

최대 흐름 정합을 이용한 실시간 음소인식 시스템 구현

이상엽*, 박성원**

삼육대학교 경영정보학과*, 한국콘텐츠진흥원 산업분석팀**

zikimi@syu.ac.kr, seongwon@kocca.kr

Real-time Phoneme Recognition System Using Max Flow Matching

Sangyeob Lee*, Seongwon Park**

Dept. of Management Information System, Sahmyook University*,

Industry Research Team, Korea Creative Content Agency**

요 약

스마트 디바이스를 이용한 다양한 게임들이 증가하고 있는데, 음소인식은 스마트 디바이스를 사용한 효율적인 입력 방법은 음성이 될 수 있다. 게임에서 음성인식은 매우 빠르게 인식되면서 구동 되어야하는데, 본 연구에서는 게임 분야에서 유용하게 활용할 수 있는 최적화된 음소 인식 방법을 개발하였다. 본 논문에서 제안하는 음소 인식 방법은 음성 파장을 FFT로 전환하고, 해당 값을 Z평면에 도식한 후, 영역 데이터를 추출한 후 데이터베이스에 저장한다. 그리고 해당 값을 가중치 있는 두 갈래 그래프 최대 흐름 정합을 사용하여 음소 인식을 한다. 제안된 방법은 게임 또는 로봇과 같은 분야에서 빠른 음소 인식을 하고자 할 때 매우 유용한 방법이다.

ABSTRACT

There are many of games using smart devices. Voice recognition is can be useful way for input. In the game, voice have to be quickly recognized, at the same time it have to be manipulated promptly as well. In this study, we developed the optimized real-time phoneme recognition using max flow matching that it can be efficiently used in the game field. Firstly, voice wavelength is transformed to FFT, secondly, transformed value is made by a graph in Z plane, thirdly, data is extracted in specific area, and then data is saved in database. After all the value is recognized using weighted bipartite max flow matching. This way would be useful method in game or robot field when researchers hope to recognize the fast voice recognition.

Keywords : Weighted Bipartite Max Flow Matching: WBM(두 갈래 그래프 최대흐름 정합), Phoneme Recognition (음소인식)

접수일자 : 2012년 01월 30일 일차수정 : 2012년 02월 14일 심사완료 : 2012년 02월 18일

교신저자(Corresponding Author) : 이상엽

1. 서론

정보 시스템 기술의 발달에 따라, 입력 및 출력 장치에도 다양한 기술이 적용되고 있다. 과거에는 입력 장치라 하면 키보드, 마우스, 터치패드 등을 일반적으로 떠올렸으나, 그러나 요즈음의 진보된 시스템에서는 영상 인식, 음소 인식 등의 기술이 입력 장치에 탑재되는 경우가 많아지고 있다[1]. 특히 음소 인식은 많은 분야에서 다양하게 사용되고 있다[2]. 이전의 음소 인식은 그래프 체인을 이용한 방법을 주로 사용하였다. 특히 히든 마르코프 모델(Hidden Markov Model: HMM)을 이용한 음소 인식은 많은 시스템에 탑재되어 운용되고 있다[3]. HMM은 음소인식의 정확도가 매우 높아서 정확한 음소 인식에는 현재 나와 있는 알고리즘 중에서 최고의 성능을 가지고 있다 판단된다. HMM은 음성의 시작점을 정확하게 인식하고 그 이후부터 음소로 분석하는 단점이 있다. 음성의 시작을 정확하게 인지하지 못했을 경우 에러율이 매우 크다. 따라서 음성 입력을 하기 전에 시작 신호를 시스템에서 제공해야 한다[4].

스마트 디바이스를 이용한 다양한 게임이 개발되고 있다. 스마트 디바이스의 입력 장치로는 주로 터치패드가 많이 사용되고 있는데, 그 외의 입력 장치로는 자이로 센서와 마이크가 많이 사용된다. 스마트 디바이스를 이용한 게임을 할 경우, 마이크를 사용한다면 명령을 매우 빠르게 입력 받을 수 있어서 게임을 할 때 유용하게 사용될 수 있다. 그러나 HMM은 음성의 시작 점을 잡아서 분석 하는 방법이기 때문에 게임에 사용하기 힘들다. 또한 음성을 분석하기 위하여 그래프 체인을 따라 가기 때문에 연산 시간이 길어 바로 게임에 적용하기 힘들다[5].

