http://dx.doi.org/10.14400/JDC.2016.14.11.249

고차 미분에너지 기반 노인 음성에서의 음성 구간 검출 알고리즘 연구

이지연
중원대학교 의료공학과

Development of Voice Activity Detection Algorithm for Elderly Voice based on the Higher Order Differential Energy Operator

JiYeoun Lee
Department of Biomedical Engineering, Jungwon University

요 약 노인 음성은 연령에 따른 호흡, 발성, 공명 등의 생리적 변화에 의하여 다량의 잡음이 발생된다. 따라서 노인 음성으로 음성인식 및 합성, 분석 소프트웨어등과 같은 융복합 헬스케어 기기를 동작시키고자 할 때, 성능을 저하시키는 결과를 야기한다. 그러므로 노인 음성을 분석하여 그들의 목소리로 다양한 헬스케어 기기를 잘 운영할 수 있는 위한 연구 개발이 필요하다. 따라서 본 연구는 노인 음성 잡음을 고려하여 기존의 대칭 구조 고차 미분 에너지 함수를 이용하여 노인 음성에서의 음성 구간 검출 알고리즘을 연구하였으며, 자기상관함수와 AMDF 방법과 비교하여 노인 음성에서의 음성 구간 검출에 보다 우수한 성능을 가지는 것을 확인하였다. 본 논문에서 제시하는 음성 구간 검출 알고리즘은 노인을 위한 음성 인터페이스에 적용함으로써 노인들의 스마트 기기에의 접근성을 높이고, 더 나아가 노인들을 위한 융복합 웨어러블 디바이스 성능 개선 및 다양한 개발이 가능할 것으로 전망한다.

주제어 : 노인 음성, 자기상관함수, AMDF, 고차 미분 에너지 함수, 음성 구간 검출

Abstract Since the elderly voices include a lot of noise caused by physiological changes in respiration, phonation, and resonance, the performance of the convergence health-care equipments such as speech recognition, synthesis, analysis program done by elderly voice is deteriorated. Therefore it is necessary to develop researches to operate health-care instruments with elderly voices. In this study, a voice activity detection using a symmetric higher-order differential energy function (SHODEO) was developed and was compared with auto-correlation function(ACF) and the average magnitude difference function(AMDF). It was confirmed to have a better performance than other methods in the voice interval detection. The voice activity detection will be applied to a voice interface for the elderly to improve the accessibility of the smart devices.

Key Words : Elderly voice, Auto-corration function(ACF), Average magnitude difference function(AMDF), The symmetric higher order differential energy operator(SHODEO), Voice Activity Detection

* 이 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 No. 2014-00540001)
Received 28 September 2016, Revised 31 October 2016
Accepted 20 November 2016, Published 28 November 2016
Corresponding Author : JiYeoun Lee(Jungwon university)
Email: jylee@jwu.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

음성 연구 분야는 다양한 정보를 효과적으로 처리하고자 하는 현대의 지식 정보화 사회 추세에 따라 인간과 기계의 소통을 원활하게하기 위한 차세대 인터페이스 기술로 각광받고 있다[1]. 또한 음성 신호는 화자의 의사뿐만 아니라 심적 상태, 건강 상태, 언어적 능력 또한 기계에 전달할 수 있다는 특징을 가지고 있다. 따라서 이를 스마트 기기에 적용하여 의료 산업의 웨어러블 디바이스(Wearable device)를 개발하고자 하는 연구가 세계적으로 확장되고 있다[1,2,3].

현재 한국의 65세 이상 노인인구는 2000년에 7.2%를 넘어 고령화 사회에 도달하였으며, 2010년 10%를 넘어서 2018년 노인인구가 14% 이상으로 추산된다[4]. 고령사회임에도 불구하고, <Table 1>에서 볼 수 있듯이, 현재 노인들의 스마트 기기 이용률은 열악하기 때문에, 한국 노인들 대부분이 웨어러블 디바이스를 비롯한 스마트 기기의 의료 서비스 혜택을 거의 받지 못하고 있는 실정이다[5].

