1993년도 한국음향학회 학술논문발표회 논문집(제 12권 1(s)호)

음성신호의 진폭분포를 이용한 유/무성음 검출에 대한 연구

o
(*) 배 성균 (**) 백금란 (**) 배 명진 (*) 김 형래
(*) 건국대학교 전자공학과 (**) 숭실대학교 정보통신공학과

(The Magnitude Distribution method of U/V decision)

(*) Seonggyun-Bae (**) Guemran-Baek (**) Myungjin-Bae (*) Hyungrae-Kim
(*) Kon-kuk University (**) Soongsil University

ABSTRACT

In speech signal processing, The accurate detection of the voiced/unvoiced is important for robust word recognition and analysis. This algorithm is based on the MD in the frame of speech signals that does not require statistical information about either signal or background-noise to decide a voiced/unvoiced. This paper presents a method of estimating the Characteristic of Magnitude Distribution from noisy speech and also of estimating the optimal threshold based on the MD of the voiced/unvoiced decision. The performances of this detectors is evaluated and compared to that obtained from classifying other paper.

Ⅰ. 서 론

음성신호는 그 발성음원에 따라 유성음, 무성음 및 묵음으로 구분할 수 있고, 유성음의 음원은 시간영역에서 준주기적인 성질을 갖는데 이 주기를 피치라 한다. 특히, 피치를 구할 때 필요한 아주 중요하고 어려운 문제 중에 하나는 음성신호의 유·무성음구간을 정확히 결정하는 문제이다.

지금까지 유성음구간을 검출하기 위해서 사용하는 파라미터들은 유성음에서 성대의 떨림에 의해 발생하는 거의 안정된 주기적 성질을 이용하거나 성도에서 나타나는 공명현상을 이용하기 위해서 에너지 펄스를 사용하고 있다. 그렇지만 이러한 파라미터를 사용하여 유성음구간을 분류하려고 하면 몇 가지의 문제점이 발생한다. 안정된 주기를 측정하는 경우에 파열음이나 천이구간이 존재하는 음성에서는 안정된 주기가 구해지지 않아 에러가 발생하게 된다. 또한 에너지 펄스를 사용하여 유성음구간을 분류할 때는 에너지가 낮은 유성 자음구간에서는 배경 잡음 에너지와 구분이 어렵게 되고, 저주파수의 에너지와 고주파수의 에너지가 서로 구분이 되지않기 때문에 유성음구간을 분류하기가 어려워지게 되는 문제점을 가지게 된다.

이러한 문제점에서 유·무성음 및 묵음구간 검출 알고리듬은 음성신호가 아닌 모든 신호들을 분리해 내어야 하고, 주변 잡음 뿐만 아니라, 임펄스성 잡음에도 강인하여야 한다. 또, 음절 간에 생기는 묵음구간도 음성신호에 포함시켜서 검출하여야만 한다.

따라서 본 논문에서는 음성신호를 긴시간 동안에 관찰해 보면 Gamma 분포에 가까운 분포 특성을 갖는 것으로 알려져 있으나, 단시간으로 보게 되면 그 특성이 유·무성음 및 묵음에 따라 서로 다른 분포를 가진다는 것을 이용하여 유·무성음 및 묵음구간 검출 알고리듬을 새로이 제안하고자 한다.

먼저, Ⅱ 장에서 음성생성 모델에 대해, Ⅲ 장에서는 음성신호의 단시간 진폭분포 특성을, Ⅳ 장에서는 단시간 진폭분포를 이용한 유·무성음 및 묵음구간 검출 알고리듬에 대한 설명, Ⅴ 장에서는 실험 및 결과, Ⅵ 장에서는 결론의 단계로 기술하였다.

Ⅱ. 음성생성 모델

음성신호는 발성 모델에 따라 유·무성음 및 묵음으로 분류될 수 있다. 유성음은 준주기적인 성대 펄스가 성도를 통해 감으로서 발생되기 때문에 유성음 각 원소마다 성대에서 고유한 공명이 일어난다. 따라서 유성음의 스펙트럼은 음소마다 고유한 공명 봉우리를 갖게 된다. 이러한 공명 봉우리를 포먼트라하며 낮은쪽 주파수에서 부터 두드러진 포먼트를 차례로 제 1, 제 2, 제 3 포먼트 등으로 불리어 진다. 유성음의 스펙트럼에서는 보통 제 1 포먼트의 주파수가 250-750Hz에 존재하며, 또한 공명현상 때문에 무성음에 비해 에너지가 크고, 성대의 진동에 의해 성도의 여기가 되기 때문에 준주기성을 띠게 된다.