본 연구에서는 게임과 로봇 분야에서 유용하게 사용될 수 있는 음소 인식 방법을 개발 하였다. 본 논문에서 제안하는 음소 인식 방법은 주파수 정보를 두 갈래 그래프(Bipartite Graph)로 도시하여 최대 흐름 정합 (Max Flow Matching)을 얻어 음

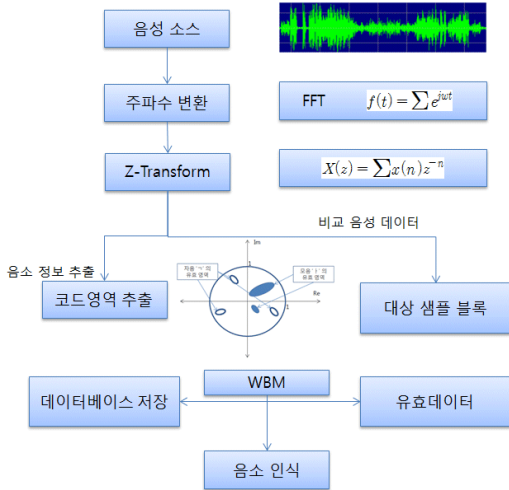
소 인식을 하는 방법이다. 본 연구에서 개발한 방법은 음성의 시작점과 끝점이 존재 하지 않아도 음성을 인식 할 수 있다. 뿐만 아니라 음소 인식의 수행 시간이 매우 빨라 실시간 명령 인식에 매우 유용하다. 그러나 HMM에 비해 음소 인식률이 떨어지는 단점이 있다. 우리가 개발한 시스템은 게임과 로봇 분야에서 실시간으로 명령을 주고받을 수 있도록 해준다. 게임과 로봇 가진 분야의 경우는 음소 인식률이 약간 떨어지더라도 빠르게 음소를 인식하여 전달하는 것이 보다 우선적이기 때문에, 제안된 방법을 사용하게 되면 게임분야와 로봇 분야에 효과적으로 적용 될 것으로 판단된다.

본 논문은 2장에서 전체 시스템 구조를 설명하고, 3장에서 Z 평면에서 음소 정보를 얻는 방법을 설명한다. 4장에서 시스템 설계 방법을 설명하고, 5장에서는 가중치 두 갈래 그래프 최대 흐름 정합 (Weighted Bipartite Max Flow Matching: WBM)을 이용한 음소 인식 방법을 설명하고, 6장에서 성능평가를 하고 결론을 맺는다.

2. 시스템 전체 구조

[그림 1]은 본 논문에서 제안하는 시스템의 전체 구조를 보여 준다. 우선 입력되는 음성 주파수를 FFT(Fast Fourier Transform)를 한다. 그 다음 전환된 주파수 데이터를 Z 평면에 도시를 한다. Z 평면에 있는 수많은 값들 중에는 모음에 기여하는 주파수, 자음에 기여하는 주파수들이 있는데, 이 기여 값들은 타원 형태의 여러 영역으로 설정되며, 겹쳐져서 설정되기도 한다. 이 영역 정보를 이용하여 자음과 모음에 기여하는 영역 정보 데이터를 추출한다. 그리고 이 값을 데이터베이스에 저장하는데, 비교 검색하기 위한 음성 정보가 입력되면 Z 평면에 도시한다. 여기에서 2차원 배열로 정보 영역을 변환 시키고 해당 정보 영역과 데이터베이스 값을 WBM을 이용하여 음소 인식을 한다. 음소 정보를 가지고 있는 코드 영역을 추출하기 위해서

많은 샘플이 필요하다. 같은 음소의 다양한 샘플 데이터를 추출 하게 되면 최적화된 음소 샘플을 얻을 수 있다.

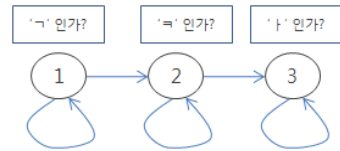


[그림 1] 시스템 전체 구조

HMM에서 사용하는 음소 샘플은 주파수 정보인데, [그림 2]는 HMM을 사용하는 음소 인식 방법의 예이다. HMM에서는 체인을 따라 가면서 해당 주파수가 있는가를 확인하기 때문에 여러 주파수 블록을 비교해야 한다.

그러나 본 논문에서 사용하는 방법은 주파수 영역 값의 매칭이기 때문에 주파수 블록의 비교가 필요하지 않다. 따라서 HMM 보다 매우 빠르게 비교가 가능하다.

WBM은 두 갈래 그래프에 비교 음성의 주파수 영역 값을 가중치로 두고 정합 결과 값을 유추하는 방법이다. WBM의 특징은 최대 흐름 값을 유추 하는 것이다. 따라서 특정 부분에 에러가 있다 하더라도 전체적인 흐름이 유사하면 유사도가 높은 결과를 얻을 수 있다. 이것은 음성의 시작점이 없이 일정 블록 구간에서 음소 인식을 할 때 빠르게 음소를 추출 할 수 있는 장점을 가진다. 따라서 본 방법을 연속적인 오디오 신호에서 명령에 필요한 음소를 유추할 때 매우 빠르게 할 수 있다[6].



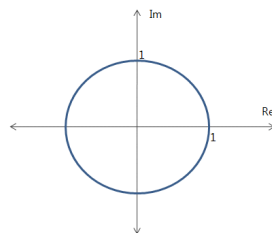
[그림 2] HMM의 음소 인식 방법의 예

3. 음소 정보의 Z 평면 전환

다음과 모음에 기여하는 값을 추출하기 위해서는 음성 샘플을 시간 영역에서 주파수 영역으로 전환시켜야 한다. 입력되는 음성 파장을 대역별 주파수로 전환 한다. 이때 사용하는 주파수 변환 함수는 FFT(Fourier Transform)이다. FFT를 하게 되면 고주파 영역부터 저주파 영역에 분포되어 있는 다양한 정보 값을 규칙적으로 분석할 수 있다. (식 1)은 음성을 주파수 변환하는 FFT 함수 이다.

