

## 의사 형태소 단위의 음성언어 형태소 해석<sup>1)</sup>

이경님

서강대학교 전자계산학과  
서울시 마포구 신수동 1, 우: 121-742  
knlee@nlpeng.sogang.ac.kr

정민화

서강대학교 전자계산학과  
서울시 마포구 신수동 1, 우: 121-742  
mchung@ccs.sogang.ac.kr

### Morphological Analysis of Spoken Korean Based on Pseudo-Morphemes

Kyongnim Lee

Department of Computer Science  
Sogang University

Minhwa Chung

Department of Computer Science  
Sogang University

#### 요약

본 논문에서는 언어학적 단위인 형태소의 특성을 유지하면서 음성인식 과정에 적합한 분리 기준의 새로운 디코딩 단위인 의사 형태소(Pseudo-Morpheme)를 정의하였다. 이러한 필요성을 확인하기 위해 새로이 정의된 40개의 품사 태그를 갖는 의사 형태소를 표제어 단위로 삼아 발음사전 생성과 형태소 해석에 초점을 두고 한국어 연속음성 인식 시스템을 구성하였다.

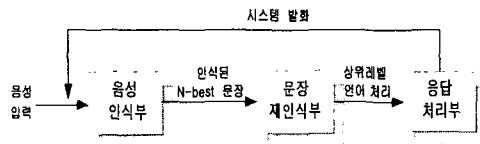
#### 1 서론

일반적인 한국어 음성 대화 처리 과정은 [그림 1]과 같다. 사람의 음성을 입력으로 받아 신호적 특성을 고려하여 특징추출을 하고 기존에 학습한 데이터와 비교를 통해 인식된 단어열의 결과를 가지고 문장 인식을 한다. 이 인식된 문장들은 음성신호 처리상의 오류를 포함한 상태에서 문장을 인식하기 때문에 100% 올바른 결과로 내놓을 수 없으므로 N 개의 최적 문장을 가지고 문장을 재인식한다.

문장 재인식 방법으로 사용하는 대표적인것들은 화자의 다음 발화 의도를 예측하여 인식 오류를 감소 시키는 방법[5]과 단어 trigram 같은 언

어모델을 이용하여 문장을 재배열함으로써 최종 문장을 얻는 방법[9] 등이 있다. 첫번째의 경우 표제어 단위에 따라 후처리시 사용되는 규칙들도 수시로 바뀌게 되며, 영역이 바뀔 때마다 단어 범주의 재분류가 필요하게 되므로 문장 요소들의 재분석 등 많은 수정이 요구된다. 두번째 경우 등록된 데이터의 부족 현상에 대한 대비책이 필요하게 되며, 대용량 연속음성 인식의 경우 문장의 수가 늘어남에 따라 사전의 크기가 증가하고, 그 경우의 수가 많아지기 때문에 인식이 급속히 떨어지며 탐색 시간이 길어지는 단점이 발생한다.

본 논문에서는 형태소 해석을 수반하는 한국어 연속음성 인식 시스템을 구현하였다. 음성언어 형태소 해석을 위해 문장을 구성하는 새로운 디코딩 단위가 필요함에 따라 의사 형태소(Pseudo-Morpheme)를 정의하였고, 이를 기반으로 인식실험을 수행하였다.



[그림 1] 음성 대화 처리 시스템 구성도

본 논문의 구성은 다음과 같다. 2 장에서는 음성인식시 언어이해 측면에서 형태소 해석이 필요

1) 본 연구는 1998 년도 LG 전자의 연구비 지원에 의하여 이루어졌습니다.

한 이유와 일반적인 문자기반의 형태소 해석과 관점 차이를 살펴보고, 이를 기반으로 3 장에서는 새로운 문장의 디코딩 단위인 의사 형태소를 정의하였다. 구체적인 품사태그와 선정기준 등에 관한 내용을 기술하였다. 4 장에서는 본 논문에서 제안하는 한국어 음운변화 현상을 반영하여 음성언어 형태소 해석을 수반하는 의사 형태소 단위의 연속음성 인식 시스템 구성을 기술하였다. 5 장에서는 구현 및 실험결과를 분석하고, 마지막으로 6 장에서 결론 및 향후 연구과제를 제시하였다.

## 2 음성언어 형태소 해석

### 2.1 형태소 해석의 필요성

일반적인 한국어 자연어 처리의 과정은 다음과 같다. 서구 언어와는 달리 한국어에서는 그 구조적 특성상 형태소 해석이 선행되어야 한다. 한국어는 교착어로서 의미를 나타내는 실질 형태소에 조사와 어미 같은 어법적 관계를 나타내는 형식 형태소가 붙음으로써 문법 기능을 한다. 또, 각 낱말의 어미 변화에 의해 문장의 성분이 결정되며, 접용과 활용이 자유롭고, 불규칙 현상 및 음운 현상이 발달해 있다. 형태소 해석 과정을 거치지 않고 사전에 모든 형태소들의 결합을 그대로 수록하여 처리한다는 것은 한계가 있기 때문에, 형태소 해석은 한국어 처리를 위해 필수적인 요소라 할 수 있다.

