음성인식을 위한 새로운 포만트트랙킹 알고리즘의 제안과 평가

An Proposal and Evaluation of the New Formant Tracking Algorithm for Speech Recognition

송 정 영*
Jeong-Young Song

요 약

본 논문에서는, 음성인식을 위한 한가지 방법으로 새로운 포만트 트랙킹 알고리즘을 제안한다. 본 연구에서는 실험을 위한 인식 데이터로, 한국어 숫자음성을 사용하였다. 새롭게 제안한 알고리즘을 사용하여 인식실험을 한 결과, 숫자음성 300개에 대한 인식률은 91%의 결과를 얻었다. 본 연구의 새로운 알고리즘은, 인식실험을 통하여 그 유효성이 확인되었다.

Abstract

For the speech recognition, this paper proposes a improved new formant tracking algorithm. The recognition data for the simulation on this paper are used with the Korean digit speech. The recognition rate of the improved algorithm for the Korean digit speech shows 91% for 300 digit speech. The effectiveness of this research has been confirmed through recognition simuluations.

1. 서 론

음성은 인간의 대뇌에서 만들어진 언어 정보가 조음기관에 신경명령으로 전달되어 음성파형이 형성된다. 이러한 음성의 음원(Speech Source) 은 폐에서 시작되어 성대를 거쳐 압축된 공기로 만들어진다. 이 공기의 흐름은 성대보다 위에 있는 음성기관과 여러 가지 조음기관들의 조음동작에 의해 결정되는 음향 전달 특성의 영향을 받아 음원의 주파수 성분이 선택적으로 공명함으로 음성이 생성된다[1,2,3,4,5,6,7]. 이러한 음원은 여기신호(Excitation Signal)에 대한 성도의 전달함수를 결정한다. 전달함수는 여러 개의 공명과 반공명으로 이루어지는 포만트(Formant)로 표현 가능하다.

일반적으로 포만트 주파수의 시간적인 변화에 대한 정보를 구하는 포만트 트랙킹은 음성인식을 위한 중요한 수단으로 알려져 있다. 포만트트랙킹에 대한 연구는 보고되어 있지만, 간단하고 적절하게 트랙킹을 행하는 방법은 현재 연구가 활발하게 진행중이다[8,9,10,11].

지금까지 알려진 포만트트랙킹의 알고리즘은 스펙트럼(Spectrum)의 피크(Peak)가 4개 이상 발생한 경우에만 고찰을 행하였기 때문에, 특히 비음구간에서 피크가 3개일 경우에는 참값을 적절하게 선택할 수 없는 것이 문제점으로 남아 있다[4,5,9]. 더욱이 한국어는 비음을 많이 포함하고 있기 때문에 한국어의 포만트트랙킹 알고리즘은 그대로 사용 할 수 없다.

본 논문에서는, 음성인식을 위해서 한국어의 비음의 특징을 적절하게 취하여 포만트트랙킹 알고리즘의 문제점을 보완하여 새로운 포만트트랙킹 알고리즘을 제안하고, 한국어 숫자음성을 입력 데이터로 하여 분석, 평가한 후 그 유효성을 확인한다.

* 종신회원 : 배재대학교 컴퓨터공학과 교수
jysong@mail.paichai.ac.kr

2. 한국어 숫자음성과 포만트트랙킹 알고리즘

2.1 한국어 숫자음성

단독으로 발성된 한국어 숫자음성은 표 1에 보인다. 이러한 숫자음성은 다음과 같은 성질을 갖는다.

(1) 단독숫자음은 모두 단음절이다.
(2) 무성자음 's', 'chi', 'p', 'k'는 어두부에만 나타난다. 또한, 6 'juk' 에대한 음소 'k'는 단독음의 어미이기 때문에 급하게 숨을 멈추는 동작으로 무음에 가깝다.
(3) 모든 숫자음성은 비음의 'm', 'ng', 유음의 'l' 인 유성자음 이거나 모음으로 끝난다.
(4) 숫자음성 '일', '이', '삼', '사'에 대하여는 비음 그자체가 각 숫자음성의 차이를 나타내는데 결정적인 요인이 된다.

