ARS와 신경회로망을 이용한 장애음성의 수집, 분석 및 식별에 관한 연구

김광인*, 조철우*, 김대현*, 왕수건**, 전계록***, 안시훈***, 김기련***, 김용주***
* 창원대학교 제어계측공학과, ** 부산대학교 이비인후과, *** 부산대학교 의공학과

Collection, Analysis and Classification of Pathological Voice from ARS using Neural Network

Kwang-In Kim*, Cheol-Woo Jo*, Dae-Hyun Kim*, Soo-Geon Wang**,
Kye-Rock Jun***, Si-Hun Ahn***, Ki-Reyn Kim***, Young-Ju Kim***
* Dep. of Control & Instrumentation Engineering, ** Dep. of Medicine Pusan National University,
*** Dep. of Medical Engineering Pusan National University
E-mail : karisman@hanmail.net

요약문

본 논문은 음성신호를 이용해 성대의 질환이 있는 환자를 진단하고 병명을 판별하게끔 유도하는 자동 진단 시스템을 개발하기 위한 연구의 일부로, 그중 ARS를 이용하여 환자의 음성을 수집, 분석, 식별의 실험에 대한 연구이다.

본 연구 팀에서는 이미 CSL을 이용한 장애음성 데이터의 수집과 식별에 관한 연구 결과를 발표한바 있다. 하지만 선행연구에서는 방음실에서 디지털 녹음기를 이용하여 수집한 음성을 사용했기 때문에, ARS를 통하여 녹음한 음성과는 샘플링 주파수나 대역폭, 잡음성분등의 데이터의 특성이 상당한 차이가 있다. 이러한 이유로 ARS를 통하여 녹음한 음성에 보다 적합한 파라미터 분석프로그램을 작성하여 파라미터를 구하였다. 이 파라미터들은 Kay사의 MDVP를 기초로하여 작성하였고, 대부분 80%정도의 신뢰성을 가졌다.

수집한 음성의 식별은 정상음성과 양성음성의 두가지 경우로 분리하였다. 식별기법으로는 신경망을 이용하였고, 식별파라미터는 구한 파라미터중 6개의 파라미터를 선별하여 식별한 결과 약 90%정도의 식별율을 가졌다.

1. 서론

산업사회가 가속화됨에 따라 대기오염, 수질오염등의 환경오염은 필연적으로 많은 환자를 낳게 되었다. 그중 그중 인간의 음성에 영향을 끼치는 후두질환자는 과거에 비해 상당히 증가하였다. 이런 후두질환은 초기에 발견하게 되면 간단한 수술로 회복할 수 있으나, 그렇지 못하면 자신의 음성을 잃는 것은 물론 생명까지 위협받을 수가 있다. 따라서 후두질환의 조기치료는 생명뿐만 아니라 본래의 음성을 보존하는데 상당히 중요한 역할을 한다. 이러한 조기치료를 위해 본 연구는 ARS를 이용한 원격진단 시스템을 목표로 한다. 이렇게 ARS 음성을 이용한 식별 시스템은 검사가 신속, 간단한데다가 환자에게는 고통없이 검사가 가능하다. 1)

본 논문은 이러한 시스템을 위해 후두질환음성의 파라미터분석과 식별을 하고자 한다. 선행된 연구에서 CSL을 이용한 장애음성의 분석과 식별의 결과는 이미 나와있다.2) 허나 선행연구에서 사용된 음성의 데이터와 수집한 ARS 음성 데이터 사이에는 상당한 차이가 있다. 샘플링데이터를 비롯하여 대역폭이라든지, 잡음성분의 포함정도는 이전에 CSL로 분석한 음성데이터와는 비교할 수 없을 정도로 다르다. 이에 기존의 MDVP에서 사용되던 파라미터를 윈도우상에서 ARS데이터를 분석할 수 있도록 작성된 프로그램3)으로 파라미터를 분석 전체 또는 몇 개를 선별하여 신경회로망을 이용하여 정상과 양성의 구분을 한다.

2장에서는 특징 파라미터에 대해, 3장에서 데이터의 수집 4,5장에서는 신경회로망을 이용한 식별과 그결과에 대해서 이야기 한다.

2. 장애음성의 특징 파라미터

장애음성의 특성을 나타내어 주는 파라미터들로는 여러 가지가 있지만 본 논문에서는 MDVP에서 사용된 33가지의 파라미터중 12가지를 사용하려 한다. 나머지 파라미터들은 장애음성의 특성과는 무관한 특성을 나타내는 파라미터 또는 특정음성에서는 구해지지 않는 성질이 있는 파라미터이므로 장애음성의 특성을 나타내어 주는 파라미터들로 보기는 어렵기 때문에 제외하였다. 그리고 선행된 연구에서 좋은 결과를 내었던 캡스트럼 방식의 음원분석 파라미터인 HNRR(Harmonic-to-Noise Ratio : Residual)을 추가로 식별에 사용하려고 한다.