기존의 음성인식 시스템의 경우는 대부분 인식된 문자열을 나열하는데 의의를 두어왔다. 그러나 단순히 올바른 문장을 찾아내는 데에 그치는 것이 아니라 언어이해를 수행하기 위해서는 형태소 분석을 통해 문장의 구성성분이나 구조를 파악하는 구문분석 단계를 거쳐 의미해석을 수행하며, 더 나아가서는 담화분석 등의 자연어 처리 기법을 통해 대화 시스템을 구축할 수 있다. 이러한 언어이해를 위해서 기존의 방식은 어절이나 단어 단위로 인식을 수행하며, 오류가 포함된 100% 완전하지 않은 문장을 기반으로 형태소 해석을 거쳐 상위레벨 언어처리를 수행하게 된다. 이런 경우 옳은 문장을 구성하기 위해 영역지식과 문장성분 지식 등 활용 가능한 정보들을 이미 사용하여 얻은 인식 결과로 다시 형태소 해석을 수행한다는 것은 반복된 작업의 결과가 된다.

따라서 음성인식 측면에서의 형태소 해석이 가

지는 의미가 크므로, 형태소를 입력으로 인식을 수행함으로써 그 결과값이 반복된 형태소 해석 기법을 거치지 않고 다음단계의 입력값으로 쓰일 수 있는 점에 의의를 가지며, 또한 음성인식 단계에서 형태소 및 형태소 접속정보 등의 다양한 언어적인 지식을 활용할 수 있게 된다.

### 2.2 문자기반 형태소 해석과의 차이점

음성언어 처리의 경우는 사람의 자연스러운 음성을 입력으로 받기 때문에 입력 단위, 즉 인식 단위의 경계가 모호하고, 신호처리의 결과가 이미 오류를 포함하고 있기 때문에 문자기반의 언어처리와는 달리 정확한 입력값을 기대하기 어렵다. 따라서 보다 정교한 음성언어 처리를 위해 형태소 해석 단계에서부터 음성인식 기술과 자연어 처리 기술을 접목하여 시스템을 구성하도록 하는 접근 방법을 채택한다.

그러나 문자 기반의 형태소 해석 기법[8]을 그대로 음성 인식 시스템에 적용하기란 그리 쉽지 않는데, 일례로 형태소 단위의 음성인식 시스템[11]에서는 다음과 같은 문제점을 볼 수 있다. 일반적인 언어학에서 말하는 형태소들은 단음소나 단음절로 이루어져 있다. 예를 들면, ‘ㄴ’, ‘ㄹ’, ‘이’ 등과 같은 단음소와 대부분 의존명사, 접미사의 경우는 단음절로서 하나의 형태소이다. 이와 같은 형태소는 음성인식에서 매우 짧은 시간 동안에 발생되기 때문에 이를 인식하기에는 많은 어려움이 있다. 따라서 음성인식에서는 이를 보완하기 위한 방법이 강구되어야 한다.

이러한 문제점을 피하기 위한 한가지 방법으로, 주로 형태소를 분리하지 않고 어절을 그대로 사용하는 경우를 들 수 있겠다. 발성의 지속 시간의 제법 길기 때문에 인식률을 어느 정도 이상으로 높일 수 있으나, 인식 대상 단어수가 늘어나고 활용에 많은 제약을 받는다는 단점 때문에 역시 잠정적인 해결책이 될 뿐이다. 결국 인식률을 높이기 위해서는 적절한 길이의 발성 시간과 적절한 수의 단어를 가질 수 있는 단위가 새로이 정의되어야 한다. 이를 위해서 가끔적으로 언어학적인 단위인 형태소를 유지하면서 음성인식에 그다지 무리가 가지 않는 범위 내에서 적절한 분리 기준에 따른 새로운 디코딩 단위인 의사 형태소의 정의가 필요하게 된다[6][7].

(제 10 회 한글 및 한국어 정보처리 학술대회)

3 의사 형태소의 정의

디코딩 단위는 사전의 표제어 단위이며 문장의 구성단위가 된다. 연속 음성 인식 시스템에서 인식률을 높이기 위해서는 적절한 길이의 발성시간과 적절한 수의 사전 표제어를 가질 수 있는 디코딩 단위가 정의되어야 한다. 이 장에서 설명하는 의사 형태소는 발음사전의 표제어, 더 나아가서는 음성언어 사전의 표제어 단위가 되기 때문에 선정 기준이 매우 중요하다. 본 논문에서 제안하는 의사 형태소는 기존 문자기반의 형태소 단위와 기본 골격은 같다.

이에 따라 의사 형태소 태그는 '한국어 표준 형태 통사 태그 표준안'[10]에서 제시하는 54 개의 품사태그와 KTS[8]에 사용된 52 개의 품사에 기반하여 음성언어 처리시 고려되는 새로운 기준에 적합한 38 개의 품사태그로 정의하였다.

음성인식의 특성상 구체적으로 고려되는 사항을 나누어서 살펴보면 다음과 같다.

첫째, 표준안의 대분류에서 기호(s)와 외국어(f)는 의사 형태소 분류에서 제외하였다. 마침표, 따옴표 및 묶음표, 이음표 등은 발화시 소리내어지는 음성신호 상의 특징이 아니므로 품사태그로 고려하지 않았다. 단, 쉼표의 경우는 음성신호상의 특징으로 인식시 모델링을 하며, 단위 기호(su)의 경우는 단위성 의존 명사(nbu)로 분류한다. 외국어의 경우도 문장상의 표현양식일 뿐 음성상의 특징이 아니므로 품사태그에서 제외하였다.

