

음성 인식을 위한 소구간 경로 제약

The Local Path Constraint for the Recognition of Speech

안 태 옥*, 김 순 협*

(Tae Ock Ann, Soon Hyob Kim)

요 약

본 논문에서는 인식을 증진시키기 위해 새로운 소구간 경로 제약을 제안하였다. 파라메타로써 자동 상관 및 LPC 계수를 사용함으로써 입력 음성 신호를 분석하였다. 제안된 형의 소구간 경로제약이 기존의 5가지 형과 비교되었다.

본 연구에서 사용된 음성은 전철역명이고, 두명의 남성화자와 1명의 여성화자에 의한 서로 다른 13개의 단어를 10번 발음한 130개의 단어를 시험하였다. 그 결과로서 본 논문에서 제안한 형이 가장 좋은 형이라는 것을 알게 되었고 또한 이에 따른 인식율은 94.6%를 얻었다.

ABSTRACT

In this paper, a local path constraint is proposed in order to increase the speech recognition rate. An input speech signal is analyzed by autocorrelation and LPC coefficient as parameters. The local path constraint of the proposed type was compared with the conventional five types.

The speeches used in this search are the subway stops, and the 130 words pronounced 10 times for the different 13 words consisting of 11 characters of syllable by 2 male and 1 female are tested. As a result, we proved that this proposed type is the most optimal type and the recognition rate of 94.6% is obtained.

1. 서 론

인간어 통신을 하는 가장 우선적인 방법은 음성이다. 또한 인간만이 초보 단계를 벗어나 정보를 부호화하고, 전달 할 수 있는 음성 기관을 발달 시켜왔다. 인간과 컴퓨터 간의 정보 교환을 위한 음성 인식의 과정에서 음성은 화자의 발음하는 속도 차이

로 인하여 음성 패턴에서 시간축의 비선형적 변동을 일으킨다. 이러한 패턴의 길이에 차이에 의한 영향을 흡수하기 위한 방법으로 다양한 DTW 알고리즘(Dynamic Time Warping Algorithm)을 이용하여 효과적으로 수행할 수 있다.

따라서, 한국어 인식을 위한 DTW 알고리즘 수행 시 인식율과 밀접한 관계가 있는 소구간 경로 제약 방법의 선택을 위해 기존의 방법들을 실험 비교하

*광운대학교 전자계산기공학과

고, 기존의 방법을 응용한 새로운 소구간 경로 제약을 제시하였다.

본 연구에서는 서울 시내에 있는 전철명에 대한 단독어 인식을 위해 기존의 소구간 경로 제약 및 본 연구에서 제안한 소구간 경로 제약을 DTW 알고리즘에 적용시켜 인식율을 비교하였다.

II. 제안된 소구간 경로 제약

1. 기존의 소구간 경로 제약

가로축을 m 이라 하고, 세로축을 n 이라 할때 그림 1에서 A형은 그리드 점 (m,n) 에 이르는 타당한 경로가 점 $(m-1,n-2)$, 점 $(m-1, n-1)$ 또는, 점 $(m-2,n-1)$ 에서 부터 시작되는데 점 $(m-1,n-2)$ 에서는 점 $(m,n-1)$ 의 매개점을 통하여 점 (m,n) 에 도달하고, 점 $(m-2,n-1)$ 에서는 점 $(m-1,n)$ 의 매개점을 통하여 점 (m,n) 에 도달한다.

B형은 같은 그리드 점 $(m-1,n-2)$, 점 $(m-1,n-1)$ 또는 점 $(m-2,n-1)$ 에서 시작하지만 매개점을 통하지 않고 직접 점 (m,n) 에 도달한다.

C형은 점 $(m-1,n-1)$ 과 점 $(m-1,n-2)$ 에서는 직접 점 (m,n) 에 이르고, 점 $(m-2,n-2)$ 와 점 $(m-2,n-1)$ 에서 시작하는 경로는 점 $(m-1,n)$ 를 매개점으로하여 점 (m,n) 에 도달한다.