또한, 한국어의 숫자음성에는 표 2에 보이는 바와 같이 어두부가 유성음인 경우와 무성음인 경우로 나누어진다. 본 연구에서는 이러한 특징을 이용하여 패턴 매칭을 행하도록 한다.

(표 1) 한국어 숫자음성

한국어 숫자음	발음
일	'il'
이	'i'
삼	'sam'
사	'sa'
오	'o'
육	'juk'
칠	'chil'
팔	'pal'
구	'gu'
영	'young'

(표 2) 어두부의 유성음과 무성음

유성음	무성음
'일', '이', '오', '육', '영'	'삼', '사', '칠', '팔', '구'

2.2 포만트트랙킹 알고리즘

2.2.1 수정된 알고리즘

음성의 기본적인 특징량인 포만트주파수의 시간적인 변화를 구해보면, 음성을 구별할 수 있는 1개의 기본적인 인자가 얻어진다. 그러나 어떤 방법으로 해서 스펙트럼으로부터 포만트주파수를 구하여 그 시간적 변화를 적절하게 살릴 수 있는가가 중요한 문제가 된다. 한국어 음성을 인식하는 경우에는 그 변화량을 결정하는 요소가 비음구간에서 영점의 영향을 받아 스펙트럼의 피크 대역폭이 넓어져서, 3개 이상의 피크를 얻을 수 없는 경우가 종종 일어난다. 한국어에 관한 대표적인 포만트트랙킹 알고리즘으로는, 대역폭을 조절하여 조건에 맞는 낮은 주파수로부터 3개의 피크를 포만트주파수로 선택하는 방법이 보고되어 있다. 그러나, 피크가 4개 발생한 경우에만 대역폭의 조건을 사용하기 때문에 비음구간에서 피크가 3개 발생되었을 경우에는 참값의 피크를 적절하게 찾아낼 수가 없게된다. 그리고 각 프레임마다 포만트주파수의 변화량을 500Hz 이하의 조건으로 하여, 이 조건에 맞지 않는 경우에는 1개 이전의 프레임으로 추정된 포만트주파수를 그대로 현재의 포만트주파수로 하는 방법을 택하고 있는데, 짧은 프레임간의 급격한 변화가 있는 경우에는 추적해서 찾아내기가 어렵게 된다.

그러므로, 본 논문에서는 각 프레임마다 조건에 맞는 포만트주파수가 2개 이하인 경우 바로 1개 이전의 프레임에서 추정된 포만트주파수의 대역폭과, 제외된 피크의 대역폭을 고려하여 일정한 조건을 만족하는 피크를 재 추출하여 비음의 특징을 적절하게 표현하는 수정된 알고리즘을 제안한다.

수정된 알고리즘은 분석 프레임에 나타난 피크 주파수와 바로 전의 프레임의 추정된 포만트주파수와의 차에 의해 분석 프레임의 포만트주파수를 추정하도록 한다. 그러므로, 트랙킹 시작 프레임의 포만트 주파수가 중요하다. 일반적으로 음성신호는 에너지가 최대의 프레임주위에서 정상적(stable)

인 포만트주파수 패턴을 갖기 때문에 본 알고리즘에서는 이 구간에 대한 다음과 같은 스텝1-스텝5의 시작 프레임에 있어서 포만트주파수의 추정이 기초정보가 된다. 수정된 포만트트랙킹의 알고리즘은 다음과 같다.

스텝 1 :

먼저 분석프레임에 대하여 20ms씩 대수 에너지를 구하여 에너지가 최대가 되는 프레임을 찾는다. 이때,

$$0 < \text{주파수} < 3400\ \text{Hz}$$
$$0 < \text{대역폭} < 500\ \text{Hz} \quad (1)$$

를 만족하는 주파수를 다음 (2)식과 같이 한다. 피크 주파수와 대역폭은 피크 피킹법을 사용하였다.

$$PF_{i,n}\ (i = 1, 2, \ldots\ldots, NP_n) \quad (2)$$

여기에서, NP_n은 식 (1)을 만족하는 n번째 프레임의 피크수이다. 또한, 대수 에너지가 양의 부분을 음성신호로 하여, 그 부분을 분석 대상으로 한다.

스텝 2 :

$NP_n = 3$ 이면, n번째 프레임을 시작 프레임으로 하여, 스텝 6으로 이동하고, 그 이외에는 j=1로 하여 스텝 3으로 이동한다.