둘째, 합성어의 경우로 복합어나 파생어의 관계점에서의 'ㄴ'첨가 현상의 발생과 한자어의 결합에서만 발생하는 두음법칙, 사이 'ㅅ'의 비첨가 현상, 역행적 유음화, 그리고 한자어의 선택적 경음화 현상 등 이를 위한 역추적 규칙은 그 노력에 비해 손실이 크기 때문에 지금까지 이는 발음 생성시 예외사전 처리를 해왔다. 이와 같은 것은 의사 형태소로 묶고, 이와 같은 특징을 형태소 정보에 신고 이를 이용하여 그 특성에 맞춘 발음 열을 생성하고, 또한 분석시 의사 형태소로 구성된 이 정보를 이용하여 효율적인 분석도 꾀할 수 있다.

셋째, 숫자 및 시간 처리를 위해 태그를 수정하였다. 하나는 시간성 보통명사(nct)로 KTS에서는 있었으나 표준안에서는 제외된 태그로 음성언어 처리시 자주 사용되는 시간이나 월, 일 등의 시간명사구를 위해 복구시킨 것이다. 양수사(nnc)

의 경우 인식단위를 '숫자+단위성 의존명사'를 결합하여 하나로 보았다.

[표 1] 의사 형태소 태그 집합

대분류	태그	명칭	결합규칙
체언 (n)	ncpa	동작성 보통명사	
	ncn	비서술성 보통명사	
	nct	시간성 보통명사	← nnc+nbu
	nq	고유명사	
	nbu	단위성 의존 명사	
	nbu	비단위성 의존명사	
	npp	인칭대명사	
	npd	지시대명사	
	nnc	양수사	← nnc+nbu
nno	서수사		
용언 (p)	pvd	지시동사	
	pvg	일반동사	
	pad	지시형용사	
	paa	성상형용사	← ncps + xsm
px	보조용언	← 양상정보	
수식언 (m)	mmd	지시관형사	
	mma	성상관형사	← paa + etm
	mad	지시부사	
	maj	접속부사	
	mag	일반부사	← +xsa
독립언	ji	감탄사	
관계언 (i)	jcs	주격조사	
	jco	목적격조사	
	jcc	보격조사	
	jcm	관형격조사	
	jcv	호격조사	
	jca	부사격조사	
	jcj	접속격조사	← jct
	jxc	통용보조사	
	jp	서술격조사	
어미 (e)	ecc	대등적 연결어미	← ep + ecc
	ecs	종속적 연결어미	← ep + ecs
	ecq	인용형 연결어미	← ep + ecq
	etn	명사형어미	← ep + etn
	etm	관형사형어미	← ep + etm
	ef	종결어미	← ep + ef + jxf
접사 (x)	xsn	명사파생접미사	
	xsv	동사파생접미사	

이런 경우 사전의 크기 증가 문제를 들 수 있는데, 1부터 99999999까지 표현하는 경우 14개의 형태소 조합으로 표현 가능한 것이 의사 형태소로 보았을 경우 45개의 조합으로 표현 가능하며, 99999999999999까지는 15개의 형태소로 표현 가능

(제 10 회 한글 및 한국어 정보처리 학술대회)

한 반면에 54 개의 의사 형태소로 표현 가능하다. 천 단위의 증감에 대해 9 개의 의사 형태소의 증가로 모든 수가 표현 가능하다. 이렇게 함으로써 ‘칠십, 팔십’과 같이 경계점에서 발생하는 경음화 현상등이 발음사전으로 해결된다.

넷째, 용언구를 살펴보면 ‘감사합니다, 수고하세요’와 같은 경우 ‘감사/ncps+하/xsm+ㅁ니다/ef, 수고/ncps+하/xsm+세요/ef’의 분석결과를 얻기 위해 쪼개어 인식하게 되는데, 실제 인식 실험 결과 앞부분에 강세가 있어 형용사 파생접미사 ‘하’가 인식이 잘 안 될 뿐만 아니라 실제 대화 시스템에서 많이 사용되지 않으므로 ‘하’가 결합된 하나의 형용사 단위로 본다.

다섯째, 한국어에서는 양상을 나타내는 보조용언과 어미 등의 굴절접사가 대단히 발달되어 있어 다양한 형태와 본토, 의미를 보이고 있다. 또한 문장에서 대부분의 문법 기능을 담당하고 있으므로 문장을 이해하는데 매우 중요한 역할을 한다. 일반적인 한국어 술어의 구조를 나타내면 다음과 같다.

술어 = 어간+보조용언+선어말어미+어말어미  
 보조용언 = (피동|사동|상|법성|부정)  
 선어말어미 = (존칭 |ε)+시제+경칭  
 어말어미 = 서법

서술어의 경우 어미의 활용에 따라 인식성능의 저하를 가져오기 때문에 그 기준 설정이 매우 중요하다. 대부분 음성인식 수행시 용언구 전체를 하나의 표제어로 삼아 인식을 수행하게 되는데 자유로운 문형을 표현하기 어려울 뿐만 아니라 화자의 자연스러운 문장 구성을 어렵게 만드는 요인이 된다. 이를 고려하여 한국어 술어 구조를 의사 형태소 단위로 재정리하면 다음과 같다.

술어 = 어간+보조용언+어말어미  
 보조용언 = 양상정보  
 어말어미 = 선어말어미+종결어미+종결보조사

용언구를 하나의 어절로 묶어서 사전에 기재하는 것 보다는 다양한 어미의 활용을 위해 어간과 어미를 분류하여 인식과정에서 결합하고, 다양한 문형의 표현을 위하여 양상정보를 포함하고 있는 보조 용언구를 하나의 인식단위로 선정하였다. 이렇게 어간과 보조용언, 그리고 어말어미의 결

합으로 활용이 자유로운 의사 형태소 단위의 문장을 구성할 수 있게 된다.