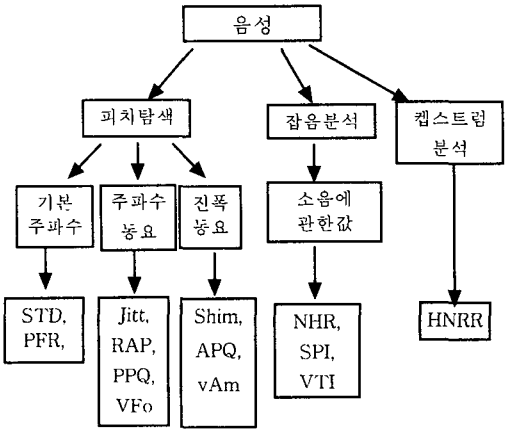

그림 1. 특징파라미터

12가지의 파라미터들로는 우선 크게 네가지로 기본주파수, 주파수 동요, 진폭 동요, 소음관련으로 구분할 수 있다. 기본주파수와 관계된 파라미터들로는 기본주파수들의 표준편차를 나타내는 *STD*와 기본주파수의 범위를 나타내는 *PFR*이 있다.

주파수 동요에 관련된 파라미터로써 우선 *Jitt*는 피치주기의 변화율을 나타내는데 사용되는 파라미터로 연속적인 피치주기사이의 평균변화율을 나타낸다.

$$Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|To^{(i)} - To^{(i+1)}|}{\frac{1}{N}\sum_{i=1}^{N}To^{(i)}} \quad (1)$$

식 1에서 $To^{(i)}$는 i번째 피치의 주기이고, N은 측정된 피치의 개수이다. 그리고 이 *Jitt* 값으로부터 *Jitt*와 비슷하지만 3개의 스무딩 계수를 가지는 *RAP*, 5개인 스무딩 계수를 가지는 *PPQ*를 구할 수 있다. 그리고 *Jitt*의 표준편차를 나타낸 값인 *vFo*가 있다.

진폭의 동요에 관련된 파라미터로써 우선 *Shim*는 진폭의 변화율을 나타내는데 사용하는 파라미터로 연속적인 진폭변화율의 평균값을 나타낸다.

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A^{(i)} - A^{(i+1)}|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \quad (2)$$

식 4에서 $A^{(i)}$는 i번째 피치주기에서의 진폭의 값이고, N은 측정된 진폭의 개수이다. 그리고 이 *Shim*값으로부터 *Shim*과 비슷하지만 11개의 스무딩 계수를 가지는 *APQ*, *Shim*값의 표준편차를 나타내는 *vAm*를 구할 수 있다.

소음 관련 파라미터로는 우선 노이즈와 하모닉 에너지의 비를 나타내는 *NHR*로 1500-4500Hz의 하모닉 에너지의 값과 70-4500Hz의 인하모닉 에너지의 비이다. *VTI*는 음성의 난조를 나타내는 파라미터로, *NHR*과 비슷하지만 그 범위가 달리 2800-5800Hz의 하모닉 에너지와 70-4500Hz의 인하모닉 에너지의 비이다. *SPI*는 연발성(輦發聲)을 나타내는 파라미터로 저주파(70-1600Hz)에서의 하모닉 에너지와 고주파(1600-4500Hz)에서의 하모닉 에너지의 비이다.3)

마지막으로 *HNRR*은 음성에서 선형예측분석으로부터 구해진 예측오차신호로부터 켑스트럼을 구한 뒤 하모닉 성분과 잡음성분을 분리하여 그 비율를 나타낸 파라미터이다.1)2) 선행실험에서 *HNRR*은 *Jitt*r나 *Shim*r에 비해 높은 변별력을 갖는 것이 확인되었다.4)

3. 음성의 수집

음성의 수집은 다음과 같은 과정을 거친다.