〈Table 1〉 Elderly using smart devices (%)

Age	Utilization	Non-utilization	Total
The age of 55-64	3.9	96.1	100
Over the age of 65	0.7	99.3	100

노인들의 스마트 기기에 대한 높은 진입 장벽의 주된 원인으로 불편한 인터페이스가 꼽힌다. 즉, 스마트 기기에서 제공하는 음성 인터페이스는 청년/장년층의 평균적인 발성을 기준으로 기기를 개발하였기 때문에 노인 음성의 발성으로 음성 인터페이스가 제대로 동작하지 않는다[5].

노년기에 접어들게 되면 기능이 감소된 성대의 움직임, 얇아지고 각질화 된 성대의 상피점막 등의 원인에 의해 성대의 공명 속성이 변화되어 말속도가 느려지고 묵음의 횟수와 길이가 늘어나게 된다[5]. 따라서 노인 음성은 정상 음성과 구별되는 장애 음성의 하나의 종류로 생각할 수 있다. 위의 변화들은 음성의 부정확성과 잡음 등을 야기함으로써 음성 인터페이스 기반 융복합 기기들의 성능을 감소시킨다[5,6,7]. 따라서 노인 음성 데이터베이스 구축을 통한 알고리즘의 개선 필요성이 증대되고 있다.

음성 구간 검출(Voice activity detection)을 위한 방법 중의 하나인 Pitch 검출 알고리즘은, 시간영역, 주파수영역, 켑스트럼 영역에서 다양하게 구할 수 있다[8]. 일반적으로 자기상관함수(Auto-correlation function, ACF), AMDF (Average magnitude difference function), ZCR(Zero crossing rate) 등의 방법을 이용한다[9]. 이와 같은 방법들은 음성이 시불변(Time-invariant)이라는 가정 하에 이루어지며, 현재까지 개발된 음성의 시변적 특성을 반영한 알고리즘 중에서 대칭구조를 갖는 고차의 미분에너지 함수(The symmetric higher order differential energy operator, SHODEO)를 이용한 주파수 추정기가 다른 방법과 비교하여 우수한 주파수 추정 성능을 보인다고 알려진다[10, 11]. 대칭 구조 고차의 미분 에너지 함수를 이용한 기본 주파수 추정기는 적은 계산량을 요구함에도 음성의 순간 주파수를 실시간으로 분석하여 유성음과 무성음 구간의 판별에 우수한 성능을 보인다[8].

이러한 사실에 근거하여 본 논문은 기존의 대칭 구조의 고차 미분 에너지 함수(SHODEO)를 응용하여 노인 음성의 특징을 고려한 음성 구간 검출 알고리즘을 개발함으로, 노인 음성 분석에 적합한 융복합 소프트웨어 및 기기를 개발하기 위한 기초 자료로 사용하고자 한다[12].

2. 음성 구간 검출 알고리즘

2.1. 자기상관함수

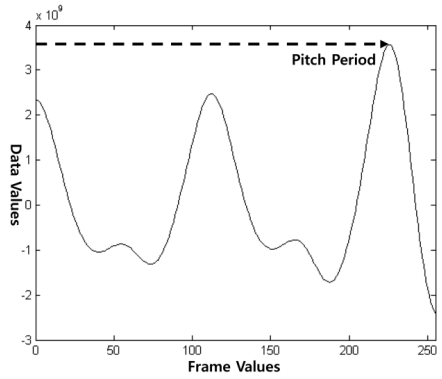
자기상관함수는 특정 신호의 한 시간과 다른 시간에서의 상관 관계를 통해 음성 신호의 Pitch를 추출하는 알고리즘으로, 수식(1)와 같이 정의한다[12].

$$D_{ACF}(m) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)x(n+m) \quad (1)$$

수식(1)에서 N는 데이터 길이, x(n)는 특정 지점 n에서의 데이터 값, x(n+m)은 n에서 m까지의 값이다. Pitch를 구하기 위한 Frame 길이는 256, Overlapping은 128로 설정하였다[14].

256 Frame마다 음성 구간의 자기상관함수를 분석하였을 때 [Fig. 1]와 같이 256 Frame에서 Maximum peak

를 가지는 파형이 나타나며, 이 Peak가 나타나는 지점을 Pitch 주기라 판정한다[15]. Pitch 주기를 결정하기 위한 후처리 작업으로는 3의 길이를 가지는 중간 값 필터(Median filter)를 사용하였다. 잡음이 다량 산재되어 있는 데이터베이스를 사용하였으므로 250Hz 이상의 주파수 대역을 제한하였다.