D형은 점 (m,n) 에 도달하는 경로가 2개의 매개점을 포함하여 9가지의 경로를 가지며, E형은 Itakura 방식이라하는데, 1프레임 어상 같은 축을 지나지 못하도록 경로 제약을 한 형이다.

2. 제안된 소구간 경로 제약

본 연구에서는 3프레임 단위로 거리 계산을 하는 D형이 인식 실험의 과정에서 많은 계산량을 필요로 하는 것을 1프레임 단위로 만듬으로써 개선하고 또한 각 화자의 발음하는 속도 차이로 인한 음성 패턴에서 시간축의 비선형적인 변동을 융통성 있게 보완해 주는 의미에서 그림 2와 같이 1프레임 단위로 계산을 하는 새로운 형을 제안하였고, 참고로 인식 실험 과정에서 새로운 형 제안을 위한 중간 단계로 1프레임 단위의 그림 3과 2프레임 계산 단위의 그림 4의 방법으로도 실험을 하였다.

형	소구간 경로 제약	최대프레임 수
A		2 프레임
B		2 프레임
C		2 프레임
D		3 프레임
E		2 프레임

그림 1. 기존의 소구간 경로 제약
The variations of the local path constraint.

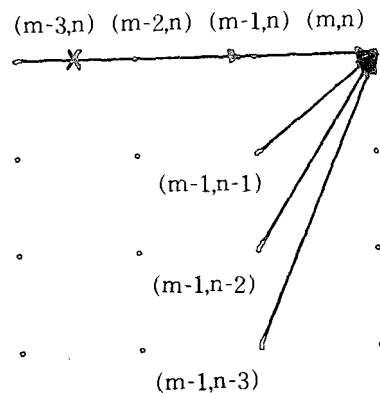


그림 2. 제안된 형의 소구간 경로 제약
Local path constraint of the proposed type

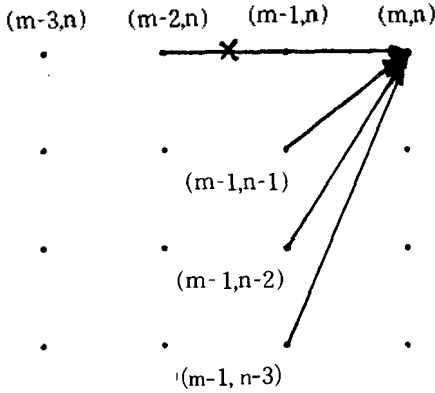


그림 3. 1프레임 계산 단위의 소구간 경로 제약
Local path constraint of 1 frame calculation unit

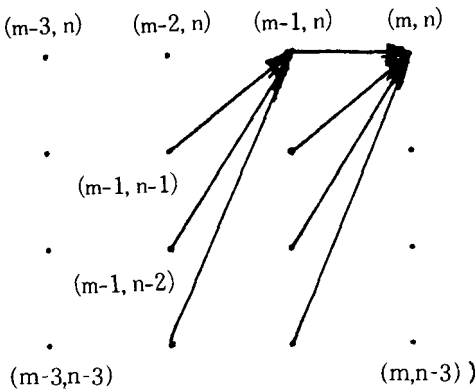


그림 4. 2프레임 계산 단위의 소구간 경로 제약
Local path constraint of 2 frames calculation unit

그 중에서 제안된 새로운 형의 누적된 거리 $D(m,n)$ 은 다음 식으로 표현된다.

$$D(m,n) = \min \begin{cases} D(m-1,n-3) + d(m,n), \\ D(m-1,n-2) + d(m,n), \\ D(m-1,n-1) + d(m,n), \\ D(m-1,n) \cdot g(k) + d(m,n) \end{cases}$$

여기서

$$g(k) = \begin{cases} \infty, & j(k-1) \neq j(k-2) \neq j(k-3) \\ 1, & \text{그 외의 경우} \end{cases}$$

III. 인식 실험 및 결과

인식 시스템을 통해 성인 남성 2인(각자 13단어를 4번씩 반복)과 여성 1인(13단어를 2번씩 반복)이 발음한 130개의 단어를 대상으로 DTW 알고리즘을 적용시켜 실험한 결과는 다음과 같다.

