

유사지명 인식시의 성능 개선 연구

백승권, 양희식, 한민수
한국정보통신대학원대학교

A study on the improvement of speech recognition for similar place names

Seung-kwon Beack, Hee-Sik Yang, Minsoo Hahn
Information and Communications University
skbeack@icu.ac.kr

요 약

본 연구에서는 DAB(Digital Audio Broadcasting) 시스템의 교통정보 검색 서비스를 위하여 경부선 및 호남선의 톨게이트가 위치한 49 개의 지명을 대상으로 이를 인식하고자 할 때 인식율을 개선하였다. 지명 어휘의 특성을 분석한 결과 전체 지명의 81.6%가 2 음절이었으며 동일한 음절을 포함하는 지명이 전체의 구성된 어휘가 61%로 조사 되었다. 시스템에서 인식율을 개선하기 위하여 인식 대상어휘를 3 개의 set 로 재분류하고 인식 대상 어휘로 판정된 후보 어휘에 대하여 인식 성공여부에 핵심이 되는 음절의 위치에 따라 가중치 윈도우를 적용하였다. 그 결과 화자 독립의 인식율 테스트에서 남성의 경우 7.2%, 여성의 경우 5.1%의 인식율 향상을 보였다.

I. 서론

일반적으로 음성인식 시스템에는 벡터 양자화

(Vector Quantization)를 이용하는 방법과 동적 시간 정합(DTW)을 이용하는 방법, 신경회로망(Neural Network)을 이용하는 방법, 은닉 마코프 모델(Hidden Markov Model, HMM)을 이용하는 방법 등이 사용되고 있다[1][2][3][4].

본 연구는 DAB 시스템에 사용될 음성인식 기술의 인식율 개선에 관한 연구로 인식 대상어휘가 경부선 호남선의 톨게이트가 위치한 49 개의 지명으로 한정된다. 따라서 하드웨어 구현이 보다 용이하며 고립단어 인식에 좋은 성능을 보이는 동적 시간정합 인식 기술을 채택하였다. 동적 시간정합 알고리즘은 음소적 특성을 나타내는 특징 벡터들 간의 정합거리를 구하고 최소 누적에러를 갖는 대상 어휘를 인식된 어휘로 판별하므로 음소적 특성이 유사한 대상어휘들 간의 인식율은 저하됨이 당연하다. 그러므로 서로 유사한 음소 및 음절을 갖는 지명어휘를 인식 대상으로 할 때 오인식이 빈번하게 발생한다. 이에 DAB 용 고립

단어 인식기의 인식율 향상을 위해 각 지명의 음운 특성을 분석 하고 지명들을 음절 특성에 따라 3 개의 set 으로 분류하였으며 각 set 에 적합한 가중치 윈도우를 적용하였다.

본 논문의 구성은 다음과 같다. 2 절에서는 인식 대상 지명의 음소 특성을 분석하였고 3 절에서는 제안된 가중치 윈도우 적용방법에 대해 설명하였다. 4 절은 실험 결과에 대해서 기술하였으며 마지막 절에서 본 연구에 대한 결론을 맺었다.

II. DB(Data Base) 분석

DAB 에서 인식 대상으로 하는 고립단어는 경부선 및 호남선의 틀게이트가 위치한 49 개의 지명을 대상으로 하고 있다. 각 지명의 음절 특성을 살펴보면 40 개의 단어가 2 음절로 구성되어 있으며 이는 전체 DB 의 81.6%를 차지한다. 또한 이들 중 동일한 음절을 포함하는 단어, 예를 들면 “부산”, “익산”, “경산”,은 모두 22 개였으며 전체 인식 대상 어휘의 61%나 되었다. 표 1 은 인식대상어휘들간의 음소적 특성에 따라 분류한 것이다. 이는 전체 대상 어휘중 2 음절단어에 한하여 조사하였으며 방법은 각 음절의 유성음 무성음의 구성순서의 패턴에 따라 조사한 것이다. 이러한 조사를 하게된 이유는 각 단어의 음절의 초성이나 종성이 각기 다른 유, 무성음으로 구성되어 있다 할 지라도 음소 특성 패턴이 유사할 경우 구분어 힘들기 때문이다. 예를 들어 두 입술 사이에서 나는 소리인 “ㅂ”, “ㅍ”, 혀끝과 윗잇몸이 만나서 소리가 나는 “ㄷ”, “ㅌ”, 경구개와 혀바닥 사이에서 나는 소리인 “ㅈ”, “ㅊ” 그리고 연구개에서 나는 소리인 “ㄱ”, “ㅋ”, 같은 자음과 조음위치가 같고 울림소리인 “ㄴ”, “ㄹ”인 경우는 그 구별이 매우 어렵다[5].