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]e^{i\omega(n-k)} \quad (\text{식 1})$$

FFT는 복소수로 나오기 때문에 실수부(real part) 와 허수부(image part)의 값이 따로 분리 된다. 이 두 개의 값의 결합의 특성을 분석 하여야 정확한 음소 분석이 가능하다. 두 개의 상관관계를 이용하여 음소를 분석하기 위해서 FFT 결과를 Z 평면으로 전환해야 한다. FFT 이후 실수부와 허수부를 Z 평면으로 전환 하게 되면 [그림 3]과 같은 ROC(region of convergence)가 형성 되는데 FFT는 단위원 형태로 ROC가 형성 된다.

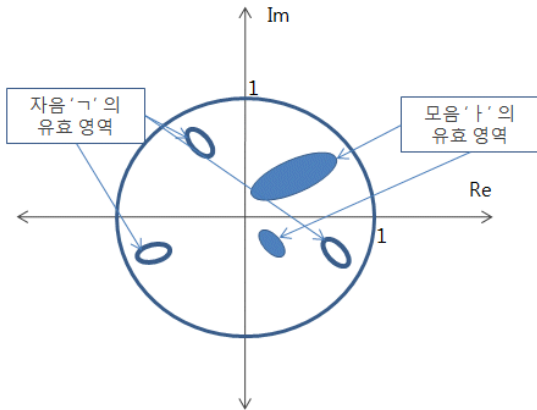


[그림 3] 음성 주파수 Z 평면 전환

본 논문에서 사용하는 Z 평면의 기반 함수를 FT(Fourier Transform) 하였다. (식 2)는 Z Transform 의 기본형이다.

$$H(z) = \sum_{n=-\infty}^{\infty} h[n]z^{-n} \quad (\text{식 2})$$

음성 주파수는 다양한 정보와 함께 유효 정보 영역으로 나누어진다. 특정 주파수의 영역이 모음과 자음에 유효한 값이 되는데 이 영역들은 Z 평면 안에 타원 영역으로 표시 된다. [그림 4]는 “가” 음에 대한 주파수 분포를 예를 들어 보여 준다.

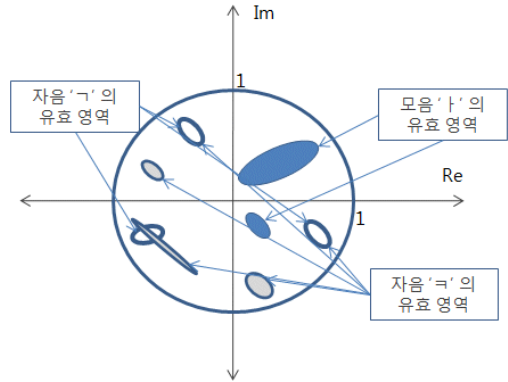


[그림 4] ‘가’ 음의 Z 평면에서 유효 영역

모음 ‘ㅏ’의 경우 주파수 평면의 1사 분면의 긴 원형의 영역과 2사 분면의 작은 타원 영역이 기여가 큰 주파수 영역이 되며, 자음 ‘ㄱ’은 2, 3, 4 사 평면에 작은 원형으로 분포 된다. 그러나 위의 영역에 주파수 값을 매칭 하였다고 무조건 그 음소가 정확하게 ‘가’로 나오지는 않는다. 그 이유는 영역 안의 값이 존재하는 필요조건이지, 필요충분조건이 아니기 때문이다.

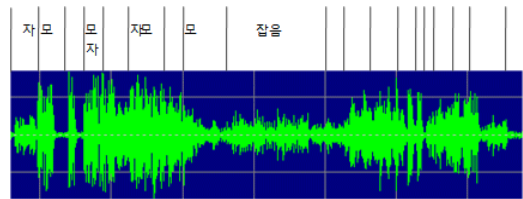
예를 들면 ‘카’와 ‘가’를 함께 도시하면 [그림 5]와 같은 형태가 된다. ‘카’의 음성은 ‘가’의 음성 영역의 일부는 겹쳐지고 특정 영역은 모든 부분을 포함하고 있게 된다. 따라서 ‘가’의 음소 영역이라

하더라도 ‘가’일 경우와 ‘카’일 경우의 중복이 형성된다. 이렇게 중복이 형성되기 때문에 효율적인 정합 방법을 사용하지 않으면 원하는 음소를 분석해 내기 힘들다. 본 문제를 해결하는 방법을 우리는 흐름 정합을 이용하였는데, 다양한 위치에 다양하게 형성되는 타원들의 값을 하나의 가중치로 두고 두 개의 흐름을 비교하는 방법이다.