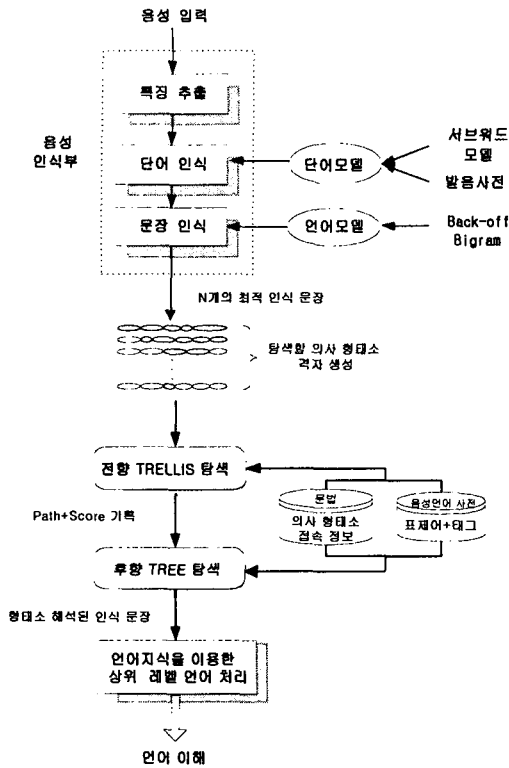
여섯째, 빈번히 생략이 가능한 조사의 경우로 다음과 같이 나누어 살펴 볼 수 있다. 격조사는 여러 개의 결합된 형태 자체를 하나의 단위로 본다. 종결 보조사는 앞의 형태소와 결합한 형태로 인식단위를 선정한다.

일곱째, 서술어의 경우 어미의 결합은 필수적이다. 음성인식에서 용언구의 인식은 전체 성능 평가에 커다란 영향을 미치므로 가능한 한 그 음운변화를 그대로 사전에 반영하는 것이 시스템의 성능에 도움이 된다. 따라서 여기서는 선어말어미는 연결어미와 전성어미, 종결어미와 결합된 형태로 사전에 등록한다.

또한 기존의 한국어 형태소 품사태그 집합으로는 표기하기 힘든 경우들이 대화체 문장에서 자주 발생하게 되는데, 대표적으로 준말 표현의 경우 음성언어 분석을 위한 태그가 필요하다. 대화체에서 자주 발생하는 ‘뭐지’, ‘뭐라고’, ‘해줘’ 등은 ‘무엇이지’, ‘무어라고’, ‘무엇이라고’, ‘하여주어’의 준말로써 원형으로부터 음성 인식을 수행하기 힘들 뿐만 아니라 그러한 노력을 통하여 반복된 되돌림 작업을 수행할 필요가 없으므로 처음부터 구어체 기반의 사전을 구성하여 준말을 그대로 인식단위로 사용하는 것이 경제적이다. ‘뭐’, ‘무어’의 경우는 하나의 지시 형용사로 보며, 이처럼 발음상 정확한 줄임말 표현이 되어 정보를 서로 공유될 수 있는 것들이 있는가 하면 ‘마지막 계 뭐지’에서 ‘계’의 경우 ‘것’의 준말로 실질 형태소와 형식 형태소가 발음상의 경계가 없어 나타나게 되고 또한 그 음운변화상의 규칙을 찾을 수가 없게 되어 ‘명사+조사’의 구성으로 품사를 설정한다는 것이 매우 모호해진다. 현재는 이와 같은 경우에 대해 조사가 생략된 명사 형태로 보고 실험을 수행하였으나, 앞으로 준말에 대한 새로운 품사의 선정 및 도입이 필요하며 자주 사용되는 어휘에 대한 준말사전이 필요하게 된다. 이에 따른 규칙이 설정되면 대화체 문장의 분석이 용이해질 것이다.

4 의사 형태소 단위의 한국어 연속음성 인식

본 논문에서 제안한 음성언어 형태소 해석을 수반하는 의사 형태소 단위의 한국어 연속음성 인식 시스템의 구성도는 [그림 2]와 같다.



[그림 2] 전체 시스템 구성도

한국어의 경우 어형의 변화와 첨가가 심하고 음운변화 현상이 다양하기 때문에 많은 음소 문맥이 고려된다. 따라서 조음현상에 민감하고 공유성을 통해 혼련성의 단점을 극복할 수 있는 트라이폰과 같은 문맥중속 단위를 대용량 어휘 인식 시스템을 구현하기 위한 인식단위로 보았다. 데이터 부족현상 발생을 해결하기 위해 조음 방식에 따른 분류에 따라 자음군들의 노드를 공유하도록 선정한 Tied-state 트라이폰을 인식단위로 삼았다.

인식기는 입력 발화에 대해 발음 사전을 참조하여 확률적으로 높은 HMM 모델 순서와 매칭이 되는 의사 형태소를 인식하게 된다. 의사 형태소 단위로 인식이 이루어지면 언어 모델을 이용하여 문장을 구성해 나간다. 이 때 언어 모델을 사용하는 것은 특정 단어와 단어의 결합 가능도가 정의된 문법에 의해 제한함으로써 탐색해야 할 단어의 수를 줄일 수 있을 뿐만 아니라 좀 더 정확한 인식이 이루어지도록 하기 위해서이다. 문장 인식 단계를 거쳐 인식된 문장을 기반으로 형태소 해석을 수행하게 된다.