스텝 3 :

n-j 번째 프레임의 피크수를 계산한다. $NP_{n-j} = 3$ 이면, n-j 번째 프레임을 시작 프레임으로 하여 스텝 6으로 이동하고, 그 외에는 스텝 4로 이동한다.

스텝 4 :

n+j 번째의 프레임의 피크수를 계산한다. $NP_{n+j} = 3$ 이면, n+j 번째 프레임을 시작 프레임으로 하여 스텝 6으로 이동하고, 그 외에는 스텝 5로 이동한다.

스텝 5 :

j = j + 1로 하여 스텝 3으로 이동한다.

스텝 6 :

구해진 시작 프레임을 n번째의 프레임으로 하여 포만트트랙킹을 행한다. 이것은 다음과 같은 조건을 만족하는 피크를 찾아서 스텝 7로 이동한다.

$$0 < PF_{i,n} \leq 3400\ \text{Hz}\ (i = 1, \ldots., NP_n)$$
$$0 < PB_{i,n} \leq 500\ \text{Hz} \quad (3)$$

여기에서, $PF_{i,n}$은 n 번째 프레임의 i번째의 주파수이고, $PB_{i,n}$은 n번째 프레임의 i번째 프레임의 대역폭이다.

스텝 7 :

스텝 6을 만족하는 피크가 3개 이상 존재하면, 스텝 8로 이동하고, 그 외에는 스텝 6에서 제외된 피크에 대하여 그 대역폭을 다음과 같이 계산한다.

$$DAk,m = | FF_{k,n-1} - PF_{m,n} |$$
$$(k = 1, 2, 3\ ,\ m = 1, 2) \quad (4)$$

여기에서 $DA_{k,m}$이 570 Hz보다 작거나 같고, 다음과 같은 식(5)의 조건을 만족하는 피크를 발견할 수 있으면, 스텝 8로 이동한다.

$$PB_{m,n} < 2FB_{k,n-1}$$
$$(k = 1, 2, 3\ \ ,\ m = 1, 2) \quad (5)$$

여기에서, $FP_{k,n-1}$은 n-1 번째의 프레임에서 추정된 k번째의 포만트주파수이고, $FB_{k,n-1}$은 n-1번째의 프레임에서 추정된 k번째의 포만트주파수의 대역폭이다.

스텝 8 :

스텝 7을 만족하는 피크수를 MNP라고 하여, MNP가 2보다 작거나 같으면 스텝 9로 이동하고,

MNP가 3보다 작거나 같으면 스텝 10으로 이동하고, MNP가 4보다 작거나 같으면 11로 이동한다.

스텝 9 :

$DB_{k,l} = | FF_{k,n-1} - PF_{l,n} |$

$(k = 1, 2, 3,\ l = 1, 2)$ (6)

를 계산하여

$FF_{k,n} = PF_{l,n}$; $min\ DB_{k,l}$ (7)

$FF_{k,n} = FF_{k,n-1}$ (8)

$FB_{k,n}$ = $FF_{k,n}$의 밴드 폭 (9)

으로 하고 스텝 12로 이동한다.

스텝 10 :

$FF_{k,n} = PF_{l,n}(k = 1, 2, 3,\ l = 1, 2)$ (10)

$FB_{k,n} = PB_{l,n}(k = 1, 2, 3,\ l = 1, 2)$ (11)

와 같은 3개의 피크를 n번째 프레임의 포만트 주파수로 하여 스텝 12로 이동한다.

스텝 11 :

$DB_{k,l} = | FF_{k,n-1} - PF_{l,n} |$

$(k = 1, 2, 3,\ l = 1, 2, ..., MNP)$ (12)

를 계산하여 다음 식에 대입한다. 그 다음에 스텝 12로 이동한다.

$FF_{k,n} = PF_{l,n}$; min $DB_{k,l}$ (13)

$FB_{k,n}$ = $FF_{k,n}$ 의 밴드폭 (14)

스텝 12 :

음성의 끝부분까지 또는 무성음이 나타날 때까지 n = n ∓ 1 로 하여 스텝 6으로 이동한다. 그 이외에는 스텝 13으로 이동한다. 여기에서 +는 후향이고, -는 전향을 의미한다.