[그림 5] 카와 가의 음소 공통 겹침 형태

[그림 6]은 자음과 모음이 순차적이면서 결합적인 모형을 보여 주고 있다. 이런 음성 주파수에서 모음과 자음을 추출하여 순차적으로 결합 하고자 할 때는 컨벌루션 곱을 사용하여야 한다. 컨벌루션 곱을 적용하면 많은 복잡도를 유발한다. 따라서 무설정 일련의 음성 신호에서 명령을 실시간으로 이용하기 어렵다.



[그림 6] 자음과 모음이 순차적 결합적인 음성 주파수

본 연구에서는 이 문제를 해결하기 위해서 우선 단어 리스트를 설정하고 해당 단어를 비교한 후에 비교 에러가 발생 하였을 때 전체 데이터베이스에

서 음소를 검색 하는 방법을 사용 한다. 본 방법을 사용하기 위해서 모음에 기여되는 영역 정보, 자음에 기여되는 영역 정보 값을 데이터베이스에 저장 한다. (식 3)은 Z 평면에서 유효한 영역 정보를 추출 하는 기본 함수 이다.

Z 평면에서 도시되는 유효 정보는 타원 방정식 이다. 이 타원 방정식의 장축과 단축의 정보 값이 음소의 정보 값이 된다. 영역 정보의 비교는 주파수 실제 값을 비교하는 것 보다 매우 빠르다. 따라서 기존의 주파수 값의 비교 거리 차 보다 매우 빠른 결과를 얻을 수 있다.

$$D(p_i, r_i) = \sum \{(x-p)^2 + (y-r)^2\} \quad (\text{식 3})$$

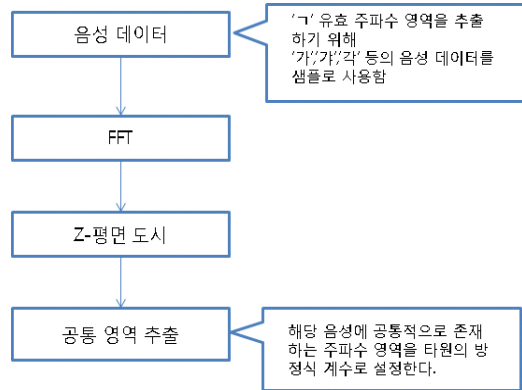
4. 시스템 설계 및 구현

제안된 방법으로 음소 인식을 하기 위해서는 두 개의 시스템이 구현되어야 한다. 첫 번째 시스템은 음소 값을 정의하고 정의된 음소 값을 추출하는 시스템이다. 두 번째 시스템은 시작 신호 없이 입력되는 음성 신호에서 음소 영역의 주파수 대역 정보를 추출 하는 시스템 이다.

4.1 유효 음소 주파수 대역 추출

음소의 유효 주파수 대역을 정의하기 위해서는 다량의 데이터 샘플링이 필요하다. [그림 7]은 유효 음소 주파수 대역을 추출하는 시스템의 전체 구조 이다. 제안하는 시스템에서 추출하고자 하는 계수는 타원의 방정식의 중점 좌표와 장축과 단축의 반경 값과 해당 영역에 분포되어 있는 주파수 가중치이다. 이 값을 추출하기 위해서 많은 음성 데이터를 분석해야 하는데, [그림 7]은 ‘ㄱ’의 자음 주파수 공통 영역의 계수를 추출하기 위한 예를 보여준다. 먼저 ‘ㄱ’의 값이 있는 음성 샘플을 FFT를 한 후에 Z 평면에 도시 한다. ‘ㄱ’의 값이 있는 많은 음성 블록의 값을 사용하면 공통된 영

역을 추출 할 수 있다.



[그림 7] 유효음소 주파수 대역 추출

제안된 방법으로 최적화 된 음소 주파수 대역을 얻기 위해서는 서로 다른 음성에 많은 음소 샘플을 이용해야 한다. 우리는 본 실험의 샘플을 추출하기 위해 20명의 서로 다른 사람에 공통 자음이 있는 음소 정보를 각각 100개씩 사용하여 유효 음소 영역을 추출 하였다.

4.2 비교 음성의 유효 주파수 계수 추출

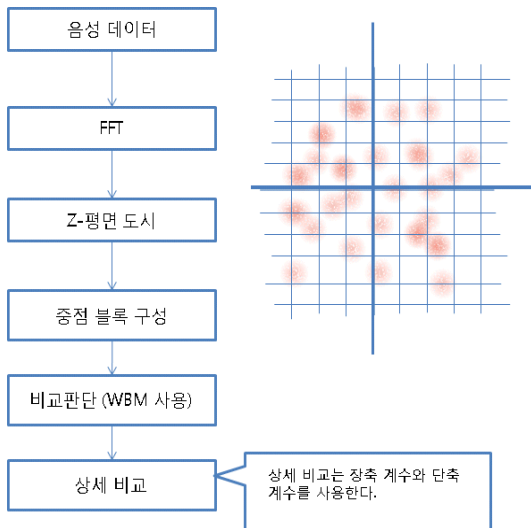
유효 음소 주파수 대역 계수를 얻은 후 해당 정보를 데이터베이스에 저장한다. 음소를 분석하기 위한 음성 비교 시스템은 [그림 8]과 같다. 입력된 데이터를 FFT한 후 Z평면으로 전환하는 것은 이전 장에서 소개한 방법과 같다. 이후 데이터를 정방형 원형 블록으로 나누고 해당 블록의 중점과 계수를 추출 하여 데이터베이스에 있는 계수와 비교 판단을 한다.