#### 4.1 한국어 음운변화 현상을 반영한 발음사전의 구성

음운변화 현상을 반영하기 위한 처리 방법에는 대표적으로 두 가지 방법을 들 수 있다. 하나는 인식된 음소열 간의 결합을 보고 음운변화 역추적 규칙을 사용하여 예상되는 표제어를 사전에서 찾아내는 규칙적 처리 방법과 표제어에 대해 가능한 모든 발음열을 사전에 기재하여 사전에 정의된 표제어 단위로 인식하는 사전적 정의에 의한 처리를 들 수 있다. 본 논문에서는 다중 발음열 발음사전을 사용한 사전적 처리 방식으로 한국어 음운변화 현상을 처리하였다.

Short pause와 silence를 포함하여 45개로 구성된 유사 음소 단위(PLU, phone-like unit)로 발음사전을 표현하였으며, 이중 모음의 경우 반모음과 단모음의 결합으로 보았다. 음성언어의 경우 두 개의 형태소가 하나의 어절을 이룰 때 음운이 생략되거나 축약되는 현상이 빈번히 발생하므로 형태소의 경계가 모호하게 된다. 이때 축약되는 어절의 경우 음운적으로 볼 때 이중모음이 대부분이므로 반모음 이후를 중심으로 형태소의 경계를 삼기 위해 반모음과 단모음의 결합형태로 본다.

표제어 내부에 대해서는 18가지 음운변화 현상[2]을 모델링 하였으며 표제어 간, 즉 의사 형태소 간에 발생하는 음운변화 현상은 크게 체언과 조사, 어간과 어미 또는 보조 용언, 서술성 보통명사와 동사파생 접미사 간에 생기는 음운변화 현상을 모델링 하여 발음사전을 구성하였다. 또한 용언의 불규칙 현상의 경우, 발음에 기반한 어간과 어미의 이행태를 사전에 등록하여 해결하는 방식으로 발음사전을 모델링 하였다. 다음은 표제어간 음운변화 현상을 고려한 발음사전의 예이다.

[표 2] 표제어간 음운변화 현상을 고려한 발음사전의 구성 예

표제	발음열	태그	표제	발음열	태그
크	KH WW	pa	다	D AA	ef
맑	M AA N	pa		TH AA	
작	TS AA KQ	pa		TT AA	
고	G OW	ecc	게 하	G EY HI AA	px
	KH OW			KH EY HI AA	
	KK OW			KK EY HI AA	

위의 표에서 보는 바와 같이 어미나 조사의 경우 그 수가 한정되어 있기 때문에 사전에 증가되

(제 10 회 한글 및 한국어 정보처리 학술대회)

는 발음열은 한계가 있으나 대용량 인식 측면에서는 가능한 발음을 네트워크에 모두 깔아놓고 인식한다는 점에서 비효율적인 면이 있다. 따라서, crossword triphone 과 같이 인식시 음소문맥을 고려한 dynamic triphone 을 생성하며 인식하는 방법이 요구된다[3].

음성언어 사전은 형태소 해석시 사용하기 위해 발음사전을 확장하여 구성하며 의사 형태소 태그 정보 뿐만 아니라 사용될 수 있는 정보들을 포함함으로써 후처리시 용이하게 사용할 수 있다.

4.2 제안된 음성언어 형태소 해석

[그림 2]의 하단부는 Tree-Trellis 알고리즘[1]에 기반하여 인식된 N 개의 최적 문장을 탐색하며 음성 언어 형태소 해석을 수행하는 부분이다. 여기서는 인식된 의사 형태소 격자를 기반으로 인식을 수행하게 되며, 각 격자에는 시작 시점과 끝시점, 발견 가능도(Probability)가 포함되어 있다.

기본적으로 두가지 탐색 단계를 거친다. 첫번째 단계는 Modified Viterbi 알고리즘의 변형인 Trellis search 알고리즘으로 전향·시간 동기적 탐색을 수행하며, 결과로 N 개의 최적 격자들을 생성하고 그에 대한 부분 경로(Partial Path)와 Score 를 저장한다. 이때 음성언어 사전에서 표제어의 품사를 찾아낸 뒤 서로 인접한 의사 형태소 쌍이 결합 가능한지를 의사 형태소 접속정보를 참조하며 수행한다. 두번째 단계는 A\* 알고리즘 또는 stack 알고리즘의 변형인 tree search 알고리즘으로 후향·시간 비동기적 탐색이다. 전방향 탐색에서 구해진 부분 경로값과 현재까지의 역방향 부분 경로값의 합을 사용하여 전 경로(Full Path)에 대한 정확한 확률값을 알 수 있게 되므로 불필요한 경로의 확장이 줄게 된다. 이를 기반으로 격자들의 결합을 문장 결과로 내놓는다. 이때 격자가 가지고 있는 시간 정보와 의사 형태소 태그를 함께 출력한다.

