활용되어 객관적인 척도를 만드는데도 이용될 수 있는 등 교육적인 효용성이 클 것이라고 생각된다.

본 저자들은 Dr. Speech Science의 음성합성 프로그램을 이용하여 /에/모음을 합성하였는데(Fig. 3) 실험적으로 합성한 다른 모음들은 비교적 불안정하여 병적음성 합성 시 분석의 한계를 넘는 경우가 많았다.

저자들은 잡음을 첨가하였을 때 주로 신호대 잡음비의 결과에 크게 영향을 미칠 것으로 예상을 하였으나 Table 2에서 보는 바와 같이 적은 잡음으로는 결과에 큰 영향을 못 미치는 반면 일정한 값이 넘어가면서 크게 음성의 길이 나빠지면서 신호대 잡음비의 결과 뿐만

Table 4. Result of voice analysis after all parameters (noise, jitter, shimmer) were changed

	Mean FO	Jitter (%)	Shimmer (%)	NNE ratio
ae0(normal)	124.93	0.21	0.46	-25.22
reference value		0.5이하	3이하	-10이하
ae17(All=15)	124.95	1.03	4.43	-12.22
ae18(All=30)	125.18	2.67	9.24	-11.88
ae19(All=45)	125.61	3.36	13.81	-10.09

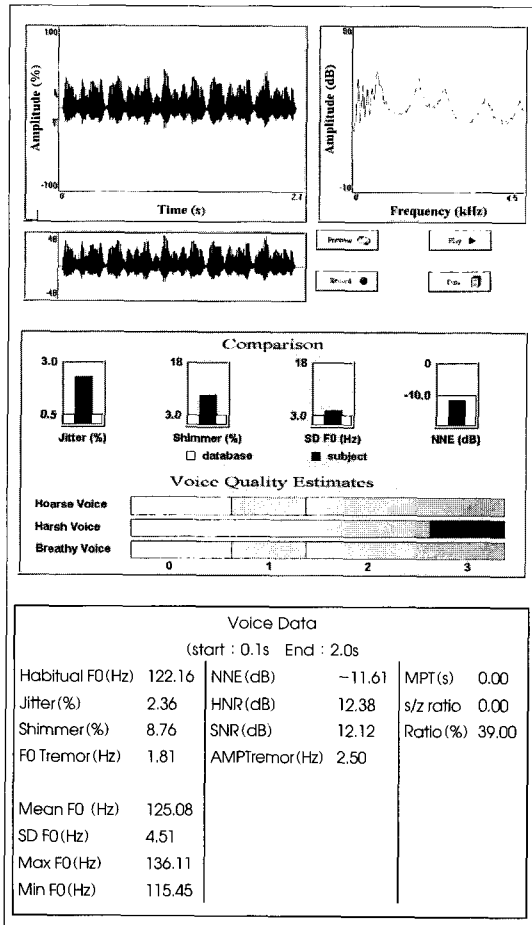


Fig. 4. Analysis of synthesized pathologic /ae/ voice with changing the all parameter (high frequency gain, low frequency gain, pitch flutter means jitter value, amplitude of flutter means shimmer value).

아니라 jitter나 shimmer도 나쁜 결과를 보였으며 비교적 shimmer값이 더 크게 영향을 받음을 알 수 있었다. 또한 낮은 잡음보다는 높은 잡음에서 약간 더 나쁜 결과가 나왔다.

또한 Table 3에서 보는 바와 같이 jitter값과 shimmer값의 변화를 주면 jitter값의 변화는 전체 결과에 영향을 미치는 반면 shimmer값의 변화는 다른 결과에는 큰 영향을 미치지 못함을 알 수 있었다. jitter값과 shimmer값에 동시에 변화를 주면 각각의 단순히 하나의 값에 변화를 준 결과보다는 더 나쁜 결과를 나타내는 것으로 보아서 두 변수가 상호 작용함을 알 수 있었다.

Table 4는 잡음과 jitter, shimmer모든 값을 변화시킨 결과로 이 또한 하나의 값에 변화를 준 결과보다는 더 나쁜 결과를 보여 주었고(Fig. 4), 스펙트로그램상에서도 정상(Fig. 5)에 비해 매우 불규칙한 모습을 보였다(Fig. 6).

이상의 결과를 볼 때 각각의 변수는 서로에게 영향을 미치는 것을 볼 수 있었으나 어떤 값이 어떻게 영향을 주는 지는 보다 많은 데이터를 분석해 보아야 할 것으로 사료된다.