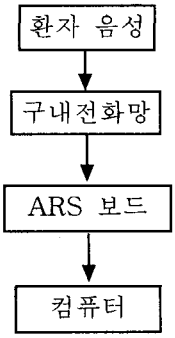

그림 2. 음성수집 과정

환자의 음성은 전화기를 통해 구내전산망을 타고서 컴퓨터실의 ARS보드와 연결된 컴퓨터로 들어오게 된다. 이때 음성은 샘플링 주파수 11025Hz에 샘플링 레이트 8bit의 모노로 저장된다. 이렇게 저장된 음성은 각 병명별로 정상(normal), 양성후두질환에서 낭종(Intra cordal cyst), 라인케 부종(Reinke's edema), 후두염(Laryngitis), 성대결절(Vocal nodule), 성대마비(Vocal

cord palsy), 용종(Vocal polyp), 기타, 그리고 악성후두질환에서 성대암 1기 (T1), 2기(T2), 3기(T3) 의 모두 3종류의 11개의 병명으로 분류된다.

4. 장애음성의 식별방법

패턴인식방법으로는 대량의 복잡한 데이터를 병렬처리하는데 유용한 방법중의 하나인 신경회로망을 사용한다. 신경회로망은 학습규칙에 따라 Hopfield network, Hamming network, Boltzmann machine, error propagation model 등 여러 가지의 방법이 있다. 그 중 perceptron은 feed-forward 연결구조를 가지며 패턴 식별의 기능을 갖는 간단한 형태의 신경회로망이다. 초기의 단층 구조의 perceptron은 그 구조의 간결성으로 이론을 증명하거나 학습규칙을 만들어 내는데 효과적이어서 그후 Rumelhart등에 의하여 back-propagation의 학습 알고리즘이 제안됨으로써 다층구조의 perceptron으로 확장되게 되어 많은 문제에 응용될 수 있게 되었다.

Back-propagation의 특징으로는 Delta Rule과 비슷하지만 multi layer를 사용한다는 점에서 차이가 난다. 그렇기 때문에 Multilayer Networks로 학습하고 Chain Rule을 사용하여 미분계산을 좀더 개선할 수 있다.

Back-propagation은 기대되는 출력층의 node들의 값과 한층 낮은 층의 node들의 값과의 차를 LMS를 적용시켜 값이 감소하도록 가중치를 조절하는 식으로 반복한다.5)

본 논문에서는 3개, 6개, 15개의 입력을 가지고 2개의 layer를 갖고 2개의 출력을 갖도록 구성되어 있다. 각 layer의 전달함수는 1번째 layer에서는 Hyperbolic Tangent Sigmoid를 사용하였고, 2번째 layer에서는 Linear를 사용하여 학습하였다.

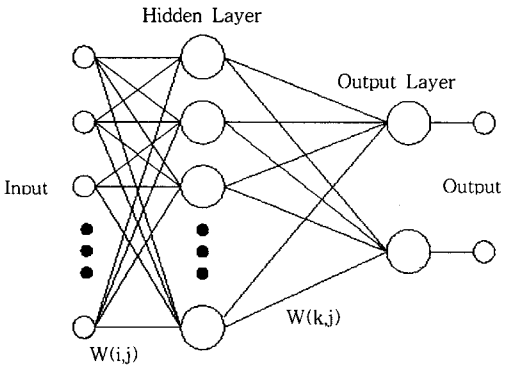

그림 3. 2 layer 신경회로망

5. 실험 및 검토

본 논문은 ARS를 통해 수집한 한국인 장애음성데이터를 대상으로 정상음성과 장애음성을 식별하는 실험을 수행하였다. 이 데이터베이스에서 정상음성 38개, 장애음성 119개를 이용하여, 전체 157개의 데이터에 대해 신경망 훈련 및 식별 실험을 하였다. 실험에 사용한 각 음성은 모음 /아/를 일정한 시간동안 발음한 것으로 샘플링 주파수는 11025Hz이며, 샘플링 레이트는 8bit로 되어있다.

실험에 쓰인 데이터는 정상인 데이터중 무작위로 2/3, 비정상인 데이터중 무작위로 2/3를 선출한 데이터로 학습에 사용했고, 나머지 1/3을 식별에 사용했다. 그리고 실험은 파라미터의 개수를 달리하여 세 번에 걸쳐 실행하였다.

첫 번째는 과거 CSL로 녹음하여 식별한 결과와 비교하기 위해 Jitter와 Shimmer 그리고 HNRR의 3개의 파라미터만으로 식별하였다.

두 번째는 15개의 모든 파라미터를 사용해 식별하였다. 이 방법은 보다 많은 정보를 적용할 수 있다는 장점을 가진다.