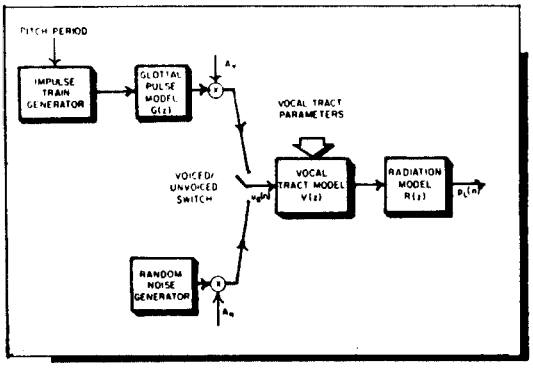


그림 2-1 음성생성 모델
Fig. 2-1. A model for speech production

성대의 진동주기는 남녀노소 및 주변 환경에 따라 다르지만 대략 2.5-25msec 정도가 된다.

그러나, 무성음은 불규칙한 잡음이 성대를 자극하는 입력으로 되어 성대를 통과하는 동안 성대의 협착점에서 공명이 발생하게 되며, 따라서 무성음의 스펙트럼에서 2500Hz 근처에서 주된 공명 봉우리가 존재하게 된다. 지금까지 제안된 알고리듬은 이러한 분석특성을 이용하여 유·무성음을 분류하였다.

Robert Ito와 Donaldson은 전력 스펙트럼 밀도와 영교차율의 수학적인 관계를 이용하여 음성신호에 대한 스펙트럼 측정과 유·무성음을 분류하는데 사용하였다. 특히, 모음의 포먼트 파라미터를 측정할 때 영교차율의 관계식을 이용하였다. 그리고, 측정된 영교차율로서 모음, 무성음, 마찰음 등을 서로 다른 주변 환경에서 분류를 하였다[10].

Atal과 Rabiner는 유·무성음 및 묵음구간 분류를 위한 특징 파라미터로 다음 다섯 가지를 제안하였다.

(1) 영교차율
(2) 신호의 에너지
(3) 1차 자기 상관 계수
(4) 선형 예측 계수
(5) 예측 오차 에너지

이 다섯 개의 파라미터 중 필요에 따라 2-3개를 사용하여 분류하였고, 특징 파라미터를 사용하여 잡음이 심한 환경에서 좋은 결과를 얻을 수 있었다[8]. 그러나 Robert Ito와 Donaldson 그리고 Atal와 Rabiner의 분류 방법은 처리과정이 매우 복잡하고, 많은 시간이 소요되는 단점을 가지고 있다.

III. 음성신호의 단시간 진폭분포 특성

음성신호의 유·무성음 및 묵음구간 검출 알고리듬은 주변 잡음의 변화나, 음성신호의 크기나 상황의 변화에 따른 영향을 받지않고 바르게 검출해야 하며, 또 전처리과정의 일부로서 유·무성음 및 묵음구간 검출 알고리듬이 사용될 때는 그 처리과정이 간단해야 하며 실시간 처리가 가능해야 한다.

특징 파라미터 추출에 의한 유·무성음 및 묵음구간 검출 알고리듬은 수백 msec 동안의 training 데이타에 의한 주변 잡음의 통계적 특성을 추출하게 한다. 이것은 프레임에 대한 평균 에너지나 영교차율의 최대, 최소값, 평균, 표준편차들을 측정하여 분류과정에 사용하게 된다. 그러나 이것은 주변 환경이 시간에 따라 변하지 않을 때에 가능하기 때문에 주변 환경 변화에 적응하기 위해 유·무성음구간을 검출할 때마다 그 파라미터의 통계적 특성을 측정하는 것은 불합리한 것이다. 따라서, 간단하면서도 주변 환경의 변화에 무관한 파라미터를 사용하는 유·무성음 및 묵음구간 검출 알고리듬이 필요하게 된다. 또한, 음성신호의 단시간 평균 에너지는 유성음일 때는 그 값이 크므로 유·무성음 및 묵음구간 검출이 쉬우나, 특히 다음 경우에는 주변 잡음과 구별하기가 곤란하다.