[Fig. 1] Pitch extraction method by Auto-correlation function

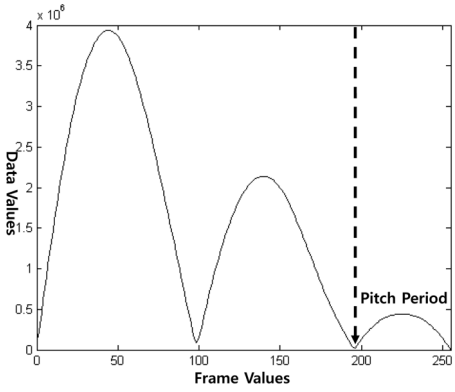
2.2 AMDF

AMDF는 자기상관함수와 마찬가지로 시간 영역에서 음성 신호의 Pitch를 검출하는 알고리즘으로, 다음 수식(2)과 같이 정의한다[13,16].

$$D_{AMDF}=\frac{1}{N}\sum_{n=0}^{N-1}|x(n)-x(n+m)| \tag{2}$$

식 (2)에서 신호 $x(n)$는 원 음성 신호와 임의의 길이 N의 창함수(Windowing function) $w(n)$와의 연산 결과로 AMDF의 입력 신호로 사용되었다[16]. Pitch를 구하기 위한 Frame 길이는 자기상관함수와 마찬가지로 256, Overlapping은 128 로 설정하였다.

AMDF의 경우 [Fig. 2]와 같이 음성 구간의 256 Frame범위 내에서 나타나는 파형의 Minimum Peak지점을 Pitch주기라 판정한다[13,16]. 후처리 작업으로는 3의 길이를 가지는 중간 값 필터를 사용하였으며 250Hz 이상의 주파수 대역을 제한하였다.



[Fig. 2] Pitch extraction method by AMDF

2.3 고차 미분에너지(SHODEO) 함수를 이용한 기본 주파수 추정기

순간주파수는 시간의 함수인 신호의 위상에 미분을 취한 것으로 정의한다[9]. 또한 연속신호의 k차 미분에너지 함수는 다음 수식(3)과 같이 나타낸다[10].

$$\begin{aligned}\Gamma_k\{x(n)\}=&x(n)x(n+k-2)\\&-x(n-1)x(n+k-1)\end{aligned} \tag{3}$$

k는 임의의 차수, n는 신호의 Sampling된 범위, x(n)는 이산 변수 n에 따른 데이터 값을 나타낸다. 고차 미분에너지 함수는 다음과 같이 임의의 차수 k에 따라 두 가지 수식으로 표현된다[10].

$$\Xi_k\{x(n)\}=\begin{cases}\dfrac{\Gamma_k\{x(n)\}+\Gamma_k\{x(n-k+2)\}}{2} & k=odd\\ \Gamma_k\left\{x(n-\dfrac{k}{2}+1)\right\} & k=even\end{cases} \tag{4}$$

위 수식 (4)와 수식 (5)을 통해 순간 주파수를 산출한다[10].

$$f(n)=\frac{1}{2\pi}\frac{1}{(k-1)}cos^{-1}\left(\frac{\Xi_{2k-1}\{x(n)\}}{2\cdot\Xi_k\{x(n)\}}\right) \tag{5}$$

k는 임의의 차수, x(n)는 현재 시점 n에서의 음성 데이터 값 , $\cos^{-1}\theta$의 $\Xi_{2k-1}\{x(n)\}/(2\cdot\Xi_k\{x(n)\})$는 차수 k에 따른 고차 미분에너지 함수의 비율이라 정의한다. 순간주파수는 무성음 구간에서는 주파수 패턴이 불규칙하게 나타나나 500Hz 이내의 유성음 구간에서는 주파수 패턴이 일정한 특성을 보인다[9].