표1. 신경회로망 결과

		3개의 파라미터		6개의 파라미터		15개의 파라미터		CSL을 분석한 3개의 파라미터	
		정상	비정상	정상	비정상	정상	비정상	정상	비정상
Training Data	정상	22	4	21	5	23	3	21	1
	비정상	4	77	4	77	8	73	4	18
	식별율(에러/전체)	92.52%(8/107)		91.59%(9/107)		89.72%(11/107)		88.64%(5/44)	
Test Data	정상	11	1	11	1	8	4	10	1
	비정상	6	33	5	34	6	33	1	10
	식별율(에러/전체)	86.27%(7/55)		88.24%(6/51)		80.39%(10/51)		90.9%(2/22)	

세 번째는 6개의 파라미터를 선별하여서 식별하였다. 이 경우는 모든 파라미터의 군집도를 통해서 식별시 도움이 된다고 판단한 파라미터만을 선별한 다음 식별하였다.

표1의 정상음성과 장애음성간의 식별에서 세 가지 방법을 모두 사용해서 나온 결과이다. 결과에서 학습에 사용된 데이터인 Training Data는 전부 90% 정도로 충분히 학습되었다고 여겨진다. 그렇게 학습된 데이터로 Test Data를 식별한 결과를 보면 대부분 80%를 넘는 식별율을 가지게 된다. 여기서 Test Data의 결과를 보면 가장 많은 15개의 파라미터를 사용해서 식별한 결과가 가장 작은 3개의 파라미터를 가지고 식별한 결과보다 못함을 알 수 있다. 원인은 모든 파라미터들이 장애음성의 식별에 도움을 주지 않는 다는 점이다. 오히려 역효과를 가지고 오는 파라미터도 포함되어 있기 때문에 효과있는 3개의 파라미터로 식별한 결과보다 15개의 파라미터로 식별한 결과가 더 못한 결과를 가진다. 그리고 6개의 데이터로 식별한 결과가 3개의 파라미터로 식별한 결과보다 약간 더 좋은 것은 마찬가지로 식별에 충분히 파라미터가 3개가 실제로 식별에 도움을 주었기 때문이다. 그리고 선행연구에서 CSL을 이용하여 수집한 음성의 식별결과와 비교해 보면 Training Data에 대해서는 ARS쪽이 더 나은 식별을 해내었으나 Test Data에 대해서는 ARS쪽이 약간 낮은 식별율을 보이고 있으나 거의 비슷한 식별수준을 보이고 있음을 알 수 있다. 이것은 CSL로 녹음한 음성데이터가 ARS를 이용하여 녹음한 음성데이터보다 잡음성분이 적고 대역폭이 크기 때문에 파라미터의 결과가 더 양호하기 때문이다.

6. 결론

본 논문에서는 정상음성과 장애음성을 구분하기 위해 3, 6, 15개의 입력에 3-layer, 2출력을 가지는 신경회로망을 이용하였으며 학습을 위해서는 back-propagation을 사용하였다. 실험결과 6개의 파라미터를 이용한 결과가 가장 높게 나왔고 이 결과는 88%정도로 충분히 정상음성과 질환음성을 이 방법에 의해 구분이 가능함을 알 수 있었다.

차후의 연구에서는 기존의 데이터베이스를 확장하여 더 많은 장애음성을 수집하여 일반성을 높이고 ARS로부터 오는 신호의 음향적 특성을 고려한 파라미터를 도출할 필요가 있다. 또한 단순히 정상음성과 질환음성만을 구분하는 것이 아니라 보다 세부적인 분류방법을 개발하여 유용성을 높이고자 한다.

참 고 문 헌

1) B.Yegnanarayana, C.d'Alessandro, V.Darsinos, "An Iterative Algorithm for Decomposition of Speech Signals into Periodic and Aperiodic Components",IEEE trans. on Speech and Audio Processing, Vol.6, No.1, Jan. 1998

2) 조철우, 김대현, "Cepstrum방법과 신경회로망을 이용한 정상, 양성종양, 악성종양 상태의 식별에 관한 연구", 한국음향학회 추계학술발표대회 논문집 제18권, pp.399-402, 1998

3) 조철우, 김광인, "다양한 수집방법에 의한 장애음성 분석도구의 구현에 관하여", 제17회 음성통신 및 신호처리 학술대회, pp211-214, 2000

4) 김대현, 조철우, "장애음성의 분류방법에 관한 연구" 제15회 음성통신 및 신호처리 워크샵, pp388-391,1998

5) R.P.Lippman, "An Introduction to computing with Neural Nets", IEEE ASSP Magazine, Vol.4. No 2, pp4-20, April, 1987

6) Operations Manual, 'Multi-dimensional Voice Program(MDVP)' ,Model 4305, Kay Elemetrics Corp, 1993

7) Operations Manual, 'Disordered Voice Database', Model 4337, version 1.03, Kay Elemetrics Corp, 1994