(1) 음성의 시작이나 끝 부분에서의 약한 마찰음
(2) 음성의 시작이나 끝 부분에서의 약한 파열음
(3) 음성의 끝 부분에서의 비음
(4) 음성의 끝 부분에서의 모음 에너지의 감소

그리고, 음성신호의 크기가 갑자기 작아질 때에도 평균 에너지를 파라미터로 사용하기가 곤란하다.

이런 점을 감안하여 본 논문에서는 단시간 진폭분포를 파라미터로 하는 음성신호의 유·무성음 및 묵음구간 검출을 수행하는 새로운 알고리듬을 제안하고자 한다. 음성신호를 단시간으로 처리할 때, 진폭분포를 보면 유성음인 경우에 어떤 평균값을 기준으로 넓게 분포하고, 평균에서 멀어질수록 점차 감소하는 분포특성을 가지며, 무성음인 경우에는 영을 기준으로 폭넓게 분포하게 된다. 그 반면에 묵음구간은 영을 기준으로 하여 많은 샘플들이 밀집되어 있다는 것을 알 수 가 있다(그림 3-1 참조).

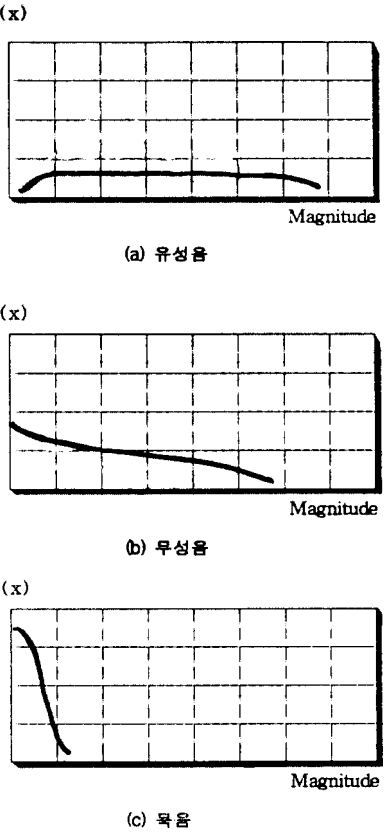


그림 3-1. 단시간 진폭분포 특성
Fig. 3-1. Characteristic of Magnitude Dist. in a frame

이러한 음성신호의 진폭분포 특성을 이용하여 유·무성음 및 묵음구간 검출을 수행할 때 잡음이나 신호의 크기에 영향을 받지 않고, 그 분포 특성만을 이용하기 때문에 주변 환경 변화에 영향을 받지 아니하며, 또한 임펄스성 잡음에도 강인하다. 처리과정도 단순하므로 인식기의 전처리과정으로 유·무성음 및 묵음구간을 검출할 수 있다.

IV. 진폭분포를 이용한 유·무성음 및 묵음구간 검출 알고리듬

진폭분포를 이용한 유·무성음 및 묵음구간 검출의 블록도는 그림 4-1에 나타내었다. 본 논문의 알고리듬은 입력 음성신호를 순차적으로 받은 데이타에 절대값을 취한후 단시간에서 진폭분포를 구하고, 문턱값을 설정하여 구간별 검출을 수행

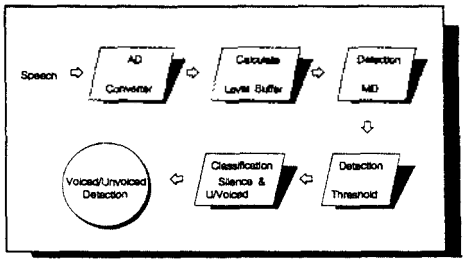


그림 4-1. 제안된 유 ' 무성음 및 묵음구간 검출 블록도
Fig. 4-1. Block diagram of proposed V/U detector.

하였다.