3. 데이터베이스

본 논문에서는 The Speech Information Technology & Industry Promotion Center(SiTEC)에서 배포한 노인 음성 데이터베이스에서 발췌한 70대 남녀 각 10명의 음성을 사용하였다. <Table 2>에서와 같이 단어 5개와 문장 2개가 실험 데이터로 사용되었다. 남녀 각각 한번씩 발화한 단어 5개, 문장 2개가 사용되어, 총 20문장과 20단어가 실험데이터로 사용되었다. 또한 그 데이터를 16Hz 샘플링하였다.

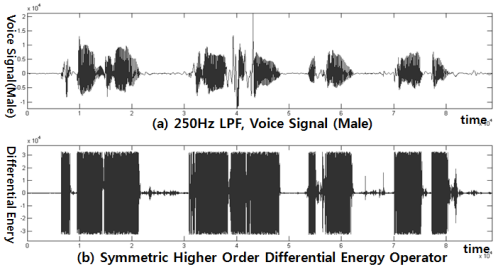
〈Table 2〉 Database information

	Male	Female
Word	Cheongwadae	
	Bogeolsungeo	
	Ccleoango	
	Uducceni	
	Boheomryo	
Sentence	Then somebody came forward to her desk.	
	Then a stranger approached and asked.	

4. 실험 결과

4.1 고차 미분에너지 함수이용 음성구간 판별

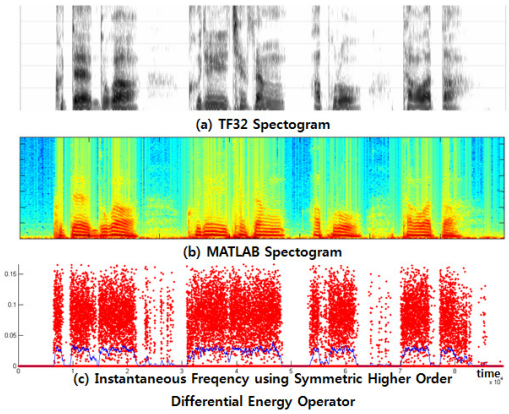
미분에너지 함수는 다음 [Fig. 3]와 같이 음성 구간에서 진폭이 크게 나타나는 특성이 [Fig. 3](b)에서 보인다. 따라서 임의의 역치 값 800을 지정하여 음성 구간을 판정한다. 저역통과필터는 250Hz, 기본 주파수를 추산하기 위한 순간 주파수의 차수는 k=2로 지정하며, 유성음과 무성음을 구분하기 위한 함수로는 차수 k=2에서 얻어낸 3차 에너지 함수를 이용한다. 마지막으로 순간주파수 값을 데이터 길이 200의 이동평균 필터로 처리하여 기본 주파수의 추정 값을 산출한다[9].



[Fig. 3] (a) Voice signal(a) and (b) SHODEO function

이동평균필터 (Moving average filter)의 길이는 200으로 설정하였으며, 기존의 대칭구조 미분 에너지 함수를 이용한 기본 주파수 추정은 필터의 입력 값으로 유성음으로 판정된 순간 주파수만을 사용하나, 본 논문에는 잡음 구간을 제거하기 위한 유성음 외의 구간을 모두 0으로 처리한 후 이동평균필터 범위를 0으로 처리된 값까지 포함하여 기본 주파수를 산출한다.

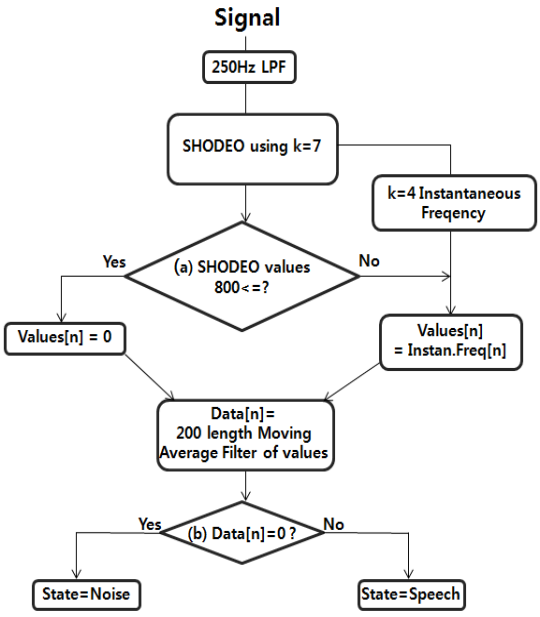
[Fig. 4]에서 추정된 고차 미분에너지 함수를 이용한 순간 주파수 [Fig. 4](c)가 [Fig. 4](a)의 TF32 와 [Fig. 4](b)의 MATLAB에서 제공하는 Spectogram과 유사한 패턴이 나타남을 확인할 수 있었다. 이는 본 논문에서 제안하는 고차 에너지 함수를 이용한 음성 구간 검출 (Voice activity detection) 알고리즘에서 주파수 값의 대소는 크게 유의하지 않음을 나타낸다.