1. 인식 시스템 구성

녹음된 음성은 200-3200Hz의 대역폭을 가진 대역 필터를 통과시킨 후 sampling 주파수 6.67KHz, 12bit로 A/D 변환하였다. 분석 프레임 길이는 128 sample(19.2 msec)로 하였다. 인식에 관한 시스템 구성도는 그림 5과 같다.

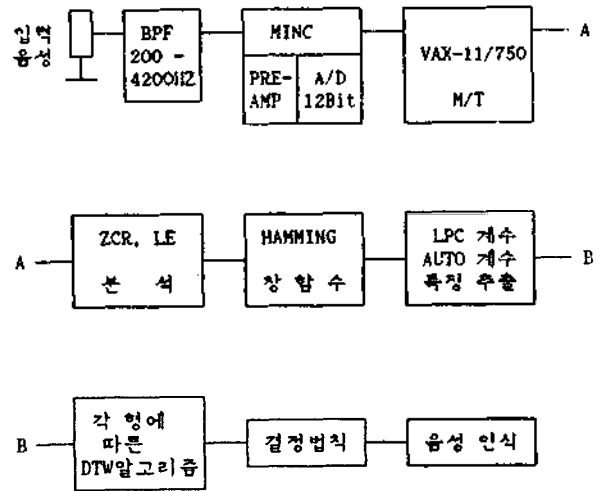


그림 5. 인식에 관한 시스템 구성도
Block diagram for the recognition

2. 거리 측정법

본 연구에서는 화자 종속 방법으로 음성 인식 실험을 하였으며, 단독어를 DTW 알고리즘을 이용하여 미리 설정된 표준 패턴과 시험 패턴간의 거리 측정은 하였는데, Itakura가 제안한 LPC 대수 확률 거리 측정법을 사용하였다.

LPC 대수 확률에 의한 거리 측정법을 다음과 같다.

$$d = \log \frac{\alpha_R \cdot V_T \cdot \alpha_R^t}{\alpha_T \cdot V_T \cdot \alpha_T^t}$$

여기서 α_R 과 α_T 는 표준 패턴과 시험 패턴의 선형 예측 계수 벡터들이고 V_T 는 시험 패턴의 자동 상관 계수 행렬이다.

3. 실험 결과 및 고찰

기존의 방법에 의한 소구간 경로 제약과 제안된 방법 및 제안된 방법의 과도기적인 방법에 의한 소구간 경로 제약을 실험하였으며 최종 인식 결정에서 시험 되어진 단어수와 인식 되어진 단어수의 비로 인식율을 나타내었다.

제안된 소구간 경로 제약에 의한 인식 실험에서 발생한 에러를 중심으로 살펴 보면, 신촌이 신사로, 신당이 신사로, 신대방이 신당으로, 대림이 대방으로, 이대가 이수로, 그리고 신당이 신촌으로 잘못 인식 되었으며, 다만 수유만 첫자가 같지 않은 신촌으로 인식됨을 알았다. 그러나 여기에서도 음소에 있어서는 s자로 같은 음소로 시작되었다.

표 1. 각 형별 인식률
Recognition rate of each type

소구간경로	웨이팅함수	인식된 단어수 총단어 수	인식률(%)
A	117	130	90.0
	130		
B	121	130	93.1
	130		
C	114	130	87.7
	130		
D	118	130	90.8
	130		
E	120	130	92.3
	130		
그림 3의 형	117	130	90.0
	130		
그림 4의 형	115	130	88.5
	130		
제안된 형	123	130	94.6
	130		

각 형별 인식률은 표1에 나타내었다. 또한, 각 개인별 각 단어에 대한 인식은 표 2와 같다.