다음으로 입력된 음성을 일정한 블록으로 분할 한다. 본 시스템에서는 0.25초 블록 거리를 설정하는데, 시간에 의해 일정한 블록으로 분할할 경우 기여되는 주파수가 서로 다른 영역에 오류가 설정 될 가능성이 있다.

제안하는 시스템은 실시간으로 빠르게 음소를 인식하는데 목적이 있기 때문에, 음소를 실시간 블록으로 나누어 인식해야 한다. 즉, 이전 연구들의

문제점을 해결하기 위해서 제안하는 시스템을 자음 기반을 둔 시스템으로 구성하였다. 자음 주파수 대역은 매우 짧은 영역에 분포되어 있기 때문에 특정 시간 분할에 잘리는 위험성이 매우 적다. 그러나 모음은 긴 시간에 주파수 값이 분포되어 있기 때문에 두 개의 블록으로 분할될 가능성이 있다. 자음 기여 주파수는 매우 작으며 중고주파 부분의 여러 개의 주파수 대역으로 흩어져서 기여된다. 여러 개의 주파수 대역으로 흩어져 있다는 것은 여러 개의 중점을 기반으로 모여 있다고 볼 수 있다. 본 연구의 시스템은 Z평면을 여러 개의 중점 블록으로 나누고 해당 중점블록에 결합 되어 있는 주파수 값을 사용하여 어떤 자음이 있는 가를 분석한다. 제안된 시스템은 64개의 블록중점으로 데이터를 분석 하였다. 각 블록의 정보와 데이터베이스에 있는 자음 정보들과 비교 판단을 할 경우 템플레이트 매칭 방법을 사용하게 되면 수행 시간과 복잡도가 커진다.

그러므로 본 연구에서의 시스템은 빠르게 비교 판단을 하기 위해서 데이터베이스에 있는 자음 주파수 영역과의 비교 판단을 가중치 두갈래 그래프의 최대 흐름 정합을 사용하였다.



[그림 8] 음성 판단 시스템

WBM을 한 후 유효 자음 값들을 추출 할 수 있다. 판단하고자 하는 음성 정보의 Z평면에서 추출된 데이터는 장축과 단축이 같은 원형의 중점 계수이다. 데이터베이스에 있는 데이터는 장축과 단축이 다른 타원 형태의 정보이다. 두 개의 정보를 비교 판단 할 경우 오차가 발생된다. 이때 유효한 자음 주파수를 한 개 이상 선정하고 해당 자음 주파수의 장축과 단축 계수를 이용하여 상세 비교를 한다. 상세 비교 이후 정확한 자음을 추출 할 수 있다.

5. 가중치 두 갈래 그래프 최대 흐름 정합 (Weighted Bipartite Max Flow Matching: WBM)

본 논문에서는 음소 인식 분야에 새로운 알고리즘 개발을 제안한다. 본 연구에서 제안 방법은 음소 판단 방법으로 자음 기반 최대 흐름 정합(Max Flow Matching)를 이용한다.

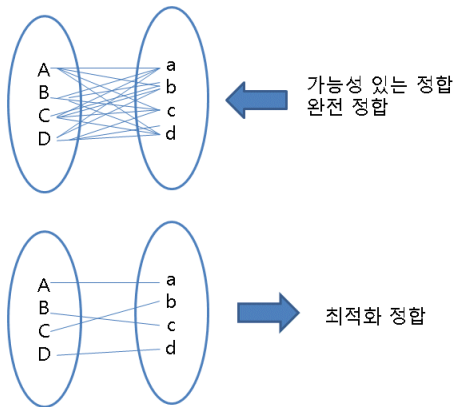
5.1 유사도 비교 방법 WBM

WBM을 구할 경우, 일반적으로는 헝가리안 메소드를 이용하여 최소 소요 비용을 얻고, 이를 통해 정합을 찾아낸다. 최소 소요 비용이란 특정 정점에서 다른 정점으로의 이동에 드는 비용을 의미한다. $G=(V, E)$ 그래프에서 임의의 에지 (a, b) 의 최대 흐름이 W_{max} 이고 (a, b) 의 가중치가 $W_{(a,b)}$ 라 할 때 최소 흐름 비용 W_{min} 은 다음 (식 4)와 같다.

