음성의 잡음레벨 추정을 위한 피치간 유사도 측정에 관한 연구

강인규, 강성모, 배명진
숭실대학교 정보통신공학과

A Study on the relation of closed pitch for Noise-Level Measurement.

InGyu Kang, SungMo Kang, MyungJin Bae
Dept. Information and Telecommunication Engr, Soongsil Univ
E-mail: skkig@korea.com

요약

인간은 "습관적 피치 레벨" 즉 자연스럽게 말할 때 평균적으로 사용하는 피치를 갖는다. 하지만 음성에 잡음이 첨가 되면 이 피치가 불규칙하게 바뀌게 된다. 이 점을 이용하여 음성의 잡음레벨을 측정할 수 있다. 본 논문에서는 입력음성의 에너지를 구하고 일정 에너지레벨 이상에서의 구간에 대해 NAMDF(Normalized Average Magnitude Difference Function)방법으로 피치를 구하고, 각 프레임을 피치단위로 분절한 뒤 인근 피치간의 유사도를 측정하여 입력음성데이터의 잡음레벨을 검출하는 방법을 제안하였다.

1. 서론

음성은 사람이 다른 도구 없이 사용하는 정보 전달 매체로서 가장 많이 이용될 뿐 아니라 가장 간편한 수단이다. 음성을 통하여 의사전달을 할 때 잡음레벨의 정도에 따라 음성을 통한 의사전달은 영향을 받게 된다. 이처럼 잡음이 끼치는 영향은 음성인식, 합성 및 분석과 같은 음성신호처리 분야에 있어서 매우 크다.

음성은 크게 유성음과 무성음으로 나눌 수가 있다. 일반적으로 유성음의 음원은 준주기적인 펄스, 무성음의 음원은 백색잡음으로 모델링을 한다. 또한 이들의 구분은 에너지비나 영교차율법을 많이 사용하며 에너지비는 에너지가 높고 영교차율법은 영교차율이 낮은 부분을 유성음으로 결정하고, 이를 제외한 구간을 무성음 구간으로 결정한다. 즉, 유/무성음 결정은 무성음의 에너지와 주기적 에너지비를 비교하여 주기적 에너지가 높은 주파수대역을 유성음으로 분류하는 반면에, 이 비가 낮은 주파수 대역을 무성음으로 분류한다. 이 때 유성음은 준 주기적인 피치를 갖게 되는데 여기에 잡음이 섞이게 되면 피치파형이 잡음과 섞여져 랜덤(Random)한 특성을 갖게 된다. 따라서 프레임을 피치주기 단위로 분절하여 프레임안에 피치파형들을 비교분석 해보면 음성에 섞인 잡음레벨을 추정할 수 있다. 2장에서는 본 논문에 사용되어진 에너지검출에 따른 음성분석구간 설정과 프레임을 피치단위로 분절하기 위해 사용된 NAMDF의 피치검출방법을 설명하고 3장에서는 본 논문에서 제안한 방법인 피치단위로 프레임을 분절한 뒤 프레임내의 인근피치단위파형 비교분석을 통해 임이의 음성시료에 대하여 잡음레벨을 추정할 수 있는 알고리즘을 제안하며 4장에서는 실험 및 결과, 그리고 5장에서는 결론을 맺는다.

2. 에너지 검출법 및 피치 검출법

에너지 검출과정과 피치검출과정은 여러 가지 음성처리 시스템에 필수적인 요소이다. 음성은 에너지비가 상대적으로 높은 유성음구간에서 준 주기적인 피치주기

형태를 갖는 특성이 있다. 이렇게 검출된 일정 에너지 레벨이상 음성구간의 피치변화도는 화자 인식용 및 발음 장애자를 위한 보조시스템용 파라미터로 널리 적용되는 등 거의 모든 음성 분석-합성(보코더) 시스템에 널리 쓰이고 있다.