먼저, 입력으로 받은 A/D 변환된 데이타에 절대값을 취한 후 2^b(b = 양자화 비트수)의 레벨로 나누어 이에 대한 진폭분포(Magnitude Distribution)를 구한다. 즉, s(n)을 입력 샘플이라고 하면,

$$0 \le |s(n)| \le 2^{b-1}-1 \qquad (4\text{-}1)$$

이 된다. 여기서 b는 양자화 비트수이다. 그리고 아래와 같이 델타 함수를 정의하면,

$$\delta(k)=\begin{cases}1, & k=0\\ 0, & \text{otherwise}\end{cases} \qquad (4\text{-}2)$$

이를 이용하여 진폭분포는 다음과 같이 구할 수 있다 :

$$MD(f,i) = \sum_{n=0}^{N-1}\delta(|s(n)|-i), \quad i = 0, 1, \ldots, 2^b-1$$

$$N = \text{프레임 길이} \qquad (4\text{-}3)$$

여기서, f 를 프레임 번호, i 는 입력 표본의 레벨이다.

이러한 진폭분포도에서 유 ' 무성음 결정알고리듬은 다음과 같다. 최대레벨의 10% 이내에 드는 빈도수가 40% 이상이면 묵음구간이라하고, 그렇지 않으면 음성신호가 존재하는 구간으로 결정한다. 음성신호가 존재하는 구간으로 판정이 되면, 다시 최대레벨의 20-30% 사이의 빈도수가 15% 이상이면 무성음, 그 이하이면 유성음으로 결정을 한다. 이러한 과정의 수치들은 배경 잡음의 영향을 고려하여 반복 실험한 결과이다.

V. 실험 및 결과

이상의 것을 컴퓨터 시뮬레이션하기 위해 마이크가 장치된 A/D 변환기를 IBM-PC/486(DX II)에 인터페이스 시키고, 아래의 발성을 8KHz의 샘플링 주파수로 양자화하여 저장한 다음, 시뮬레이션에 대한 시료로 사용하였다. 각 시료는 각각 25세 28세 32세 남성 화자가 배경 잡음이 있는 환경에서 5번씩 반복 발성한 것을 데이타로 하였으며 시료들은 아래와 같다.

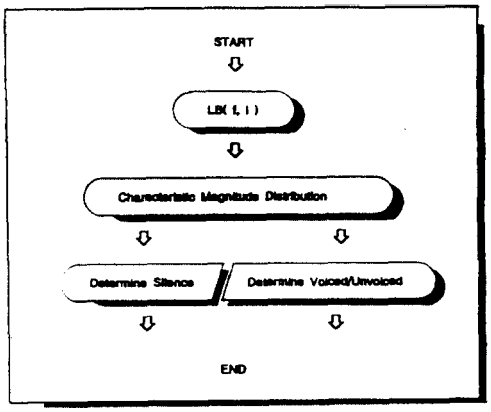


그림 5-1. 제안된 유 ' 무성음 및 묵음구간 검출 순서도
Fig. 5-1. Flowchart of proposed V/U detector

발성 1 : "인수네 꼬마는 천재 소년을 좋아한다."
발성 2 : "예수님께서는 천지창조의 교훈을 말씀하셨다."
발성 3 : "공일이삼사오육칠팔구."

그리고, 각 음성시료에 대해 한 프레임 길이를 128샘플로 하여 64샘플 단위로 겹쳐가며 분석과정을 수행하였다.

진폭분포도를 이용한 유 ' 무성음 및 묵음구간을 검출하는 알고리듬의 순서도는 그림 5-1에 나타내었다. 실험과정은 먼저 입력 음성신호를 순차적으로 받아들여 데이타에 대해서 절대값을 취한후에 A/D 변환된 데이타에 대해서 진폭분포도를 만든다. 그리고, 최대레벨의 10%이내의 빈도수가 40% 이상이면 묵음구간으로하고, 이하이면 음성신호가 존재하는 구간으로 결정한다. 음성신호가 존재하는 구간으로 결정이되면, 다시 최대레벨의 20-30% 사이의 빈도수가 15% 이상이면 무성음, 그 이하이면 유성음으로 결정을 한다.

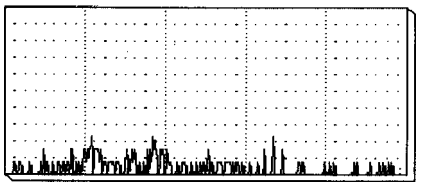
실험 및 결과에서 알수 있듯이 제안한 알고리듬은 구분하기 어려운 부분도 매우 효과적으로 구분함과 동시에 간단한 계산과정으로 처리됨을 알수가 있다.