[Fig. 4] Comparison of spectrogram generated by (a) TF32, (b) MATLAB, and (c) higher order differential energy function

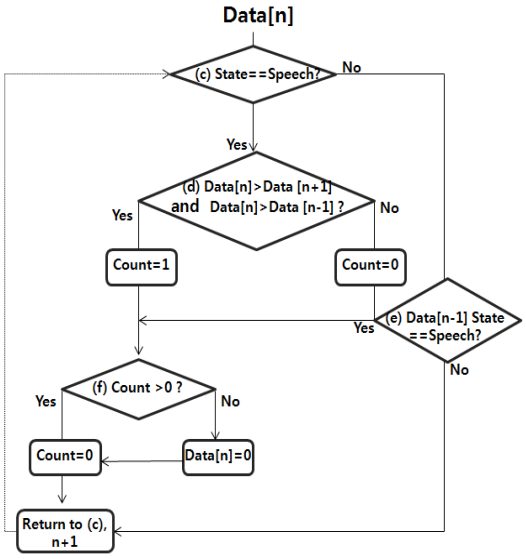
[Fig. 5]는 1차적 음성 구간 검출을 위한 알고리즘의 블록도를 나타낸다. 본 실험에서 기본 주파수를 추산하는 과정에서 유성음 외 구간을 모두 0로 처리하였기 때문에 기본 주파수가 0보다 큰 값을 가지는 구간은 음성이 존재하는 구간이라 판정할 수 있다.

높은 주파수 대역을 가지는 잡음 구간과 무성음 및 묵음 구간은 250Hz의 저역통과필터와 800의 임계값을 가지는 미분에너지 함수에 의하여 다수 제거되나, 노인 음성 특성상 불규칙적인 잡음이 음성 구간에 잔존되어 있어 추가적인 후처리 작업이 요구된다.



[Fig. 5] Voice Activity Detection Algorithm (Step 1)

앞서 수식 (4)와 (5)에 의해 0의 값을 포함한 이동평균 필터를 이용함으로 200의 데이터 샘플길이 동안 0의 값을 계속해서 더하고 나눔으로써 짧은 길이를 가지는 단발적 잡음 신호는 최종적으로 0으로 탈락되거나 Peak값이 나타나지 않는 특징을 가지게 된다.

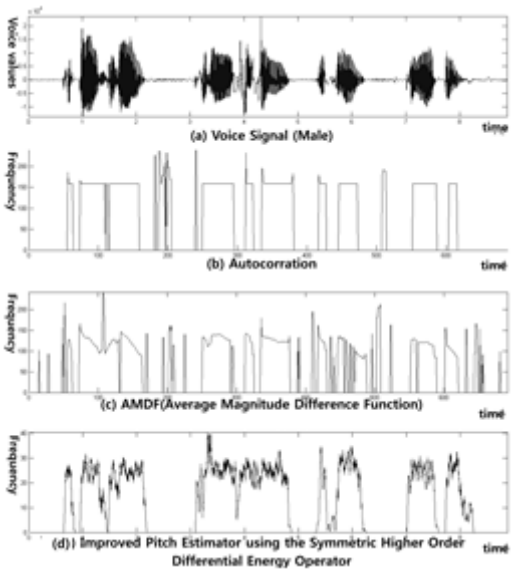


[Fig. 6] Voice Activity Detection Algorithm (Step 2)

[Fig. 6]는 앞서 1차적으로 검출된 음성 구간 판정 결과를 기준으로 [Fig. 6](c)와 같이 입력 값의 State가 Speech일 때 [Fig. 6](d)과 같이 Peak의 유무를 판단한다. 그리고 [Fig. 6](e)에서 Speech의 종료 시 Peak의 유무를 확인하여, [Fig. 6](f)에서처럼 Peak값이 나타나지 않은 구간은 잡음 구간이라 판정하여 0로 초기화시킨다.