스텝 13 :

포만트트랙킹에 의하여 추정된 포만트 주파수에 대하여, 다음과 같은 필터를 이용하여 최종적으로 부드러운 포만트주파수의 결과를 얻는다.

$FF_{k,n} = (FF_{k,n-1}) / 4 + (FF_{k,n}) / 2 + (FF_{k,n+1}) / 4$ (15)

그리고 n이 종단의 프레임이라면 다음 (16) 식과 같다.

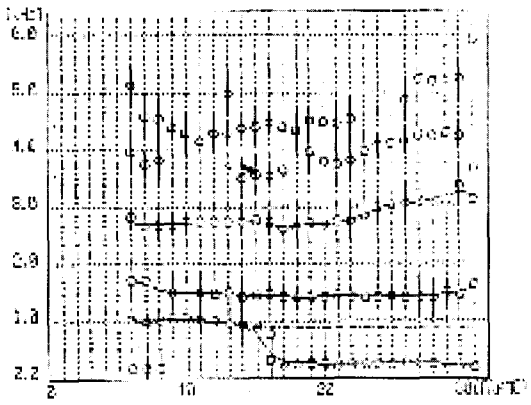
$FF_{k,n} = (FF_{k,n}) / 2 + (FF_{k,n} ∓ 1) / 2$ (16)

일반적인 포만트트랙킹 알고리즘의 경우[4,5,6], 임의의 n번째 프레임이 비음구간으로, 비음의 영향을 받아 제2, 제3의 피크 대역폭이 넓어져서 위의 스텝 6의 제한에 의해 제외되어 남은 2개의 피크에 대하여 스텝 9의 처리를 행하여, n-1 번째의 프레임의 포만트주파수와의 최단거리를 만족하는 피크를 n번째 프레임의 주파수로 하여 남은 1개를 n-1번째의 프레임으로부터 처리하고 있음을 알 수 있다. 그러나, 본 수정된 알고리즘에서는, 비음의 영향을 받아 대역폭이 넓어진 프레임에 대하여, 위의 스텝 7에서 n-1번째의 프레임의 포만트주파수와의 거리가 570 Hz이하일 경우, 그 대역폭을 계산하여 해당조건을 만족하는 피크를 다시 처리함으로 스텝 6에서 제외된 피크를 재추출 할 수 있다.

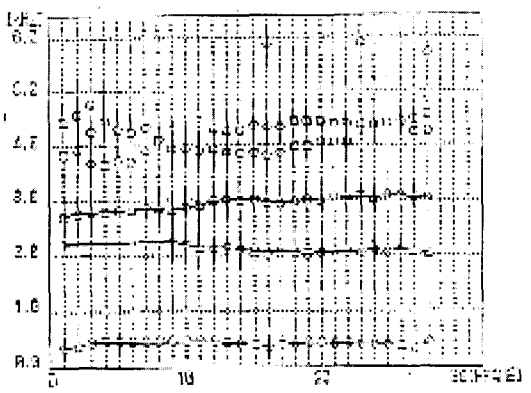
2.2.2 수정된 알고리즘의 평가

본 논문에서 제안한 수정된 포만트트랙킹 알고리즘을 평가하기 위하여, 한국어 숫자음성을 이용하였다. 그 결과를 그림 1에 보인다. 그림 1의 (a)는 숫자음성 '삼'에 대한 포만트트랙킹의 결과이다. 이 그림에서 점선은 종래의 알고리즘에 의한

트랙킹결과이고, 실선이 본 연구에서 수정한 알고리즘으로 얻은 트랙킹결과이다. 종래의 방법에서는, 16번째 프레임에서 나타난 4개의 피크중 제2피크는 대역폭의 제한에 의해 제외되고, 제1피크 역시 전의 프레임과의 변화량이 500Hz이하라고 하는 조건으로부터 제외되는 결과를 낳게되어 점선과 같은 트랙킹의 결과를 얻게 되었다. 따라서 비음과 같은 특징을 정확하게 추출 할 수가 없다. 그러나, 본 연구에서 제안한 수정된 알고리즘에서는, 제1피크를 재 추출하여 그 특징을 올바르게 트랙킹하게 되었다. (b)는 한국어 숫자음성 '일'에 대한 트랙킹결과이다. 이 경우, 모음구간의 제2, 제3포만트의 주파수가 2.1kHz, 2.7kHz의 영역에 분포