2.1 단시간 에너지 검출법

음성신호의 진폭은 시간에 따라 변한다. 특히, 무성음의 진폭은 일반적으로 유성음 구간의 진폭보다 아주 작다. 음성신호에 대한 단시간 에너지는 이들 진폭변화를 반영하는 편리한 표현법이다. 단시간 에너지는 식(2.1)과 같이 정의할 수 있다.

$$E_n = \sum_{m=-\infty}^{\infty}[x(m)w(n-m)]^2 \qquad (2\text{-}1)$$

이 표현식은 아래 식(3-2)으로 쓸 수 있다.

$$E_n = \sum_{m=-\infty}^{\infty}x^2(m)\cdot h(n-m) \qquad (2\text{-}2)$$

여기서

$$h(n) = w^2(n) \qquad (2\text{-}3)$$

이다.

식 (2-3)는 그림.2-1에 표시된 그림으로 해석할 수 있다. 즉 신호 $x^2(n)$은 식(2-3)에 있는 것처럼 임펄스 응답이 h(n)인 선형필터를 통과한다.

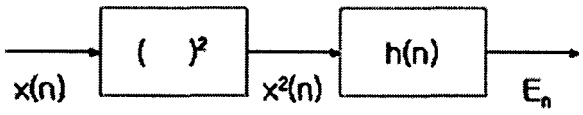

그림.2-1. 단시간 에너지의 블록도

2.2 단시간 NAMDF함수

AMDF법은 자기상관함수법과 다르게 x(n)과 x(n-k)곱 대신 다음과 같이 절대값으로 정의된다.

$$AMDF(k) = \sum_{m=-\infty}^{\infty}|x(m)-x(m+k)| \qquad (2\text{-}4)$$

AMDF법은 자기상관관계 함수법에서 수행하는 곱 연산을 절대값과 차분으로 대신하기 때문에 상대적으로 빠르다는 장점을 가지고 있다. 이러한 이유로 실시간에 많이 적용된다. 자기상관함수법 에서는 피치주기 배수에 최대값을 이루지만, AMDF법에서는 피치주기배수에 최소값을 갖는다. 그림.2-2은 유성음에 대한 AMDF법을 나타내었다.

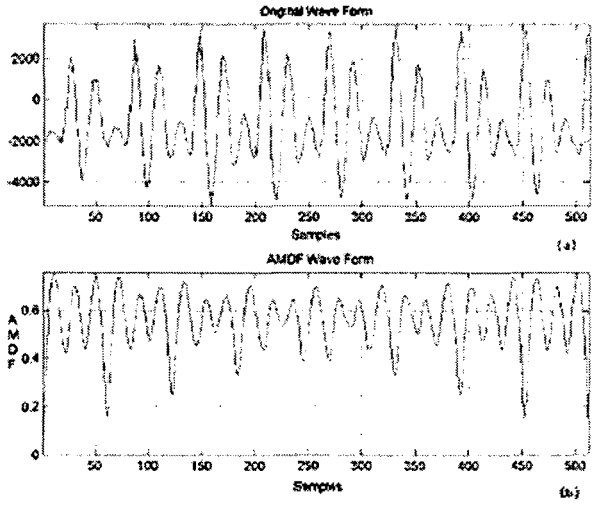

그림.2-2. 유성음에 대한 AMDF신호

(a)음성샘플 (b)AMDF

또한 현재의 프레임이 잡음구간이나 정상상태구간에 속하는지를 판정하는 방법은 현재의 프레임을 반분하였을 때 평균 진폭의 비를 측정하여 판정할 수 있다. 평균진폭의 비는 MR(fr)로 표현하고, 이것은 음성신호를 s(n)이라 할 때 식(2-5)과 같이 나타낼 수 있다.

$$MR(fr) = \frac{\sum_{k=N/2}^{N-1}s(n-k)}{\sum_{k=0}^{N/2-1}s(n-k)} \qquad (2\text{-}5)$$