5 실험 및 평가

5.1 실험환경 및 문장구성

연속 HMM 을 기반으로 한 화자 독립 시스템으로 본 실험에서는 16 KHz 로 샘플링된 음성신호에 대해 25msec 의 해밍 윈도우를 사용하여 프레임 분석을 한 후, 프레임 단위로 13 차 mel-

frequency cepstral coefficients, 13 차 delta-cepstral coefficients, 13 차 delta-delta cepstral coefficients 로 구성된 특징벡터를 학습 및 인식을 위한 벡터로 사용하였다. 단어 인식 과정에서는 발음사전과 서브워드 모델을 사용한다. 발음사전은 본문에서 설명한 바와 같이 구성하였으며 서브워드 모델은 각 유사 음소에 대해 학습된 HMM 이다. 이 HMM 은 유사 음소와 같은 모노폰의 형태에서 확장된 음소 문맥 정보를 포함하는 트라이폰으로 학습되어졌다. 단어인식이 이루어지면 언어모델을 이용하여 문장 인식을 수행한다. 여기서는 back-off bigram 모델을 사용하여 실험을 수행하였다.

실험 A 는 호텔예약 영역 29 종류의 문장에 대해 5 명의 화자가 5 번씩 발화한 725 문장과 추가로 8 종류의 문장에 대해 앞에 화자 중 동일 1 명과 다른 2 명의 화자가 5 번씩 발화한 120 문장에 대한 실험으로 총 845 문장에 대한 실험 결과이며, 실험 B 는 위의 학습 참여자 7 명의 학습에 참여하지 않은 169 개의 발화문장에 대한 인식실험 결과이다.

문장을 구성하는 디코딩 단위의 선정 기준은 다음과 같다. 어절은 띄어쓰기 단위이며, 단어(i) 은 음성인식 실험에서 일반적으로 사용되는 단위로 용언구를 제외한 어절에 대해 내용어와 기능어를 분리한 단위이며, 단어(ii)는 단어(i)과 같은 구성으로 단어의 조합으로 이루어지는 모든 단위를 나눈 경우이며, 형태소는 문자기반의 형태소로 5 가지의 경우에 대해 실험을 수행하였다.

5.2 인식률 및 사전크기 비교

[표 3] 실험 A 의 문장 인식 결과

인식률(%)/디코딩 단위		어절	단어(i)	단어(ii)	의사 형태소	형태소
		단어	Correctness	98.22	96.87	95.81
	Accuracy	96.93	95.08	92.68	94.52	88.94
문장	Correctness	91.83	83.67	71.95	79.88	53.14

[표 4] 실험 B 의 문장인식 결과

인식률(%)/디코딩 단위		어절	단어(i)	단어(ii)	의사 형태소	형태소
		단어	Correctness	98.02	96.15	94.51
	Accuracy	96.53	93.29	90.36	92.03	86.84
문장	Correctness	91.02	79.88	66.27	69.82	49.11

는 발음열은 한계가 있으나 대용량 인식 측면에서는 가능한 발음을 네트워크에 모두 갈아넣고 인식한다는 점에서 비효율적인 면이 있다. 따라서, crossword triphone 과 같이 인식시 음소문맥을 고려한 dynamic triphone 을 생성하며 인식하는 방법이 요구된다[3].

음성언어 사전은 형태소 해석시 사용하기 위해 발음사전을 확장하여 구성하며 의사 형태소 태그 정보 뿐만 아니라 사용될 수 있는 정보들을 포함함으로써 후처리시 용이하게 사용할 수 있다.

#### 4.2 제안된 음성언어 형태소 해석

[그림 2]의 하단부는 Tree-Trellis 알고리즘[1]에 기반하여 인식된 N 개의 최적 문장을 탐색하며 음성 언어 형태소 해석을 수행하는 부분이다. 여기서는 인식된 의사 형태소 격자를 기반으로 인식을 수행하게 되며, 각 격자에는 시작 시점과 끝시점, 발견 가능도(Probability)가 포함되어 있다.

기본적으로 두가지 탐색 단계를 거친다. 첫번째 단계는 Modified Viterbi 알고리즘의 변형인 Trellis search 알고리즘으로 전향·시간 동기적 탐색을 수행하며, 결과로 N 개의 최적 격자들을 생성하고 그에 대한 부분 경로(Partial Path)와 Score 를 저장한다. 이때 음성언어 사전에서 표제어의 품사를 찾아낸 뒤 서로 인접한 의사 형태소 쌍이 결합 가능한지를 의사 형태소 접속정보를 참조하며 수행한다. 두번째 단계는 A\* 알고리즘 또는 stack 알고리즘의 변형인 tree search 알고리즘으로 후향·시간 비동기적 탐색이다. 전방향 탐색에서 구해진 부분 경로값과 현재까지의 역방향 부분 경로값의 합을 사용하여 전 경로(Full Path)에 대한 정확한 확률값을 알 수 있게 되므로 불필요한 경로의 확장이 줄게 된다. 이를 기반으로 격자들의 결합을 문장 결과로 내놓는다. 이때 격자가 가지고 있는 시간 정보와 의사 형태소 태그를 함께 출력한다.

### 5 실험 및 평가

#### 5.1 실험환경 및 문장구성

연속 HMM 을 기반으로 한 화자 독립 시스템으로 본 실험에서는 16 KHz 로 샘플링된 음성신호에 대해 25msec 의 해밍 윈도우를 사용하여 프레임 분석을 한 후, 프레임 단위로 13 차 mel-

frequency cepstral coefficients, 13 차 delta-cepstral coefficients, 13 차 delta-delta cepstral coefficients 로 구성된 특징벡터를 학습 및 인식을 위한 벡터로 사용하였다. 단어 인식 과정에서는 발음사전과 서브워드 모델을 사용한다. 발음사전은 본문에서 설명한 바와 같이 구성하였으며 서브워드 모델은 각 유사 음소에 대해 학습된 HMM 이다. 이 HMM 은 유사 음소와 같은 모노폰의 형태에서 확장된 음소 문맥 정보를 포함하는 트라이폰으로 학습되어졌다. 단어인식이 이루어지면 언어모델을 이용하여 문장 인식을 수행한다. 여기서는 back-off bigram 모델을 사용하여 실험을 수행하였다.

