제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

# 의사형태소 단위 대어휘 연속 음성 인식기 개발

권오욱, 박준, 황규웅

한국전자통신연구원 휴먼인터페이스연구부 음성언어팀

# Development of a Pseudomorpheme-Based Large Vocabulary Continuous Speech Recognizer

Oh-Wook Kwon*, Kyuwoong Hwang, and Jun Park

Spoken Language Processing Team, ETRI

{owkwon, hkw, junpark}@etri.re.kr

## 요약

대어휘 연속음성인식을 목표로 개발한 의사형태소 단위의 인식기를 기술하였다. 먼저 의사형태소를 정의하고, 의사형태소 태거를 간단히 기술하며, 의사형태소의 병합에 의한 인식단위 결정 방법, 의사형태소 단위 인식기에서 특히 고려되어야 할 음향모델링, 품사 정보를 이용한 언어모델 및 어절규칙의 적용 방안, 의사형태소 단위 인식을 위한 새로운 탐색기 구조를 기술한다. 약 5,500 어절의 인식어휘를 갖는 여행계획 영역의 대화체 연속음성 데이타베이스를 이용하여 초벌 인식실험을 한 결과, 의사형태소 단위의 인식기의 단어인식률은 66.4%, 어절인식률은 60.0%를 나타내었다.

## 1. 서론

한국어 대어휘 연속음성인식을 위한 인식어휘 단위로는 음소, 음절, 형태소, 단어, 어절이 가능하다. 음소 단위 인식기의 경우 어휘 개수는 음소 개수와 같으므로 인식과정은 단순하나, 인식단위의 평균 지속시간이 짧고, 음소간의 언어모델이 적용되므로 인식률이 저하되며, 단어 단위 인식결과를 얻기 위한 음소 격자의 처리가 필요하다. 어절 단위 인식기에 있어서는 텍스트 코퍼스의 양이 제한되므로 강인한 언어모델을 구하기가 어렵지만, 인식단위의 평균 지속시간이 길어지고 탐색기에서 넓은 범위의 문맥을 고려 할 수 있으므로 인식률이 향상된다. 그러나 인식단어가 증가할수록 모든 종류의 조사 및 어미가 결합된 어절을 인식어휘에 넣어야 하므로 탐색 공간이 증가하고, 어휘외단어(out-of-vocabulary word)가 증가하고, 언어모델의 강인성이 저하되므로 대어휘 인식기의 인식단위로는 적합하지 않다.

이러한 단점을 극복하기 위하여 대어휘 연속음성인식을 목표로 하는 본 연구에서는 의사형태소를 이용한 인식기를 사용하였다. 의사형태소는 주어진 어절의 소리 값을 유지하는 범위 내에서의 언어학적인 형태소를 말한다. 즉 분리된 의사형태소들의 단순 결합에 의해서 원래의 소리 값을 찾을 수 있음을 의미한다. 의사형태소를 그대로 인식단위로 사용할 경우 'ㅂ', 'ㄹ', 'ㅆ'과 같은 단음소 또는 '시', '았', '었'과 같은 단음절의 형태소가 증가하기 때문에 인식오류가 증가한다. 따라서 하나의 어절을 내용어와 기능어로 분리하여 이들을 인식단위로 한다.

제 2 절에서는 의사형태소 단위 인식기의 구조를 기술한다. 제 3 절에서는 대화체 여행계획 영역의 음성 데이터베이스를 사용하여 실험한 결과를 기술하고 토의를 한다. 제 4 절은 연구결과의 간단한 요약 및 향후 계획으로 구성된다.

## 2. 의사형태소 단위 연속음성인식기

### 2.1. 연속음성인식기

의사형태소 단위 연속음성인식기의 구조는 그림 1과 같다. 입력된 음성은 특징추출 모듈에서 특징벡터로 변환되고 탐색모듈에서 음향모델과 발음사전, 언어모델을 이용하여 가장 확률이 높은 단어열을 얻게 된다. 탐색기는 대어휘 인식을 위하여 트리 구조를 갖는다. 의사형태소 단위 인식기에서는 발음사전, 언어모델이 의사형태소를 수용하기 위하여 변경되어야 한다. 후처리모듈은 인식결과를 어절 단위로 모아 쓰고, 태그와 잡음기호를 제거하여 최종적인 텍스트를 출력한다.