4.2 음성 구간 검출 성능 비교

[Fig. 7](a)는 72세 남성이 발화한 “그때 누가 그녀의 책상 앞으로 다가왔다” 문장의 파형을 나타낸다. [Fig. 7](b), [Fig. 7](c)와 [Fig. 7](d)는 각각 자기상관함수와 AMDF, 고차 미분에너지 함수를 (SHODEO)를 이용한 기본 주파수 추정 방법을 사용한 음성 구간 검출 성능 비교 결과를 나타낸다. 기존의 방법들과 비교하였을 때 [Fig. 7](d)의 음성 구간 검출 알고리즘이 보다 정확히 음성 구간을 검출하고 있으며, 잡음 구간에도 우수한 성능을 보임을 확인할 수 있다.

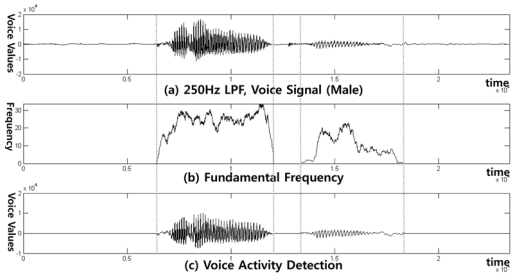


[Fig. 7] Comparison among Various Voice Activity Detections

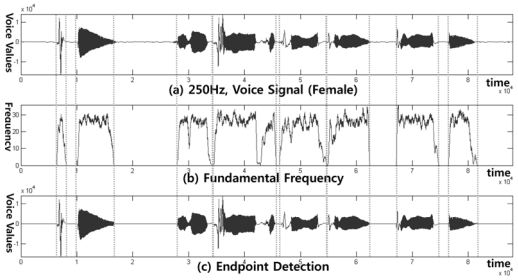
4.3 음성 구간 검출 결과

[Fig. 8]과 [Fig. 9]는 [Fig. 7](d)의 음성 구간 검출 알고리즘을 이용하여 다양한 데이터베이스에서의 음성 구간 검출 결과를 보인다. [Fig. 8](a)는 74세 남성의 “끌어

안고"라고 발화된 단어이고, [Fig. 9](a)는 73세 여성의 "그때 누가 그녀의 책상 앞으로 다가왔다"고 발화된 문장의 파형을 나타낸다. [Fig. 7](d)의 제안된 음성 구간 검출 알고리즘을 이용하여 음성 구간을 검출하였을 때 잡음 구간을 배제하고 보다 정확히 음성 구간만을 검출함이 [Fig. 8](c)와 [Fig. 9](c)를 통해 확인되었다.



[Fig. 8] Word (Male)



[Fig. 9] Sentence (Female)

5. 결론

본 연구는 노인 음성 잡음을 고려하여 기존의 대칭 구조 고차 미분 에너지 함수(SHODEO)를 이용하여 노인 음성에서의 음성 구간 검출 알고리즘을 개발하였다. 또한 그것은 자기상관함수와 AMDF 방법과 비교하여 노인 음성 검출에 보다 우수한 성능을 가지는 것을 확인하였다. 고차 미분 에너지 함수는 유/무성음 구간에서 진폭의 차이가 두드러지며 순간 주파수가 무성음 구간에 불규칙하게 나타나는 특성에 기반하여, 노인 음성 구간마다 나타나는 잡음을 제거함으로써 보다 정확한 노인 음성 검출을 가능하게 한다. 또한, 음성 신호의 시변적 특성을 반영하고 적은 계산량을 요구하여 실시간에 가까운 음성 구간 검출이 가능하다는 장점을 가진다. 따라서 본 논문에서 제시하는 음성 구간 검출 알고리즘은 노인을 위한 음성 인터페이스에 적용함으로써 노인들의 스마트 기기에의 접근성을 높이고, 더 나아가 노인들을 위한 융복합 웨어러블 디바이스(Wearable Device)의 다양한 개발이 가능할 것으로 전망한다.