(a) 숫자음성 '삼'



(b) 숫자음성 '일'

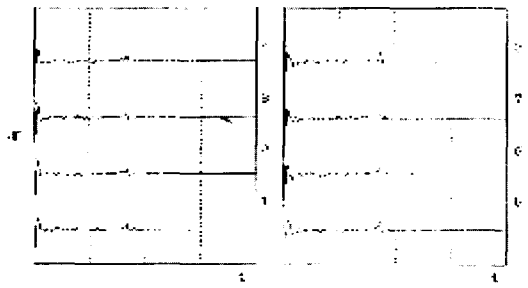
(그림 1) 포만트 트랙킹의 결과

하고 있어서, 모음으로 이어지는 비음의 영향에 의해 그 대역폭이 넓어지는 경향이 현저하게 나타난다. 따라서, 같은 프레임 내에 존재하는 대역폭의 제한만이 정확한 포만트 주파수를 추적해 나갈 수 있다고는 할 수 없으나, 본 연구에서 제안된 수정 알고리즘의 스텝 7에서 식(5)의 조건을 만족하는 피크가 선택되어, 올바르게 트랙킹이 일어남을 알 수 있다.

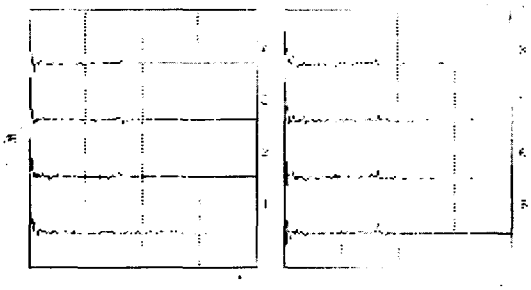
3. 어두부에 의한 분류

한국어의 숫자음성에는 어두부가 유성음 또는 무성음에 따라서 2개의 그룹으로 나눌 수 있다. 본 연구에서는 이러한 어두부에 의한 유성음/무성음의 판정으로 매칭 패턴을 2개의 그룹으로 나누어 인식을 행하기로 한다. 이와 같은 2개의 그룹으로 분류하는데는, 파형처리, 상관처리, 스펙트럼처리에 의한 분류법도 존재하지만, 여기에서는 스펙트럼처리의 켑스트럼피치(Capstrum Pitch), 대수 에너지(Log Energy)등을 이용하여 분류를 행한다. 숫자음성 '일'과 '삼'에 대한 피치 분석결과를 그림 2에 보였다. 대수 에너지가 양(Positive)인 부분을 음성구간으로 하여, 40ms의 해밍 윈도우를 통하여 20ms씩 이동시키면서 분석한 결과이다. 유성음이 시작되는 숫자음성 '일'은 첫 번째 프레임에서 피치가 나타나지만, 무성음으로 시작되는 숫자음성 '삼'의 경우에는, 두 번째 프레임에서 처음으로 피치가 나타나고 있음을 알 수 있다.

이렇게 하여, 각 숫자음성의 어두부에 대한 피치의 크기를 나타낸 그림이 그림 3이다. 피치의 크기는, 무성음으로 시작되는 경우가 0.2~0.8 사이에 분포하고 있다. 유성음으로 시작되는 경우가 1.0~2.6 사이에 분포하고 있음을 알 수 있다. 이러한 결과를 이용하여 임계값 0.9를 유성음이 시작되는 경계선으로 하고, 무성음이 끝나는 경계선으로 처리할 수 있음을 알 수 있다. 본 연구에서는 이러한 2개의 그룹으로 나누어 그 그룹 내에서 인식을 행하도록 하였다.