실험 A 는 호텔예약 영역 29 종류의 문장에 대해 5 명의 화자가 5 번씩 발화한 725 문장과 추가로 8 종류의 문장에 대해 앞에 화자 중 동일 1 명과 다른 2 명의 화자가 5 번씩 발화한 120 문장에 대한 실험으로 총 845 문장에 대한 실험 결과이며, 실험 B 는 위의 학습 참여자 7 명의 학습에 참여하지 않은 169 개의 발화문장에 대한 인식 실험 결과이다.

문장을 구성하는 디코딩 단위의 선정 기준은 다음과 같다. 어절은 띄어쓰기 단위이며, 단어(i) 은 음성인식 실험에서 일반적으로 사용되는 단위로 용언구를 제외한 어절에 대해 내용어와 기능어를 분리한 단위이며, 단어(ii)는 단어(i)과 같은 구성으로 단어의 조합으로 이루어지는 모든 단위를 나눈 경우이며, 형태소는 문자기반의 형태소로 5 가지의 경우에 대해 실험을 수행하였다.

#### 5.2 인식을 및 사전크기 비교

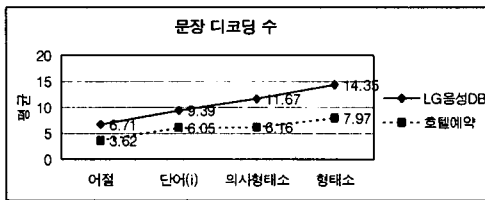
[표 3] 실험 A 의 문장 인식 결과

디코딩 단위 인식률(%)		어절	단어(i)	단어(ii)	의사 형태소	형태소
		단어	Correctness 98.22	96.87	95.81	96.55
	Accuracy 96.93	95.08	92.68	94.52	88.94	
문장	Correctness 91.83	83.67	71.95	79.88	53.14	

[표 4] 실험 B 의 문장인식 결과

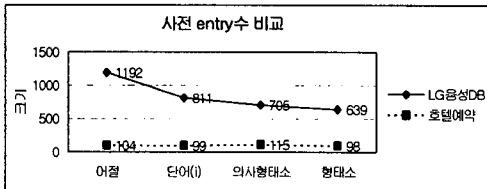
디코딩 단위 인식률(%)		어절	단어(i)	단어(ii)	의사 형태소	형태소
		단어	Correctness 98.02	96.15	94.51	94.82
	Accuracy 96.53	93.29	90.36	92.03	86.84	
문장	Correctness 91.02	79.88	66.27	69.82	49.11	

실험 결과를 보면 전반적으로 디코딩 단위에 따른 단어 인식률의 차이는 크게 나타나지 않았다. 반면 문장 인식률 면에서 보면 형태소를 디코딩 단위로 삼았을 경우에 인식률이 현저하게 떨어지는 것을 볼 수 있다. 이는 짧은 음성신호상의 문제일 뿐 아니라, 하나의 문장을 구성하게 되는 디코딩 단위의 개수에 따라 영향을 받은 것으로 구성수가 n 배로 증가되는 경우 탐색영역은  $2^n$  배로 증가하게 된다. 따라서 계산량이 많아짐에 따라 속도가 떨어지고 인식률의 정확도가 떨어지게 된다. 참고로 한 문장을 구성하는 평균 디코딩 수는 다음과 같다.



[그림 3] 문장을 구성하는 평균 디코딩 수

음성 인식기에서 고려되는 것은 인식률 뿐만 아니라 실시간 처리를 위해 그 인식속도도 중요한 고려 대상이 된다. 인식기에서의 인식속도는 일반적으로 사전의 크기에 비례하게 된다. 다음은 사전의 표제어 단위에 따른 사전 entry 수이다. 호텔예약 영역의 경우 문장의 수가 적기 때문에 사전의 크기를 비교하기에 적당하지 않으므로 대용량 음성 인식시 발생하는 문제를 살펴보기 위해 'LG 음성 DB' 중에서 1000 문장을 선정하여 분석하였다.



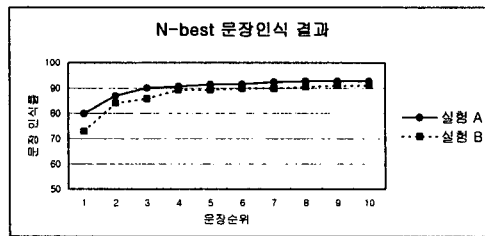
[그림 4] 사전의 표제어 단위에 따른 사전 크기

어절 단위의 경우 사전 표제어 수가 가장 크며 형태소의 경우 가장 적다. 1000 문장에 대한 크기 비교이지만 대규모 음성 인식 실험에서는 기능어의 수가 한정되어 있으므로, 사전에는 고유명사나 용언의 어간과 같은 새로운 미등록어가 첨가되므로 어절과 형태소 사이의 사전크기가 크게

차이날 것을 추측할 수 있다.