ACKNOWLEDGMENT

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No. 2014-00540001).

REFERENCES

[1] Yong-Wook Nam, Yong-Hyuk Kim, "Speed estimation of sound-emitted objects through convergence of sound information analysis and smart device technology", Journal of the Korea Convergence Society, Vol. 6, No. 5, pp. 233-240, 2015

[2] EunJeong Choi, SERI management note, vol. 117, pp. 1-14, SERI, 2011

[3] Seong-Hoon Lee, Dong-Woo Lee, "On Issue and Outlook of wearable Computer based on Technology in Convergence", Journal of the Korea Convergence Society, Vol. 6, No. 3, pp. 73-78, 2015

[4] Yunkyung Song, "Prevalence of Voice Disorders and Characteristics of Korean Voice Handicap Index in the Elderly", Journal of the Korean society of speech science, Vol. 4, No. 3, pp. 151-159, 2012

[5] Soon-Kyeom Kim, Jamg-Eui Hong, "Application of Safety Analysis and Management in Software Development Process", Journal of Convergence Society for SMB, Vol. 6, No. 1, pp. 7-15, 2016

[6] Kahane, J. C. "Anatomic and physiologic changes in the aging peripheral speech mechanism. In D. S. Beasley & G. A. Davis (Eds.)," Aging: Communication processes and disorders New York: Grune & Stratton.

pp. 21-45, 1981

[7] Lee, S.Y. “The overall speaking rate and articulation rate of normal elderly people,” Graduate program in speech and language pathology, Master these, Yonsei University, 2011

[8] Hong Jungpyo; Park Sangjun; Jeong Sangbae; Hahn Minsoo, "Robust Feature Extraction for Voice Activity Detection in Nonstationary Noisy Environments", Journal of the Korean society of speech science, Vol. 5, No. 1, pp. 11-16, 2013

[9] Byeong-Gwan Iem; "Estimation of Fundamental Frequency Using an Instantaneous Frequency Based on the Symmetric Higher Order Differential Energy Operator", The Korean Institute of Electrical Engineers, Vol. 60, No. 2, pp. 2374-2379, 2011

[10] Byeong-Gwan Iem, "An Instantaneous frequency estimators based on the symmetric higher order differential energy operator," IEICE Trans. Fundamentals, vol. E93-A, no. 1, pp. 227-232, 2010

[11] P. Maragos, and A. Potamianos, "Higher order differential energy operators," IEEE Signal Processing Letters, Vol. 2, pp. 152-154, 1995

[12] In-Kyu Seo, Sang Ho Lee, “An Efficient Hospital Service Model of Hierarchical Property information classified Bioinformatics information of Patient”, Journal of Convergence Society for SMB, Vol. 5, No. 4, pp. 17-23, 2015

[13] K. Abdullah-Al-Mamun, "A High Resolution Pitch Detection Algorithm Based on AMDF and ACF", J. Sci. Res. 1, pp. 508-515, 2009

[14] Myungkyu Ham; Sungyoung Choi; Jongcheol Park; Myungjin Bea; "On a Pitch Point Detection by Preserving the Phase Component of the Autocorrelation Function", 2000 Korea Signal Processing Conference, Vol. 13, No. 1, pp. 799-802, 2000

[15] LAWRENCE R. RABINER, “On the Use of Autocorrelation Analysis for Pitch Detection”, IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-25, NO. 1, FEBRUARY 1977

[16] Hyun-Soo Seo, "Pitch Period Detection Algorithm Using Modified AMDF", The Korea Institute of Information and Communication Engineering, Vol , 10, No. 1, pp. 23-28, 2006

이 지 연(Lee, Ji Yeoun)

・2003년 2월 : KAIST, 전자공학과 (공학석사)
・2008년 8월 : KAIST, 전자공학과 (공학박사)
・2008년 9월 ~ 2011년 2월 : UCLA, University of Wisconsin-Madison, 연구원
・2011년 3월 ~ 현재 : 중원대학교 의료공학과 조교수
・관심분야 : 생체신호처리, 의료전자, 의공학기술
・E-Mail : jylee@jwu.ac.kr