VI. 결 론

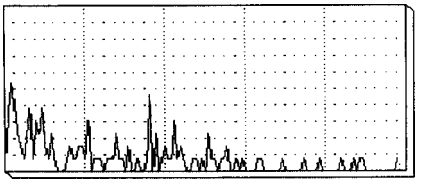
음성신호에서 유성음구간을 분류해내는 문제는 피치를 구할때 정확도를 좌우하는 중요한 문제로서 음성 신호처리 분야에서는 아주 필요하고 어려운 작업들이다. 유 ' 무성음 분류에 대한 대다수의 논문들은 피치를 구하는 과정속에 보통 포함시키고 있으며, 사용하는 파라미터도 유성음이 갖는 주기성이나 혹은 에너지의 크기에 제한시키고 있다.

그럼지만 본 논문에서는 음성신호에 대한 단시간 진폭분포 특성을 이용한 새로운 유 ' 무성음 및 묵음구간 검출 알고리듬을 제안하였다.

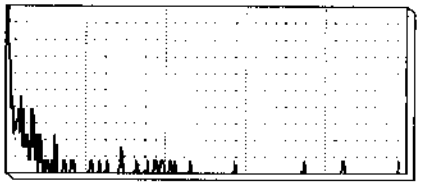
음성신호는 전체적으로 관찰해 보면 데이타가 Gamma 분포에 가까운 분포를 갖는 것으로 알려져 있으나 단시간으로 관찰해보면, 그 특성이 유 ' 무성음 및 묵음에 따라 서로 다른 진폭분포 특성을 가진다는 것을 알 수 있다. 이와 같이 단시간 진폭분포 특성을 이용한 방법은 잡음에 강인하며, 계산량이 작음과 동시에 효율적으로 구간 검출이 수행되었다.



(a) 유성음



(b) 무성음



(c) 묵음

그림 6-1. 유 · 무성음 및 묵음구간 검출 결과
Fig. 6-1. Results of silence & U/V detection.

참고 문헌

[1] 이인섭, 최정아, 배명진, 안수길, "음성 신호의 진폭 분포를 이용한 끝점 검출 알고리듬", 대한전자공학회. 하계종합학술대회, 제 12권, 제 1호, 1989.7.

[2] 배명진, 안수길, "배경잡음이 있는 음성신호에서 Explicit 끝점 검출", 대한전자공학회. 추계종합학술대회, 제 10호, 제 1호, 1987.11.

[3] H. Kobatake, "Optimization of Voicd/Unvoiced Decisions in Nonstationary Noise Environments", IEEE Trans. on ASSP, Vol. ASSP-35, No. 1, pp. 9-18, Jan., 1987.

[4] Shuzo Saito, Kazuo Nakata, Fundamentals of Speech Signal Processing : Academic-Press, 1985

[5] S.G. Knorr, "Reliable Voiced/Unvoiced Decision", IEEE Trans. on ASSP. Vol. ASSP-27, No. 3, June, 1979.

[6] L.R. Rabiner and R.W. Schafer, Digital processing of Speech Signals Englewood Cliffs, New Jersey : Prentice-Hall, 1978

[7] L.R. Rabiner, B.S Atal, "Evaluation of a Statistical Approach to Voiced-Unvoiced-Silence Analysis for Telephone-Quality Speech", B.S.T.J., Vol. 56, No. 3, pp. 455-482, March, 1977.

[8] B.S. Atal, L.R.Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Application to Speech Recognition", IEEE Trans. on ASSP, Vol. ASSP-24, No. 3. June, 1976.

[9] L.R. Rabiner, M.R. Sambur, "An Algorithm for Determing the Endpoint of Isolated Utterances", B.S.T.J. Vol. 54, No. 2, pp. 297-315, Feb., 1975.

[10] Mabo Robert Ito, Robert W. Donaldson, "Zero-Crossing Measurements for Analysis and Recognition of Speech Sounds", IEEE Trans. on A.A.E. Vol. AU-19, No. 3 pp. 235-242, Sep., 1971.