(식 4)에서 최소 소요 비용을 찾아내는 것은 최대 흐름의 정합 쌍을 찾는 방법과 같다. 유사도 비교를 할 때에도 이와 같이 최소 소요 비용을 찾는 방법을 이용한다.

$$W_{min} = W_{max} - W_{(a,b)} \quad (\text{식 4})$$

[그림 9]는 두 갈래 그래프를 이용한 정합의 예를 보여 준다. 가중치 있는 에지 정보가 상대 그래프 에지와 연결 가능한 모든 고리를 다 연결 하였을 때를 완전 정합이라고 한다. 완전 정합은 1대 1 대응 결합의 모든 경우가 된다. 이때 최적의 정합의 값을 얻어내는 1대1 대응의 값을 얻어 내는 방법이 WBM이다. WBM을 이용하여 다중 정보를 비교 판단 하는 방법은 이전 연구들[8, 9, 10, 11]에서 이루어져 왔는데, WBM을 사용하면 다중 주파수 정보를 한 번에 비교 검색할 때 매우 효과적이다. 영상과 음성을 주파수 평면으로 전환하고 주파수 평면에서 얻은 다양한 값을 비교하는 방법들로 많이 이용되어왔다.

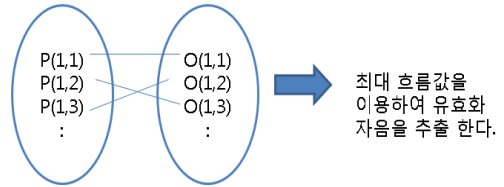


[그림 9] WBM 정합

5.2 WBM을 이용한 판단

WBM을 이용하여 최적화 된 완전 그래프를 찾아 비교하여 차이 값을 얻어내는 방법으로 음소를 찾아낸다. [그림 10]은 제안하는 알고리즘의 간단한 예를 보여준다. 데이터베이스에 있는 정보는 음소에 기여되는 Z평면의 중점 좌표, 장축과 단축 계수 및 주파수 값이다. 갈래 그래프의 요소 값으로 중점 좌표와 장축, 단축 계수, 그리고 주파수 가중치가 된다. 여러 개의 요소 값이 두 갈래 그래프의 한 그래프로 설정되고 정합을 하고 정합의 결과로부터 최대 흐름 정합 값을 얻는다. 한글은 초성이

19자로 설정되어 있기 때문에 19개의 유효 중점 영역을 최대 흐름 정합 계수를 찾는 것은 매우 빠르게 얻어 낼 수 가 있다.



[그림 10] 중점 에지 정보를 이용한 WBM

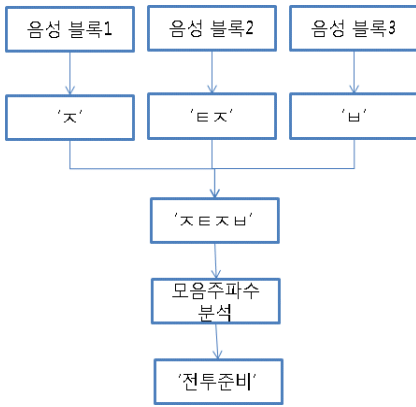
제안하는 방법은 주파수 도메인에서의 값을 추출하는데, 시간 영역에서는 순차적이거나 주파수 영역에서는 순차적이지 않다. 따라서 자음의 위치가 바뀔 수 있다. 예를 들어 “반갑다”라는 음소를 매우 빠르게 이야기하면 “ㄴ”과 “ㄱ”이 하나의 음소 블록에 존재할 수가 있다. 이럴 경우 제안된 알고리즘은 매우 효과적으로 두 개의 자음을 추출해 낼 수가 있다. 단점으로는 자음의 순서가 보장 되지 못한다. 본 문제를 해결하기 위해서 자음을 유추하여 유효 명령어에 비교할 경우에도 WBM을 사용하면 된다. WBM의 특징은 순서가 독립적인 비교 판단이 가능하다[8, 9, 10]. 따라서 자음의 순서가 바뀌어도 효과적으로 음소를 인식 할 수가 있다.

5.3 모음 주파수를 결합

자음 주파수를 이용하여 유효한 단어의 자음을 유추한다. 실시간 게임에서 본 연구에서 개발한 알고리즘을 이용하면 명령을 빠르게 인식 할 수 있다. 예를 들어 “전투준비”라는 인식을 “즈트즈니”로만으로도 판단이 가능하다.

음소인식시 명령이 아닌 문장 전체를 음소로 전환하고자 할 경우가 있다. 그럴 경우에는 자음 추출로부터 모음의 결합 위치를 판독하고 음성 블록을 재조합한다. [그림 11]은 모음을 결합하는 방법을 보여준다. 여러 개의 작은 블록으로 나누게 되

면 모음이 여러 개의 블록에 나뉘게 된다. 그러나 자음은 주파수 시간 간격이 매우 짧아서 한 블록에서 검색이 가능하다. 이런 특징을 이용하여 자음을 이용하여 음소 판단을 한다. 자음을 이용하여 음소를 판단하게 되면 전체 단어 블록을 유추할 수 있다. 다음 유추된 단어 블록을 다시 결합하여 모음을 분석한다. 모음을 분석하는 방법도 자음을 분석하는 방법과 같은 방법을 활용한다.