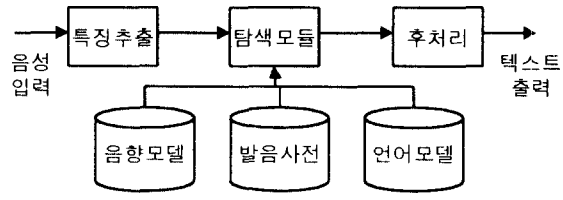


그림 1. 연속음성인식기의 구조

### 2.2. 특징추출

본 연구에서의 특징벡터로는 인간의 청각특성을 잘 반영하는 것으로 알려진 PLP (percetual linear prediction)[1]을 사용하였다. 음성의 동적특성을 반영하기 위하여 PLP계수의 1차 및 2차 미분치에 해당하는 벡터를 추가하였다. 최종적으로 인식에 사용되는 특징벡터의 차원을 줄이기 위하여 LDA (linear discriminant analysis)[2]를 적용하였다.

### 2.3. 음향모델

40개의 음소모델을 음향모델의 기본 단위로 사용하였다. 각 음소는 좌우의 문맥에 따라서 다른 파라미터 값을 갖는 3개의 상태로 구성된 HMM (hidden Markov model)으로 모델링 되었다. 각 상태에서의 특징벡터에 대한 관측확률은 가우시안 확률밀도함수로 주어진다. 이 시스템에서는 다수의 코드북이 사용되며, 각 상태마다 어떤 코드북을 사용할지가 결정되어 있으며, 코드북내의 코드워드에 대한 가중치는 모든 문맥의존 음소마다 달라진다. 즉 하나의 코드북은 다수의 모든 문맥의존 음소에서 공유된다.

인식기가 실제환경에서 사용되기 위해서는 불가피하게 발생하는 잡음에 대처하여야 한다. 여기에서는 숨소리, 주변잡음, 입술소리, 긴 묵음구간, 기침소리, 간투어 처리를 위하여 1개의 상태를 갖는 HMM으로서 이들을 모델링하였다.

현재의 인식기에서는 단어내에서는 3개까지 단어간에서는 1개까지의 음소문맥을 고려하며, 단어경계를 나타내는 정보도 동시에 이용한다.

### 2.4. 의사형태소

의사형태소는 주어진 어절의 소리값을 유지하는 범위 내에서의 언어학적인 형태소를 말한다. 즉 분리된 형태소들의 단순 결합(concatenation)에 의해서 원래의 소리값을 찾을 수 있음을 의미한다. 의사형태소는 일반적인 형태소와 매우 유사하나, 형태소의 분리에 있어서 소리값이 유지된다는 점이 매우 다르다. 따라서, 불규칙이나 음운 현상을 처리하는 데 있어서 소리값이 유지되도록 그 기준을 정한다. 의사형태소는 가능한 한 하나의 어절에 대해서 내용어와 기능어로 분리함을 원칙으로 한다. 그 분리가 의사형태소의 정의에 벗어날 경우에는 그대로 둔다. 예를 들어 '써서'라는 어절은 일반적인 형태소에서는 '쓰+어서'로 분리되나, 의사형태소에서는 '써/pvg(EUI)+서'로 분리된다. "말할 <u>게</u> 아무 것도 없다"에서 '게'의 원형은 '것+이'이다. '에'를 복모음으로 생각하고 '어+이'로 분리할 수 있다. 그러나 본 연구에서는 복모음에 대해서 분리하지 않으므로 '게'는 분리할 수 없다. 따라서 '게'는 그대로 둔다.

의사형태소는 일반적인 형태소와 달리 형태소들의 원형을 밝힐 수 없는 경우가 많으며, 또한 그들의 품사 결정도 매우 모호하다. 첫째 원형을 분명하게 말할 수 없을 경우에는 내용어의 품사를 따른다. 둘째 불규칙 현상이나 음운 현상에 의해서 분리된 용언의 경우에는 어간의 품사를 따른다. 이와 같은 원칙은 기형적인 형태소 배열 규칙을 만들게 된다. 예를 들면, 일반적으로 동사는 자신만으로 문장 내에서의 어떤 성분 역할을 할 수 없다. 그러나, 의사형태소의 경우에는 동사가 어미를 가지지 않을 수 있다. 또한 많은 경우에 명사 바로 다음에 어미를 동반하게 된다.