여기서 변수 n은 프레임이 시작되는 첫 시퀀스의 위치이며 N은 프레임의 길이를 나타낸다. 이 평균 진폭비는 프레임 길이를 1/2로 했을 때의 인근 프레임의 평균 진폭비를 나타내기 때문에 창함수의 영향을 받게 된다. 창함수의 영향에 무관하게 현재의 프레임이 어떤 상태에 존재하는지를 측정하는 방법으로는 정규화된 AMDF (Normalized Average Magnitude Difference Function, NAMDF)방법이 제안되었다. 이렇게 피치단위로 신호를 분절하는 이유는 피치에 동기시켜 각 피치 구간의 상관관계로 음성신호의 전이구간을 검색하기 위해서 이다.

식 (2-6)은 피치 검색시 사용된 NAMDF식이다. s(n)는 음성신호이고 N은 윈도우 크기이며 d는 지연 인자이다.

$$NAMDF(d)=\frac{\sum_{n=1}^{N}|s(n)-s(n-d)|}{\sum_{n=1}^{N}[|s(n)|+|s(n-d)|]} \qquad (2\text{-}6)$$

3. 잡음레벨의 측정

본 논문에서는 먼저 입력음성에 대해 프레임단위로 단시간 에너지검출을 수행하고 에너지의 문턱 값(Threshold) 이상부분의 프레임에 대해서 NAMDF법으로 피치주기를 찾은 후 각 프레임을 피치주기단위로 음성신호를 분절한다. 음소변화는 음성파형에 비해 서서히 변화하기 때문에 프레임 단위로 분석하는 것이 보통이다. 창함수의 영향에 무관하게 현재의 프레임이 어떤한 상태에 존재하는지를 측정하는 방법으로 제안된 정규화된 AMDF (Normalized Average Magnitude Difference Function, NAMDF)방법을 사용하였다. 이렇게 피치단위로 신호를 분절하는 이유는 피치에 동기시켜 각 피치구간의 상관관계로 음성신호의 전이구간을 검색하기 위해서 이다.

식 (2-6)을 이용하여 피치주기(τ)를 구하고 입력음성신호를 이 피치주기 단위로 분절한다. 분절된 음성신호에 피치동기된 상관관계를 수식을 적용한다. 현재 피치주기의 상관관계 계수는 식 (3-1)과 같이 정의된다.

$$\beta(i)=\frac{E_{ij}}{E_{ii}}=\frac{\sum_{n=1}^{\min(\tau_i,\tau_j)}s_i(n)s_j(n)}{\sum_{n=1}^{\tau_i}s_i^{\,2}(n)} \qquad (3\text{-}1)$$

여기서 $\beta(i)$는 i번째 피치주기의 상관관계 계수이고 $s_i(n)$와 $s_j(n)$는 각각의 피치주기인 τ_i, τ_j에 의해 현재 미래 피치주기로 분절된 음성신호이다. 피치 동기된 상관관계 계수는 피치주기로 분절된 두신호의 상관관계가 클수록 1에 가까운 값을 가지게 된다.

$$VR(i)=|1-\beta(i)|^2 \qquad (3\text{-}4)$$

본 논문은 잡음레벨에 민감하게 적응하기 위해서 상관관계 계수를 변화시켜 잡음레벨 검출에 필요한 파라미터를 얻어내었다. 그림.3-1은 잡음구간을 추출하는 과정의 블록도이다. 프레임단위로 분절하기 위해 Hamming Window를 사용하여 프레임별 단구간 에너지를 구하였다. 이때 만일 계산된 에너지값이 정해진 문턱 값보다 작으면 β값을 1로 설정한다. 그리고 계산된 에너지값이 정해진 문턱 값보다 높으면 단구간 예상피치를 검색한다.