[그림 11] 모음 결합 방법

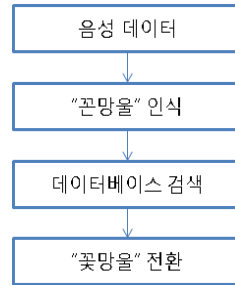
5.4 변형 음소에 대한 사전식 판단

한국어를 인식할 때 구개음화와 연음 법칙에 의해 자음 주파수가 변화될 수가 있으며, 이로 인해 음소 인식에 오류가 발생할 수 있다. 음소 인식에서 본 항목이 예외 적용 부분에 해당 한다. 제안된 알고리즘의 목적은 정확도가 다소 떨어져도, 빠르게 음소인식을 할 수 있게 하는 것이다. 따라서 실시간으로 시스템이 알고리즘을 수행할 수 있어야 한다.

본 알고리즘을 이용하여 운영하고자 할 때 시스템의 수행 능력을 고려해야 한다. 시스템의 성능이 우수 할 경우 예외 적용을 적용 할 수가 있다. 그러나 시스템 성능이 우수 하지 않을 경우 예외 적용을 하지 못할 수 있다. PC에서 본 알고리즘을 적용하고자 할 경우 예외 처리가 가능하다. 예외 처리는 예외 부분에 해당하는 음성 블록의 자음

블록을 이용한다.

[그림 12]는 예외 적용의 예를 보여 준다. 음소 인식을 하였을 때 자음을 이용하게 되면 “ㄱㄴㅇ”를 이용한다. 이때 모음 주파수를 분석하여 결합하면 “꼐망울”로 인식하게 된다. 본 단어가 예외 적용 데이터베이스에 존재하는가를 확인한 후 데이터베이스에 존재한다면 “꽃망울”로 전환한다. 본 알고리즘은 게임에 적용하기 때문에 음소를 문자로 표시할 경우가 극히 드물다. 따라서 예외 적용을 하지 않는 경우가 많다. 그러나 교육을 결합한 게임을 제작할 경우와 같이 음소인식의 정확도가 필요할 경우 예외 적용 사전을 이용한 본 방법을 적용 한다.



[그림 12] 예외 적용

5.5 노이즈 제거 알고리즘

음소 인식 시스템에서 명령하고자 하는 화자의 명령 외에 다양한 노이즈가 입력된다. 이런 노이즈를 무시하고 유효한 음소 인식을 하기 위해서는 노이즈 제거 알고리즘을 개발 하여야 한다. 노이즈에도 음소가 결합 되어 있기 때문에 단순한 주파수 제거로 노이즈를 제거하기가 매우 힘들다. 제안 대상 기술 과제에서 사용할 노이즈 제거 알고리즘은 화이트 노이즈를 제거하는 방법을 응용한 것이다. 일반적으로 노이즈는 주파수 전역에 고루 퍼져 있어서 음소를 인식하기 힘든 것을 의미 한다. 노이즈 제거 방법은 이전 블록과 현재 블록과의 대역별 주파수 평균값을 비교하여 해당 평균값이 존재할 경우 노이즈를 해석하고 주파수를 제거하는

것이다. 앞에서 제안된 음소 인식을 자음기반 두 갈래 그래프 최대정합을 사용할 경우 노이즈에 포함된 음소도 삽입이 된다. 그러나 제안 알고리즘은 노이즈에 의해서 생성된 음소는 매칭이 되지 않고 무시되기 때문에 앞 절에서 사용한 알고리즘에 의해서 음소 노이즈 제거가 가능하다.

6. 성능 평가 결과 및 결론

본 논문에서 목적으로 하는 음소 인식은 시작 설정 없이 일련의 음성 주파수에서 음소를 정확하게 인식하는가가 중요하다. 표1은 HMM과 제안하는 방법과의 성능 평가 비교표이다.

[표 1] 음소 인식 성능 평가 비교표

시간 (단위: 분)	단어 수	HMM	WBM
1	10	8	6
2	20	10	14
3	30	9	23
4	40	13	31
5	50	12	38

[표 1]의 결과는 총 300회의 음성 명령을 주고 인식 한 결과의 평균이다. 실험 시 분석 유효 시간을 30초로 설정 하였고 30초 를 넘으면 분석을 종료 하는 방법을 사용 하였다. 1분 미만일 경우 HMM이 정확도가 높았다. HMM은 짧은 영역에서 세부 검사를 할 경우 정확도가 높기 때문이다. 그러나 HMM은 샘플 데이터가 길어지면 음성 시작점을 유출하기가 어렵기 때문에 음성 데이터 시간이 길어질수록 인식률이 떨어진다. 또한 입력 데이터 시간이 길어질수록 분석 시간이 길어져서 3분 이상의 데이터는 30초 이내에 판별 하지를 못했다. HMM은 음성이 지속적으로 입력 하는 경우에는 인식을 제대로 할 수가 없다는 것으로 판단되었다. 제안된 방법도 5분 동안에 명령에서 12개의 명령을 인식 하지 못하는 단점은 있으나, 빠른 명령을 인식하는 방법에서는 최적화 된 방법이라 판단된다.