의사형태소 단위로 인식을 수행한 경우 한국어의 특성에도 맞고, 동일한 태스크에 대하여 필요한 인식단어의 개수도 감소하며, 어휘외단어가 감소하는 장점이 있다. 그 반면에 인식 오류는 증가하게 된다. 그 이유는 실제 발음과 분리된 형태소의 발음이 달라지며, 관형형어미 'ㄴ', 'ㄹ', 과거시제 선어말어미 '었'과 같은 단음소 및 음절 한 개짜리의 형태소가 많이 발생하여 삽입/삭제 오류가 많아지며, 현재의 탐색기 구조상의 제약에 의하여 인식단위간(interword)의 음향모델을 2개 이상은 적용할 수 없기 때문이다.

### 2.5. 의사형태소 태거

어절을 의사형태소 단위로 분리하기 위하여 태거를 개발하였다. 기존의 일반 형태소 태거를 최소한으로 수정하여 구현하기 위하여 발음이 변화되는 동사 및 형용사의 어간을 형태소분석기가 사용하는 사전에 추가하고, 대화체에서 특이하게 나타나는 형태소간 연결규칙을 보완하였다.

### 2.6. 인식단위 결정

의사형태소 단위 인식기의 인식률 저하를 막기 위하여 의사형태소를 적절히 결합한 확장의사형태소를 인식단위로 하였다.

독립적으로 사용될 수 있는 내용어는 하나의 인식단위로 한다. 하나의 내용어 뒤에 나타나는 조사들은 하나의 인식단위로 결합된다. 용언의 어미 부분은 모두 결합되어 하나의 인식단위로 된다. 접미사는 하나의 단위로 된다. 다만 동사형접미사 '하'의 경우 동작성/상태성 명사에 붙어 동사로서 빈번히 사용되므로 이를 어미와 결합하여 새로운 인식단위로 한다. 관형형어미는 보통 단음소로 이루어지는 경우가 많아서 인식오류의 주원인이 되므로 관형형어미를 나타내는 단음소 형태소는 바로 앞의 형태소와 결합하여 새로운 인식단위로 된다. 기본 의사형태소를 사용한 인식기에서 인식오류를 많이 알으키는 것으로 보조용언이 있다. 한국어에서 보조용언은 앞에 오는 본용언에 기대어 사용되며 빈도가 상당히 높다. 보조용언의 어간은 한 음절짜리가 대다수이므로 이들도 어미와 결합하여 인식단위로 한다.

### 2.7. 발음사전

확장의사형태소 단위 인식기에서 사용되는 발음사전은 일반적인 인식기에서의 발음사전과 유사하다. 그러나, 같은 이름을 가지는 확장의사형태소가 복수의 품사를 가질 수 있으며, 하나의 내용어도 뒤에 나타나는 조사나 어미에 따라서 발음이 달라지는 경우가 있으므로 이를 고려한 다중발음사전이 필요하다.

기능어의 경우 앞 단어와 연결되어 발음되는 경

우가 대부분이므로 단어 경계를 나타내는 태그를 붙이지 않는다. 그림 2는 발음사전의 일부를 보여주고 있다. 여기에서 'WB'는 단어경계를 나타내는 태그이며, '/'뒤에 이름은 그 단어의 품사를 나타낸다. 탐색기에서는 음향모델의 점수를 계산할 때는 품사정보와 다중발음 정보를 포함하는 전체 단어 이름으로 식별하며, 언어모델의 적용에서는 다중발음 정보를 제거한 단어 이름으로 식별한다. 품사정보는 품사언어모델 및 어절규칙의 적용에 이용된다.

'먹/pvg'의 발음이 두 가지인 것은 '먹고', '먹는'이 /먹꼬/, /멍는/으로 발음되기 때문이다. 이와 같은 다중발음을 자동으로 구하기 위하여 하나의 어절을 의사형태소 단위로 분리하여 발음을 구하여야 한다. 이를 위하여 앞에서 기술한 의사형태소 태기와 형태소 단위의 발음열 변환기를 사용한다.

```
.
.
ㅂ니다/ef {m n i d {a WB}}
ㅆ다/ep_ef {d D {a WB}}
먹/ncn {{m WB} v g}
먹/pvg {{m WB} v g}
먹/pvg(2) {{m WB} v N}
하고/jcj {h a g {o WB}}
하고/px_ecc {h a g {o WB}}
하고/xsv_ecc {h a g {o WB}}
```

그림2. 확장의사형태소 단위 발음사전의 일부

### 2.8. 언어모델

본 시스템에서 사용하는 언어모델은 확장의사형태소 트라이그램을 기본으로 사용한다. 훈련 텍스트에서 나타나지 않은 트라이그램을 구하기 위하여 백오프 방법[3]이 사용된다. 단어수가 수만개로 증가할 경우 언어모델을 위한 메모리가 크게 증가하며, 트라이그램 값을 찾기 위한 탐색 시간이 길어지므로, 메모리 감소 및 고속 액세스 기법이 사용되어야 한다.