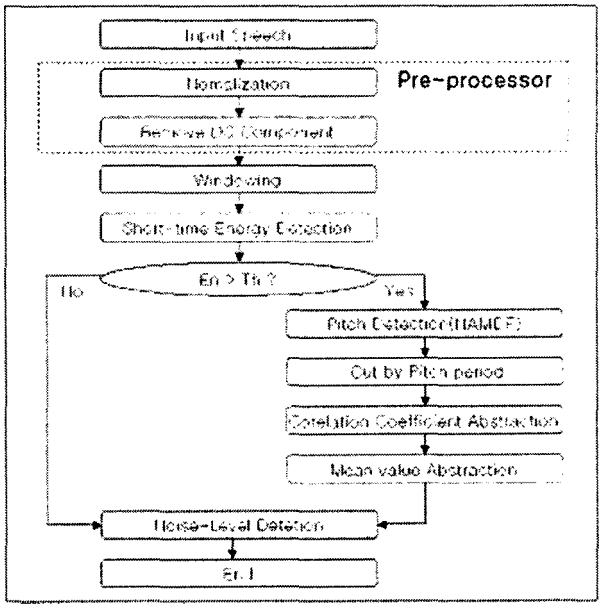

그림.3-1. 잡음레벨 검출의 블록도

이렇게 구하여진 피치를 사용하여 프레임별로 그 피치주기만큼 파형을 자른 후 피치주기단위(τ)로 분절된 파형에 대해 상관관계를 적용하여 이득계수(β), 즉 상관관계 계수를 추출한다. 그리고 이렇게 구해진 값은 안정된 구간에서는 1에 가까운 값이 나타나지만 잡음이 부가 되어지면 잡음레벨에 비례하여 변화가 일어난다. 이러한 성질을 이용하기 위해 1에서 이 값들을 빼서 제곱을 하면 새로운 파라미터(VR)가 된다. 이 VR은 상관계수의 값을 잡음구간과 안정구간에서 각각 큰 차이가 나게 한 값이다. VR은 안정구간에서는 0에 가까운 아주 작은 값이 나타나게 되고 잡음레벨에 따라 큰 변화를 나타낸다. 이렇게 하여 구하여진 각 프레임별 β값들을 합하고 전체 음성구간에 대해 평균을 내어 그 음성에 섞여있는 잡음레벨을 추정을 할 수 있다. β의 평균값인 NLF(Noise-Level Factor)는 식(3-5)와 같이 정의한다

$$NLF=\frac{\sum_{n=1}^{P}|1-\beta(p)|}{P} \qquad (3\text{-}5)$$

식(3-5)에서 β는 피치단위로 분절된 인접구간과의 상관

계수이고 P는 전체 음성구간에서의 총β값의 개수를 말한다. 이 NLF의 크기에 따라 입력음성의 잡음레벨을 측정할 수 있다.

4. 실험 및 결과

실험분석을 위해 IBP-PC/pentium(1.7GHz)에서 음성을 입력 받았다. 시뮬레이션을 위해 사용한 음성시료는 남녀 각각 30초 음성 30개씩을 사용하였으며 8kHz로 표본화 하였고 16bit로 양자화하여 사용하였다. Clean Speech에 대해 각각 30dB, 20dB, 10dB, 0dB의 SNR을 갖는 음성시료를 노이즈를 부과하여 생성하였고 이 때 사용되어진 잡음은 화이트노이즈, 핑크노이즈를 각각 첨가 시켜 실험하였다. 그리고 프로그램은 VC++와 Matlab을 사용하였다.