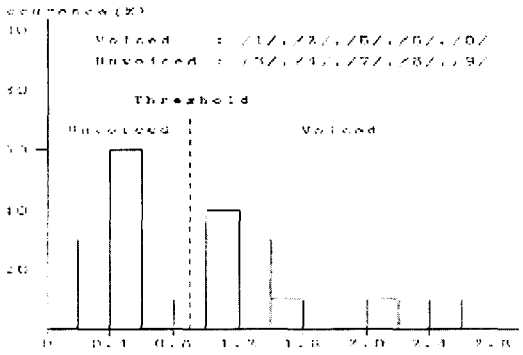


(a) 숫자음성 '일'



(b) 숫자음성 '삼'

(그림 2) 숫자음성 '일', '삼'의 피치 분석 결과



(그림 3) 숫자음성의 어두부에 대한 피치의 크기

4. 다이나믹 프로그래밍에 의한 패턴 매칭

음성인식을 하기 위하여, 표준음성 패턴과 인식해야 할 음성사이에 패턴 매칭을 하게 되는데, 화자의 속도에 대응하는 시간축의 정규화가 필요하다. 이것이 곧 다이나믹 프로그래밍으로, 비선형적인 시간축을 정규화 하여 이론적인 최적의 매칭법을 구하여, 패턴의 일부 상이한 부분이 매칭 전체에 미치는 악영향을 줄이고자 하는 방법을 사용한다. 다이나믹 프로그래밍은 다음과 같다.

표준 패턴의 특징 파라메타인 포만트 주파수 벡터 시계열은 다음 식 (17)로 놓는다.

$$A = (P_1, P_2, P_3, ..., P_i) \quad (17)$$

그리고, 입력 패턴의 특징 파라메타의 시계열을 식 (18)과 같이 놓는다.

$$B = (F_1, F_2, F_3, ... , F_j) \quad (18)$$

여기에서, i, j는 프레임수 이다.

그렇다면, 벡터 A 와 B 간의 거리는 다음과 같이 된다.

$$d(i, j) = \| A(P_i) - B(F_j) \| \quad (19)$$

단, $P_i = (P_{i1}, P_{i2}, P_{i3}, ..., P_{ik})^T$,

$F_j = (F_{j1}, F_{j2}, F_{j3}, ..., F_{jk})^T$

여기에서, 다음과 같은 식(20)의 점화식을 초기 조건으로 하여 순차적으로 G(i, j)를 계산한다.

$$G(1, 1) = d(1, 1) \quad (20)$$

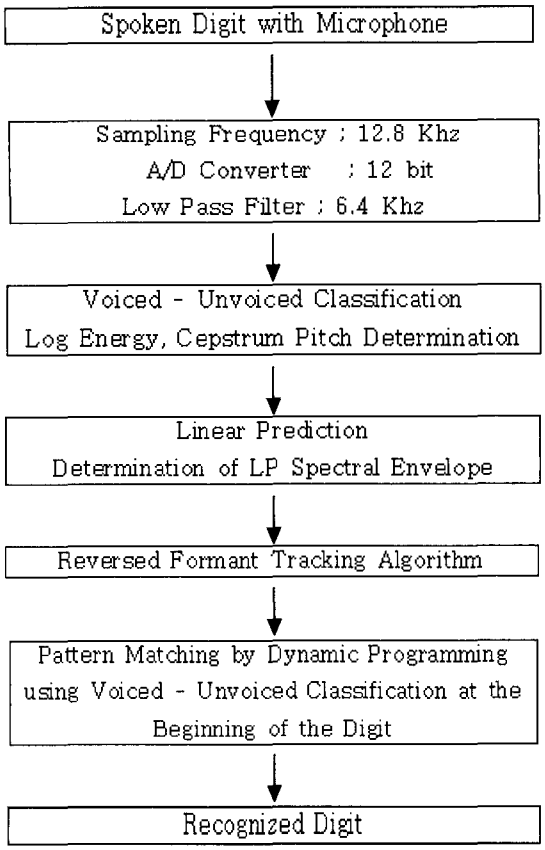
$$G(i, j) = Min \{ (d(i, j) + G(i-1, j)), (d(i, j) + G(i, j-1)), (2d(i, j) + G(i-1, j-1)) \} \quad (21)$$

위의 식(21)을 이용하여 G(i, j)를 계산하여 다음과 같은 유사도 S를 구하여 인식을 행한다.

$$S(A, B) = G(i, j) / (i + j + 1) \quad (22)$$

단, $| i - j | \leq r$, $i > 0$, $j > 0$, r = 정합창

정합창의 넓이를 정하는 파라메타 r의 값은 작을수록 정합의 능력이 저하되고, 클수록 오인식의 원인이 되기 때문에 적절한 값을 설정할 필요가 있다.