지금까지 게임에서 음소 인식을 사용하는 것은 매우 힘들었다. 게임은 빠르게 진행 되어야 하나, 음소 인식은 시간 지연이 발생되기 때문이다. 또한 지속적으로 들어오는 음성 데이터에서 음소와 노이즈를 판단하여 분리한 후 비교 하는 방법이 제안 되지 않았다[7].

지금까지는 음소인식에 있어 정확도에 집중하는 경우가 대부분 이었다. 그러나 게임과 같이 빠른 판단과 운영이 필요한 분야에서는 음소인식 속도의 향상이 필수적으로 요구된다. 음소인식이 빠르게 이루어질 수 있는 방법은 크게 하드웨어가 발전하는 방법과 음소를 빠르게 인식하는 방법의 개발이 있다. 본 논문은 기존의 방법과 다르게 빠르게 음소 인식을 하는 새로운 방법으로서, 충분한시간이 주어졌을 때의 인식은 기존의 알고리즘보다 떨어지는 단점이 있으나, 빠른 시간에서의 인식은 타 어느 알고리즘 보다 우수하다.

본 연구에서의 알고리즘은 다양한 부분에 적용 가능한데, 특히 빠른 음성 명령이 필요한 분야와 실시간으로 다양한 음성 명령이 필요한 분야에서 유용하게 적용할 수 있다. 우선 음성 명령의 빠른 전달이 가장 필요한 분야는 게임 분야와 로봇 분야로 판단이 되는데, 게이머들이 게임을 하는 모습을 관찰해보면, “빨리 빨리” 또는 “바라라~”와 같은 말을 하는 것을 알 수 있다. 그러므로 게임 시 게이머의 흥미를 높이기 위해서는 조이스틱 그리고 마우스와 같은 전통적인 입력 장치 뿐만이 아니라 빠른 음성 명령이 전달 가능한 향상된 입력 시스템이 개발되는 것은 의미가 있다. 본 연구에서 개발한 알고리즘은 음성 명령을 결합한 게임을 개발하고자 할 때 매우 효과적으로 사용될 수 있을 것이다.

참고문헌

- [1] 최원석, 이동우, 김문식, 나종화, “음성인식 및 영상처리 기반 멀티모달 입력장치의 설계”, 제어·로봇·시스템학회 논문지, Vol. 13, No. 8,

pp.743-748, 2007

- [2] 정보통신산업진흥원, “모바일과 클라우드로 재조명받는 ‘음성인식’”, 주간기술동향, 1520호, 2011년 11월
- [3] 한국콘텐츠진흥원, “음성인식기술의 동향과 전망”, 문화기술(CT) 심층리포트, 2011년 11월
- [4] 안종영, 김상범, 김수훈, 허강인, “모델적용 HMM을 이용한 모바일 환경에서의 음성인식에 관한 연구”, 한국인터넷방송통신학회 논문지, Vol. 11, No. 3, 2011
- [5] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, Proceedings of the IEEE, Vol. 77, No. 2, pp. 257-286, 1989
- [6] Sangyeob Lee, Whoiyul Kim, “Robust Character Image Retrieval Method Using Bipartite Matching and Pseudo-bipartite Matching”, Proceedings of the First International Workshop on Advanced Internet Services and Applications (AISA), London, UK, 2002
- [7] S. Sclaroff and A. P. Pentland, “Search by Shape Examples: Modeling Nonrigid Deformation”, Proceedings of 28th Asilomar Conference on Signals, Systems, and Computers, pp.1341-1344, October 1994
- [8] SangYeob Lee, Whoi-Yul Kim, “Robust Character Image Retrieval Method Using Bipartite Matching,” LNCS 2402, pp. 295-306, 2002
- [9] 이상엽, 김희율, “Bipartite Matching 을 이용한 강인한 캐릭터 영상 검색 방법,” 방송공학회 논문지 Vol.7, No.2, pp. 136-143 June 2002.
- [10] Whoi-Yul Kim and C. Kak, “3-D Object Recognition Using Bipartite Matching Embedded in Discrete Relaxation,” IEEE Trans. Pattern Anal, Machine Intell., vol. 13, no.3, pp. 224-251
- [11] J.Canny, “A computational approach to edge detection,” IEEE Trans, Pattern Anal. Machine Intell., vol. PAMI-8, pp. 679-698, Nov. 1986



이 상 엽 (Lee, Sangyeob)

1998년 한양대학교 전자계산학과 석사
2003년 한양대학교 전자공학과 박사
현 재 삼육대학교 경영정보학과 교수

관심분야 : 영상처리, 멀티미디어, 게임, 로봇



박 성 원 (Park, Seongwon)

2003년 이화여자대학교 컴퓨터학과 석사
2011년 연세대학교 정보대학원 박사
현 재 한국콘텐츠진흥원 산업분석팀 선임연구원

관심분야 : 게임, 디지털콘텐츠, IT 서비스