표 1. 음소패턴에 따른 빈도수

패턴 \ 음절	1 음절	2 음절
U + V	18	17
U + V + U	11	15
V + U	4	5
V	7	3

III. 가중치 윈도우를 이용한 지명 인식방법

앞서 언급한 바와 같이 DAB 시스템의 교통정보 검색 서비스 구현을 위한 인식 대상 어휘들의 특성을 보면 대부분의 어휘가 2 음절로 구성되어 있으며 각음절의 음소 특성이 상당히 유사한 경우가 많음을 알 수 있다.

2 음절의 지명 중 마지막 음절이 일치하는 경우가 40 개의 어휘 중 21 개로 이러한 경우 1 음절의 인식 성공율에 따라 각 지명을 인식할 수 있게 된다. 마찬가지로 첫번째 음절이 일치할 경우, 두번째 음절의 인식 성공율에 따라 올바른 지명 인식이 이루어진다. 따라서 주어진 인식 대상어휘를 3 가지 경우로 나누어 생각할 수 있는데, 첫 음절이 일치하는 경우, 두번째 음절이 일치하는 경우, 두 음절 모두 일치하지 않는 경우 등으로 나누어 볼 수 있다. 이를 유사 음절의 위치에 따라 3 개의 set 으로 나누어 표 2 에 나타내었다. 3 음절 이상의 지명 중에서는 “대전”, “서대전”의 경우와 같이 2 음절 이상 일치하는 지명 들에 한해서 포함시켰다. 표 2 에서 나눈 각각의 set 에 대하여 가중치 윈도우를 적용하게 된다. 이때 가중치 윈도우의 형태는 인식대상어휘의 중요 음절 부분에 대해 가중치를 부여할 수 있는 형태로 모델링 하였다. 예를 들어 set1 의 경우는 첫째 음절(어두부분)에 가중치를 부여하고, set2 의 경우는 둘째 음절에 가중치를 부여 할 수 있는 윈도우 형태를 지녀야 한다. 그림 1 은 모델링된 가중치 윈도우의 형태를 나타내고 있다.

$$U_i = u_i \cdot \Delta w_k \cdot i \quad k = 1, 2, 3 \quad (\text{식 1})$$

$$V_i = v_i \cdot \Delta w_k \cdot i$$

$$\Delta w_1 = A_1 - \frac{A_1}{L}, \Delta w_2 = 1 + \frac{A_2}{L}, \Delta w_3 = A_3$$

$$D = \sum_{i=0}^L \sum_{n=0}^P (U_i(n) - V_i(n))^2 \quad (\text{식 2})$$

L = 전체 프레임 수, i = 현재 프레임,
 u_i = 입력 특징벡터, v_i = 템플릿 특징벡터
 D = 누적에러

표 2. 유사 음절의 위치에 따른 구분

set1		Set2		set3
동일 음절	지명	동일 음절	지명	지명
~전	대전	경~	경산	서울
	서대전		경주	양재
~원	수원	청~	청주	판교
	청원		청원	신갈
~산	오산	영~	영천	기흥
	경산		영동	남이
	양산			독립기념관
	부산			신탄진
	논산			회덕
	익산			금강
	광산			황간
	청주			구미
~주	경주		왜관	
	전주		금호	
	서광주		언양	
	목천		구서	
~천	옥천		삼례	
	김천		금산사	
	영천		태인	
	안성		정읍	
~성	장성		백양사	
	유성			
	북대구			
~구	동대구			

제안된 가중치 윈도우를 이용하여 발생된 지명 어휘를 인식하는 과정을 그림 2의 블록 다이어그램으로 나타내었다. 먼저 입력 어휘에 대해 특징벡터(LPC cepstrum)를 추출하고 추출된 특징벡터와 훈련된 자료(template)의 특징벡터 사이의