표 2. 인식 결과
Recognition result.

1) 남성 화자 A의 인식 결과

입력 출력	이 대	이 촌	이 수	옥 수	수 유	신 촌	신 당	신 사	사 당	대 림	신 림	신 대방	대 방
이 대	4												
이 촌		4											
이 수			4										
옥 수				4									
수 유					4								
신 촌						4							
신 당							4						
신 사						1	4						
사 당								4					
대 림									4				
신 림										4			
신대방											4		
대 방												4	

2) 남성 화자 B의 인식 결과

입력 출력	이 대	이 촌	이 수	옥 수	수 유	신 촌	신 당	신 사	사 당	대 림	신 림	신 대방	대 방
이 대	4												
이 촌		4											
이 수			4										
옥 수				4									
수 유					3							1	
신 촌						1	4						
신 당							3						
신 사								1	4				
사 당									4				
대 림										3			
신 림											4		
신대방												3	
대 방										1			4

3) 여성화자의 인식 결과

입력 출력	이 대	이 촌	이 수	옥 수	수 유	신 촌	신 당	신 사	사 당	대 림	신 림	신 대방	대 방
이 대	1												
이 촌		2											
이 수	1		2										
옥 수				2									
수 유					2								
신 촌						2	1						
신 당							1						
신 사								2					
사 당									2				
대 림										2			
신 림											2		
신대방												2	
대 방													2

IV. 결 론

본 논문에서는 한국어 인식을 위해 최적의 소구간 경로 제약을 찾기 위한 인식 실험을 하였다. 대상어로서는 13개의 전철명을 선택하였으며 실험에서는 남성 2인(각각 4번씩 발음)과 여성 1인(2번 발음)이 발음한 130단어를 인식시켰다.

본 연구에서는 한국어 전철명에 대해 DTW를 이용하여 기존의 여러 소구간 경로 제약에 대해 인식 실험을 행한 뒤 본 연구에서는 새로운 최적의 소구간 경로 제약을 제안하였다. 인식 실험 결과 기존의 방법들에 비해 본 연구에서의 소구간 경로 제약이 인식이 가장 높은 94.6%임을 알았다.

참 고 문 헌

1. Hiroaki Sakoe and Seibi Chiba, "Dynamic Programming algorithm Optimization for spoken Word Recognition," IEEE Trans. on Acoustics, Speech, Signal Processing, Vol. ASSP-26, No.1, pp. 43-49, Feb. 1978.

2. Michael K. Brown and Lawrence R. Rabiner, "An Adaptive, Ordered, Graph Search Technique for Dynamic Time Warping for Isolated Word Recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-30, No.4, pp. 535-544, August 1982.

3. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoustics, Speech, signal Processing, Vol. ASSP-23, pp. 67-72, Feb. 1975.

4. Myers, L. R. Rabiner and A. E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing Vol. ASSP-28, No. 6, pp.623-634, Dec. 1980.

5. L. R. Rabiner, "Note on some Factors Affecting Performance of Dynamic Time Warping Algorithms for Isolated Word Recognition," The Bell System Technical Journal, Vol.61, No.3, pp. 363-373, Mar. 1982.

6. L.R. Rabiner and S. E. Levinson, "Isolated and Connected Word Recognition-Theory and Selected Applications," IEEE Trans. on communication, Vol. COM-29, No. 5, pp.621-639, May 1981.

7. 홍건우, 김순협, "한국어 단독 숫자음 인식을 위한 DTW 알고리즘의 비교", 한국음향학회 Vol.3, No.1, pp25~35, 1984.

▲안 태 옥

1953년 6월 24일생

1981년 2월 : 울산공과대학재
로과 졸업

1987년 2월 광운대학교전자
계산기공학과 졸업

.989년 현재 : 광운대학교전자
계산기공학과박사과정



▲김순협 : 8 권 3 호 참조