(그림 4) 음성인식 순서

(표 3) 인식실험 조건

샘플링 주파수 ; 12.8 Khz
샘플링 포인트 ; 4096 point
샘플링 시간 ; 0.078 ms
데이터 길이 ; 320 ms
프레임 길이 ; 10 ms

5. 인식실험

위에서 서술한 수정된 포만트트랙킹 알고리즘과 어두부에 의한 분류, 다이나믹프로그래밍을 이용한 한국어 숫자음성인식의 순서를 그림 4에 보인다. 그리고 인식실험 조건을 표 3에 보인다. 실험에서 사용된 숫자음성 데이터는, 성인 화자 5명이 0부터 9까지 6번씩 명확하게 발성한 숫자 데이터로 300개를 사용하였다. 그리고, 음성의 강도는 일정한 수준으로 정규화 하였다. 표준 패턴은 인식실험 데이터와는 별도로 남성화자 2명이 0부터 9까지 2번씩, 그리고 여성화자 1명이 0부터 9까지 한번씩 발성한 50개의 숫자음성으로부터 포만트 주파수를 구하여 평균화하여 작성하였다.

인식방법은, 먼저 마이크로부터 입력된 음성데이터에 대하여 샘플링주파수 12.8 kHz, 양자화 12bit, 저대역 필터(Low Pass Filter) 6.4 kHz의 전처리를 행하였다. 그 후에, 각프레임 마다 대수 에너지(Log Energy), 켑스트럼 피치(Capstrum Pitch)를 구하여 음성데이터의 시작점과 끝점을 결정하여 어두부에 의한 유성음/무성음의 판정을 행한다. 그 후에, 선형예측법에 따라서 각 프레임마다 스펙트럼 포락(Spectrum Envelope) 특성을 구하여 수정된 포만트 트랙킹 알고리즘을 이용하여 제1, 제2, 제3포만트 주파수를 구한다. 표준패턴과 입력된 패턴과의 매칭은 4장에서 서술한 다이나믹프로그래밍을 이용한다. 패턴매칭의 경우, 어두부에 의한 유성음과 무성음의 판정을 마친 다음 유성음인 경우에는 유성음으로 시작되는 숫자음성 '영', '일', '이', '오', '육'을 표준패턴과 매칭 시키고, 무성음인 경우에는 무성음으로 시작되는 숫자음성 '삼', '사', '칠', '팔', '구'를 표준패턴과 매칭 시킨다. 표준 패턴과 입력패턴과의 최단거리를 계산하여 거리가 가장 짧은 숫자음성을 인식된 결과로 한다.

인식결과는 표 4와 표 5, 표 6에 보이는 바와 같다. 표 4는 기존의 포만트 알고리즘[5]을 이용하여 인식된 결과를 나타내고 있다. 표 5는 수정된 알고리즘에 의한 인식결과이고, 표 5는 어두부에 의한 유성음/무성음에 의한 판정을 겸용한 경우의 인식결과이다. 오인식은 숫자음성 '삼', '사', '육'의 경우에 많이 나타났는데, '삼'의 경우, '삼'(sam)을 짧게 발성했을 때, 's'와 'm' 사이의 모음 'a'의 구간에 저주파수가 나타나는 원인으로 '팔'과 혼동되었기 때문이고, '사'(sa)의 경우에는, 제2포만트주파수의 피크가 나타나지 않았기 때문이며, 또한 '육'(juk)의 경우에는 발성시간이 너무 짧기 때문에 짧게 발성한 '영'(young)과 혼동되었기 때문으로 생각된다.