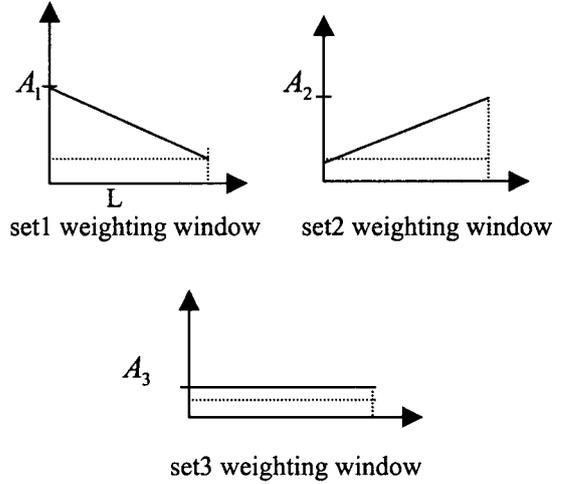


그림 1. 각 set의 가중치 윈도우

표 3. 가중치 윈도우 선정방법

후보자 어휘의 타입			선택윈도 타입
Set1	Set2	Set3	
0	0	0	Set1
0	0	-	Set1
0	-	0	Set1
-	0	0	Set2

거리를 동적 시간정합 알고리즘을 이용하여 정합 거리를 구하게 된다. 이때 최소 누적 에러를 갖는 후보 3 개를 선정하게 된다. 선정된 3 개의 후보들의 인덱스 정보를 이용하여 어느 set에 속하는 지명인가를 판별하고 각 set에 맞는 가중치 윈도우를 적용 한다. 가중치 윈도우는 입력 어휘와 선정된 후보자의 추출된 특징벡터에 적용되며 이는 식 1과 식 2와 같다. 이 때 선정된 후보자 3 개로부터 가중치 윈도우를 선택하는 방법은 표 3과 같은 룰에 의해 선택되어진다. 이는 대개의 경우 선정된 3 개의 후보자들은 같은 set에 속하게 되는데 그렇지 않을 경우 가중치 윈도우 타입을 결정하기 위해서 이다. set1의 가중치 윈도우를 우선 순위로 삼은 이유는 가장 높은 오인식 율을 발생시키기 때문이다. 식 2로부터 구해진 특징벡터들은 다시 한번 동적 시간정합 알고리즘에

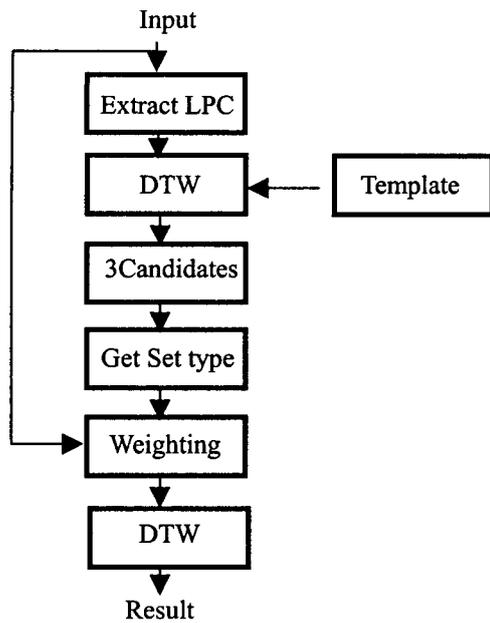


그림 2. 유사지명 인식 과정 블록다이어그램

의해 최소 누적에러를 갖는 후보자 하나를 선정하게 되고 이를 인식결과로 나타낸다

IV 실험 결과

전체적인 인식율을 조사해 보기에 앞서 제안된 가중치 윈도우를 적용하여 특징벡터들 간의 거리를 측정하는 방법이 타당한가를 조사해 보았다. 먼저 인식 대상인 지명 중 set1에 속하는 임의의 단어 14 개를 발성하여 순수하게 동적 시간 정합 알고리즘만을 이용한 기존의 시스템의 인식기와 가중치 윈도우를 적용한 인식기를 이용하여 인식 테스트를 하였다. 이 때 기존의 시스템에서의 누적에러와 가중치 윈도우를 적용하였을 때의 누적에러의 변화를 살펴 보았다. 그림 3은 기존 시스템의 인식기에서 가장 최소 누적에러를 갖는 후보자 3 개에 대해서 누적에러 값들을 나타낸 것이다. C1, C2, C3로 분류한 기준은 후보자의 인덱스 정보이용 하여 지명을 살펴보았을 때 올바른 인식을 하기위한 지명을 나타낸 후보자를 C1으로

나타내었으며 C2, C3는 임의로 정하였다. 기존의 시스템에서의 인식 방법은 최소 누적에러를 가지는 경우를 인식 결과로 삼게 된다. 따라서 그림 3을 살펴보면 C1이 선택되는 경우는 4번째의 경우를 제외하고는 다른 후보가 선택된다. 결국 set1의 발성어휘 14개중 인식된 지명은 1개였다. 그림 4는 가중치 윈도우를 사용했을 경우의 누적에러의 변화를 나타내고 있다. 이 때 발생되어진 14개 단어는 그림 3에서와 동일한 데이터를 사용하였다. 그림 4에서 알수 있듯이 14가지 발성에 대해 C1이 선택되어지는 경우는 8번째와 13번째 14번째를 제외한 모든 경우에 선택된다. 이는 가중치 윈도우를 적용하였을 때 발생패턴과 후보자들 간에 높은 유사도를 지니는 경우에 대해서 누적에러가 상대적으로 감소되기 때문으로 가장 높은 유사도를 지니는 후보자가 최소 누적에러를 갖는 된다.