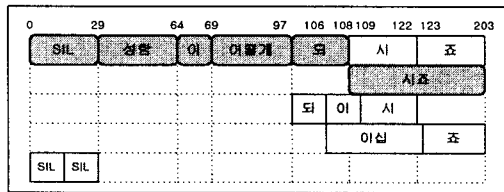
### 5.3 음성언어 형태소 해석 수행

다음 [그림 5]는 의사 형태소 단위로 인식 실험을 수행하였을 경우 N-best 문장인식 결과이다. 문장 순위를 아무리 올려도 인식률의 증가에는 한계가 있음을 볼 수 있다. 이는 기본적으로 인식이 어려운 사항들이 포함되어 있기 때문이다. 따라서 형태소 해석시 상위 5 문장을 기반으로 실험을 수행하였다.



[그림 5] 실험 A, B의 N-best 문장인식 결과

다음은 실험에 사용된 문장 중 하나로 “성함이 어떻게 되시죠?”라는 실제 음성파일을 입력으로 하여 수행된 음성언어 형태소 해석 과정이다.



[그림 6] 의사 형태소 격자

[그림 6]은 인식된 상위 5 번째까지의 문장 결과를 기반으로 재구성된 중간 탐색 격자이다. 앞장에서 기술된 tree-trellis 알고리즘에 기반하여 위의 예제 문장에 대해 다음과 같은 형태소 해석 결과를 얻었다.

성함/ncn+i/jcc+어떻게/mad+되/pvg+시죠/ef

## 6 결론 및 향후 연구과제

본 논문에서는 형태소 해석을 수반하는 한국어 연속음성 인식 시스템을 구현하였다. 음성언어 형태소 해석을 위해 문장을 구성하는 새로운 디



코딩 단위가 필요함에 따라 의사 형태소를 정의하였고, 이를 기반으로 인식실험을 수행하였다. 디코딩 단위가 어절이나 단어의 경우에는 의사 형태소나 형태소 보다 인식을 면에서는 좋으나, 형태소 결과를 얻기 위해서는 별도의 후처리 과정을 거쳐서 형태소 분석을 수행해야만 한다. 그러나 의사형태소나 형태소로 인식하는 경우에는 별도의 형태소 분석 과정 없이 형태소 단위로 결과를 얻을 수 있는 이점이 있다. 디코딩 단위가 문자기반의 형태소인 경우에는 발생시간이 짧은 기능어로 인해 인식이 현저히 낮았으나 이러한 문제점을 보완한 의사 형태소의 경우 약 20% 이상의 인식을 향상을 얻었다. 상위 5 개의 인식 문장에 대해 재탐색시 한국어 대화체 문형의 특성을 고려한 접속정보를 사용하여 가장 올바른 형태소 해석 결과를 탐색하였다.

현재 'KAIST 무역상담 DB'를 대상으로 하여 본 논문에서 제안한 의사 형태소 디코딩 단위를 채택한 대규모 연속음성 실험이 진행중이며, 이를 통해 격자 탐색시 필요한 안정된 형태소 통계 정보와 더 나아가 상위 언어 모델의 구축이 기대된다. 앞으로는 보다 많은 대화체 문형에 대한 연구를 통해 현재 다소 영역 의존적인 접속 정보를 영역 독립적인 형태소 접속정보로 규칙화 할 계획이다.

#### 참고문헌

- [1] F. K. Soong and E.-F. Huang, "A Tree-Trellis Based Search for Finding the N-best Sentence Hypothesis in Continuous Speech Recognition", *Proceeding of the DARPA Speech and Natural Language Workshop*, pp12-19, 1990.
- [2] Jehun Jeon, Sunhee Wee, and Minhwa Chung, "Generating Pronunciation Dictionary by analyzing Phonological Variations Frequently Found in Spoken Korean", *Proceedings of International Conference on Speech Processing*, 1997.
- [3] Julian James Odell, *The Use of Context in Large Vocabulary Speech Recognition*, Ph.D Thesis in Queens's College, 1995.
- [4] Kai-Fu Lee, *Automatic Speech Recognition : The Development of the SPHINX System*, Kluwer Academic Publisher, Boston, 1989.
- [5] 강현석, 정민화, "의도분석을 이용한 대화체 연속음성인식," *제 14회 한국 음향학회 학술 발표대회 논문집*, pp.63-66, 1997.
- [6] 권오욱, 박준, 황규웅, "의사 형태소 단위 대어휘 연속 음성 인식기 개발," *제 15회 음성통신 및 신호처리 워크샵 논문집*, pp.320-323, 1998.
- [7] 이경남, 정민화, "의사 형태소 단위의 연속 음성 인식," *제 15회 음성통신 및 신호처리 워크샵 논문집*, pp.309-314, 1998.
- [8] 이상호, 서정연, 오영환, "KTS:미등록어를 고려한 한국어 품사 태깅 시스템," *제 12회 음성통신 및 신호처리 워크샵 논문집*, pp.195-199, 1995.
- [9] 이승배, 이종석, "N-Best 문장탐색기법을 적용한 연속음성인식 시스템," *제 13회 음성통신 및 신호처리워크샵 논문집*, pp.151-154, 1995.
- [10] 최기선, 남영준, 김진규, 한영균, 박석문, 김진수, 이춘택, 김덕봉, 김재훈, 최경진, "한국어정보베이스를 위한 형태·통사 태그 표준에 관한 연구," *한국인지과학회 논문지*, Vol.7, No.4, pp.43-61, 1996.
- [11] 최준기, 이근배, 이종혁, "HMM에 기반한 연속음성인식에서의 형태소 그래프 생성," *제 9회 한글 및 한국어 정보처리 학술대회*, pp.500-504, 1997.