(표 4) 정(5)에 의한 인식결과

84.3%		입력데이터									
		1	2	3	4	5	6	7	8	9	0
인식결과	1	26	1								
	2	4	29								
	3			16	7		1				
	4			6	20						1
	5					25					
	6						21				
	7			2	1		1	30			
	8			6	1				30		
	9				1	5				29	
	0						7			1	29

(표 5) 수정된 알고리즘에 의한 인식결과

89%		입력데이터									
		1	2	3	4	5	6	7	8	9	0
인식결과	1	28									
	2	2	30								
	3			22	7		1				
	4			2	20						1
	5					25					
	6						21				
	7			1	1		1	30			
	8			5	1				30		
	9				1	5				29	
	0						7			1	29

(표 6) 수정된 알고리즘과 유/무성음 판정에 의한 인식결과

91%		입력데이터									
		1	2	3	4	5	6	7	8	9	0
인식결과	1	28									
	2	2	30								
	3			22	7						
	4			2	20						
	5					30					
	6						23				
	7			1	1			30			
	8			5	1				30		
	9				1					30	
	0						7				30

6. 결 론

본 논문에서는, 한국어 음성인식의 한가지 기초연구로서 한국어의 특징인 비음을 고려한 포만트트랙킹 알고리즘을 수정하여 새롭게 제안함과 동시에 숫자음성인식에 적용한 결과를 보이고 그 유효성을 확인하였다.

종래의 포만트트랙킹 알고리즘에서는 비음의 특징이 정확하게 추출되지 않았지만, 본 연구에서 제안한 수정된 포만트트랙킹 알고리즘에서는 정확하게 추출되어 숫자음성을 데이터로 하여 그 결과를 확인하였다.

또한, 어두부에 의한 숫자음성의 특징을 정보로 살려서 유성음/무성음의 판정을 행하여, 동일 그룹 내에서 매칭을 행하는 숫자음성 시스템을 확립하였다.

본 연구는 음성인식의 최종 목표인 불특정 화자 연속음성인식법에 접근하는데 사용될 수 있으리라 생각한다.

참 고 문 헌

[1] Borden, G.J., and Harris, K.S. "Speech Science Primer", 2nd edition, Williams & Wilkins, Baltimore., 1984.

[2] O'Connor, J.D. "Phonetics", Penguin, Middlesex, England, 1973.

[3] Dutoit, T., and Leich, H. "MBR-PSOLA ; Test-to-Speech Synthesis Based on an MBE Re-synthesis of the Segments Database," Speech Communication, 13, pp. 435~440, 1993.

[4] 鄭, 李, 城戶; 한국어의 단모음인식에 관한 고찰, 음향학회 강연논문집, 1986.

[5] 鄭, 牧野, 城戶; 한국어의 어두파열자음의 인식, 음향학회 강연논문집, 1989.

[6] 安居院, 中; 컴퓨터 음성처리, 1980.

[7] 般橋; 뉴럴네트워크에 의한 단어중의 모음인식 검토, 음향학회 강연 논문집, 1988.

[8] 이상호, 오영환, 서정연, "한국어 문서 음성 변환 시스템을 위한 문서 분석기", 한국음향학회지 제 15권 제 3호, pp. 50~59, 1996.
[9] 吳; A Simple Algorithm for Endpoints Determination and Voiced-Unvoiced-Silence Classification of Spoken Words, 정보과학회논문지 Vol. 2, No. 1, pp. 23~30, 1985.
[10] 김형근, 확률적 의존 문법과 한국어 구문 분석, 석사학위 논문, 한국과학기술원 전산학과, 1995.
[11] 강은영, 민소연, 배명진, "개선된 피치검출을 위한 스펙트럼 평탄화 기법에 관한 연구", 한국음향학회지 제21권 제3호 pp. 310~314, 2002.
[12] 정혜경, 김유진, 정재호, "켑스트럼으로부터 변환된 로그 스펙트럼을 이용한 포먼트 평활화 켑스트럴 평균차감법", 한국음향학회지 제21권 제4호 pp. 361~373, 2002.

◐ 저 자 소 개 ◑

송 정 영
1984년 한남대학교 컴퓨터공학과 졸업
1992년 와세다대학교 전기전자정보공학연구과 졸업(공학석사)
1995년 동대학 동연구과 박사과정 졸업(공학박사)
1995년~1997년 청운대학교 전자계산학과 전임강사
1997년~현재 : 배재대학교 공과대학 컴퓨터공학과 교수
관심분야 : 문자·음성·영상처리 및 분석, 프로그래밍언어 복잡도분석, 소프트웨어공학, Human-interface etc.
E-mail : jysong@mail.paichai.ac.kr