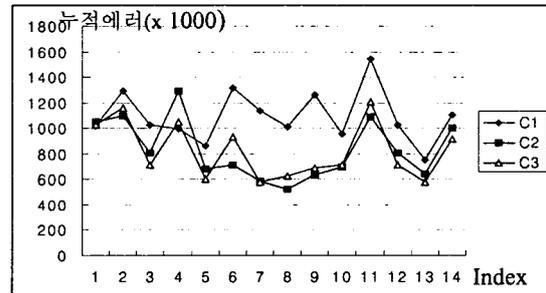


그림 3. 가중치 적용 전 누적에러

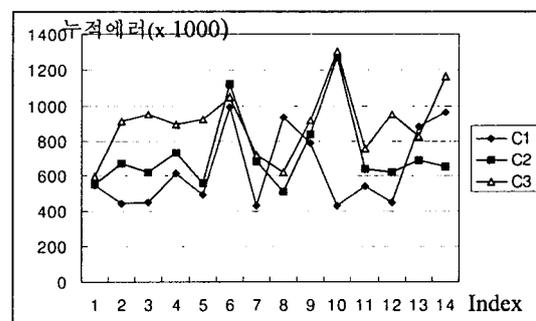


그림 4. 가중치 적용 후 누적에러

가중치 윈도우를 적용한다는 것은 이러한 유사도를 측정할 경우 어느 부분에 가중치를 두어 유사도를 측정하겠는가를 결정하는 것이다

인식율 조사는 20 대 중반의 성인 남녀 2 인을 대상으로 화자 독립으로 실험하였다. 인식율 조사 결과 본 연구에서 제안된 알고리즘이 기존 시스템에서 보다 여성의 경우 5.1%, 남성의 경우 7.2%의 향상된 인식율을 보였다. 그림 5 에 그 결과를 나타내었다

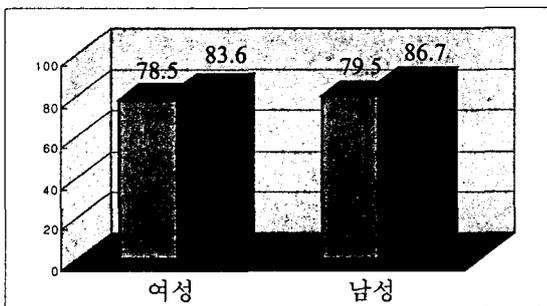


그림 5. 최종 인식율 조사 결과

V. 결론

본 연구는 유사지명인식시의 오인식율을 줄이고 자 음절의 음운 특성에 따라 가중치 윈도우를 적용하여 weighted-distance 개념을 도입하였다. 그 결과 여성의 경우 5.1%, 남성의 경우 7.2% 인식율이 향상되었다. 앞으로 90%이상의 인식율을 목표로 하고 있으며 이를 위해 보다 적합한 가중치 윈도우 형태를 연구 해야한다. 본 연구는 인식 대상 어휘가 지명으로 한정되어 있는 DAB 시스템의 인식 알고리즘으로써 달리는 차량의 소음 문제도 고려하여 인식율 저하를 막기 위해 잡음제거 기술에 대한 연구도 지속되어야 할 것이다.

참고문헌

- [1] W. Koenig, "A new frequency scale for acoustic measurements", Bell Telephone Lab, Record, Vol.27, PP.299-301,1949.
- [2] J.D. Markel and A.H. Gray, Jr.,*Linear Prediction of Speech*, Springer-Verlag, New York,1976.
- [3] A. M. Kondoz, *Digital Speech, Coding for Low Bit Rate Communications Systems*, JOHN WILEY & SONS, 1994.
- [4] S. Furui, *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker, 1992.
- [5] 윤철중 외, "음성인식을 이용한 차량 번호 조회 시스템의 구현" 한국음향학회 학술발표대회 논문집, 제 19 권,1 호, pp127-130,2000.