결론

저자들은 Dr. Speech Science의 음성합성 프로그램을 이용하여 정상음성과 병적음성을 합성할 수 있었으며, 이에 대한 음향학적 분석을 함으로써 주어진 변수에 대한 변화를 체계적으로 살펴볼 수 있었다. 이렇게 하여

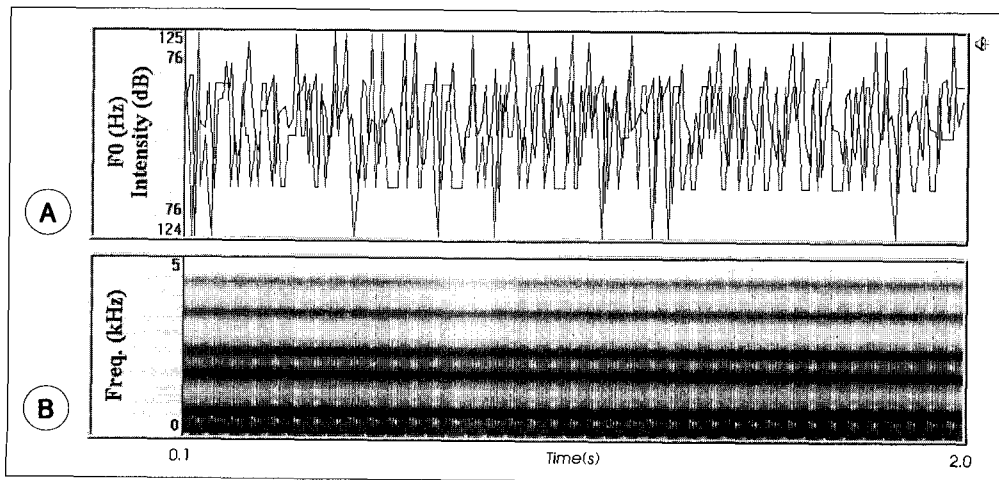


Fig. 5. Analysis of synthesized /ae/. A : fundamental frequency and intensity, B : spectrogram.

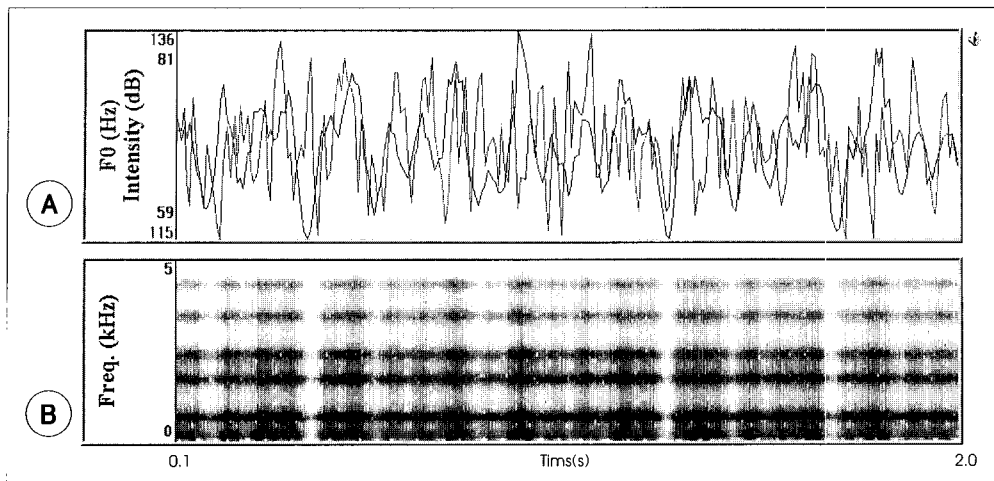


Fig. 6. Analysis of synthesized pathologic /ae/voice changed all parameter such as high frequency gain, low frequency gain, jitter, shimmer. A : fundamental frequency and intensity, B : spectrogram.

잡음이나 jitter, shimmer 등의 변수가 변화된 소리를 직접 듣게 되어, 이들 변수의 변화로 인해 음성이 어떻게 변화하는지 알 수 있었으며 이를 통해 이들 변수들에 대한 이해를 높일 수 있었다. 또한 음성질환을 가진 환자에게 되먹임작용을 통해 치료적인 목적으로 사용할 수 있을 뿐만 아니라 청각적 인지검사인 GRBAS 척도를 구할 때 표준화된 음성샘플로도 제시할 수 있을 것으로 기대하며 본 연구에서 객관적인 자료를 제시하고자 하였다.

중심 단어 : 음성합성 · 병적음성.

References

1) Hong SK : *Sound spectrography analysis. J Korean Soc*

Logo Phon. 1997 ; 8 : 128-137

2) Dunn HK : *The calculation of vowel resonances and an electric vocal tract, JASA. 1950 ; 22 : 740-753*

3) Kent RD, Forner LL : *Developmental study of vowel formant frequencies in an imitation test, readings in clinical spectrography of speech, singular publishing group. and KAY electronics corp, 1978 ; 392-401*

4) Huang DZ, Engr M, Lin S, Engr B : *Speech training user's manual ; Guide to voice synthesis and therapy, 1995 ; 1-10*