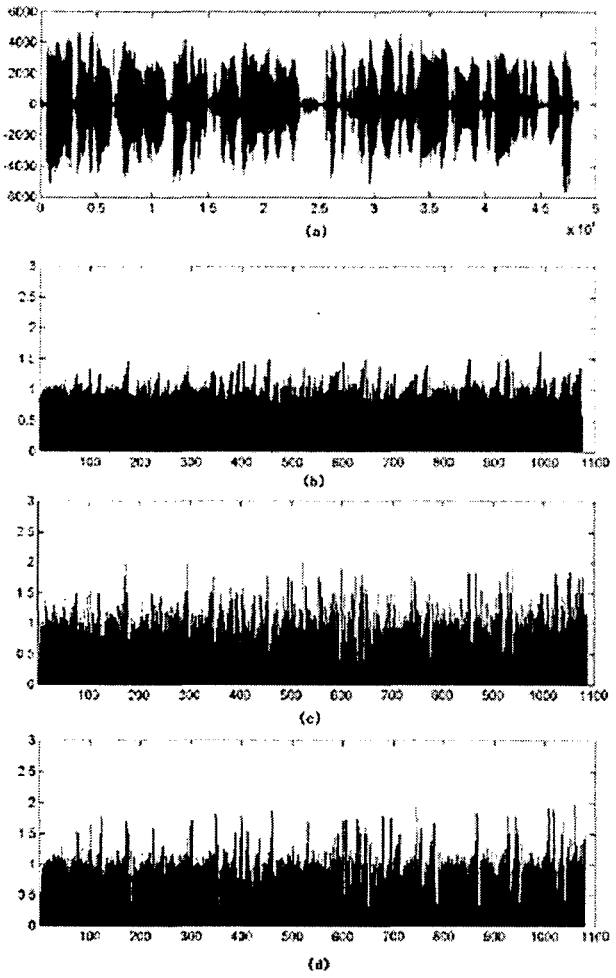

그림.4-1. 음성신호의 피치동기된 상관계수
(a) 입력음성파형, (b)상관관계 계수(β)
(c)AWGN(20dB)첨가 (d) 핑크노이즈(20dB)첨가

그림.4-1.에서 그림(a)는 실험의 환경의 원음성 이다. 그림(b)는 원음성의 인근피치분절신호간의 상관계수(유사도)로서 피치변화도를 나타내는 역할을 한다. 그림(c) 원음성에 20dB의 SNR을 갖는 화이트노이즈를 첨가시킨 시료의 β값을 나타내고 있다. 그림(d)는 원음성에 20dB의 SNR을 갖는 핑크노이즈를 첨가시킨 시료의 β값을 나타내고 있다. 현재피치와 인근피치의 값이 완전히 같다면 1이 된다. 원 음성시료에서 β값은 1에서 거의 변함없는 값을 가지며 잡음이 첨가된 음성시료에서는 매우 큰 변화를 갖는다는 것을 볼 수 있다. 표4-1은 위 실험을 30개의 음성시료와 노이즈레벨을 다르게하여 반복 실험한 결과들의 평균값이다

SNR \ Noise	AWGN	Pink Noise
Clean Speech	0.2132	
30dB	0.3312	0.3421
20dB	0.8872	1.0571
10dB	1.4302	1.9129
0dB	2.2331	3.0112

표4-1. 실험결과

5. 결론

본 논문에서는 인근 피치간의 유사도의 정도를 파라미터로 표현하고 음성시료의 잡음레벨을 측정할 수 있는 방법을 제안하고 실험을 통해 만족스러운 결과를 얻을 수가 있었다. 잡음이 없는 음성시료에서 경우에는 피치변화도계수가 1에 가까운 값을 갖는 반면 잡음이 부가된 음성시료의 경우에는 잡음의 종류에 따라 차이는 있었으나 노이즈레벨 부가량에 비례하여 NLF값이 상승하는 것을 확인하였다.

감사의 글

본 연구는 한국과학재단 특정기초 연구(과제번호 R01-2002-000-00278-0)의 지원에 의하여 이루어 졌습니다.

참고문헌

1. 한진수, "음성신호처리", 오성미디어, 2000년
2. 배명진, "디지털 음성분석", 동영출판사, 1998년
3. L.R.Rabiner, R.W.Schafer "Digital Processing of Speech Signals", Prentice-Hall, Inc.