본 인식기에서는 품사의 연결관계를 언어모델에 이용한다. 발음사전에 이미 각 단어의 품사가 저장되어 있으므로 인식기의 탐색모듈에서 단어간 천이가 있을 경우 언어모델이 적용된다. 최종적인 언어모델 값은 확장의사형태소간 트라이그램과 품사 트라이그램에 의한 언어모델 값의 가중 결합으로 주어지며, 가중치는 여러 가지 값을 단계적으로 변화하여 가면서 최고 성능을 나타내는 것으로 선택한다.

### 2.9. 어절규칙

확장의사형태소 단위 인식기의 인식오류는 주로 조사, 어미와 관련된다. 인식단위의 평균길이는 늘어났지만 음향적으로 유사한 조사, 어미가 많기 때문이다. 이를 위하여 어절 내에서는 트라이그램에 의한 언어모델보다 더 엄격한 어절 형성 규칙을 적용한다. 이 규칙에 위배되는 단어간 천이의 경우에는 벌점을 부가하는 방식으로 구현된다. 벌점의 형태로 하지 않고 가능한 단어간 천이만을 탐색하는 경우에는 오히려 인식률의 저하가 발생한다. 어느 한 단어의 인식오류 때문에 그 다음 단어의 인식에도 영향을 미치기 때문이다.

적용된 어절규칙으로는 조사 앞에는 내용어만이 올 수 있으며, 어미는 용언 뒤에서만 오며, 단위성 의존명사는 수사 뒤에만 오며, 비단위성 의존명사는 관형형어미 뒤에 온다는 것 등이 있다. 앞 형태소의 받침 유무에 따라서 뒤에 나타나는 조사의 종류가 결정되는 현상(예를 들면, 은/는, 을/를), 모음조화 현상도 규칙으로 적용한다.

연속음성인식에서는 짧은 단어의 삽입오류를 줄이기 위하여 탐색기의 단어간 천이 시점에 단어삽입 벌점을 다르게 된다. 그런데 조사나 어미, 접미사, 단위성 의존명사 등의 경우에서는 바로 앞 단어와 이어서 발성되는 경우가 대부분이다. 따라서 이와 같은 어절내의 단어간 천이 부분에서는 단어삽입 벌점을 사용하지 않는다.

### 2.10. 의사형태소 단위 탐색기 구조

어절규칙 및 단어삽입 벌점의 선별적 적용을 위하여 그림 3과 같은 새로운 탐색기 구조를 제안한다.

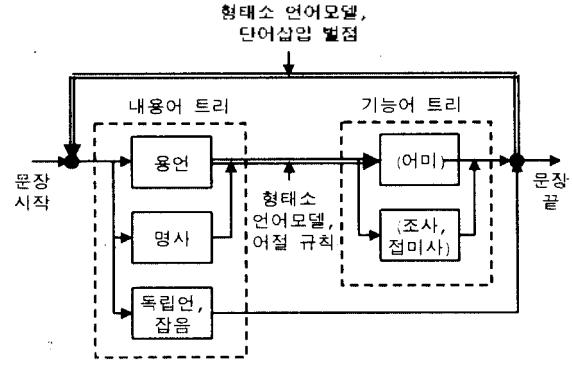


그림 3. 확장의사형태소 단위 인식기의 탐색기 구조

기존의 탐색기에서는 모든 단어들이 하나의 트리로 구성되어 있으며, 단어간 천이시에 언어모델 및 단어삽입 벌점이 부가되었다. 제안된 구조에서는 내용어와 기능어로 구분된 두개의 탐색 트리를 두었으며, 내용어에는 용언, 명사, 독립언, 잡음 등이 있고, 기능어에는 어미, 조사, 접미사가 속한다. 내용어와 기능이 트리간의 천이 시에는 형태소 단위 언어모델과 어절규칙이 적용되며, 내용어 간의 천이시 또는 기능어와 내용이 간의 천이 시에는 단어삽입 벌점과 형태소 단위 언어모델이 적용된다. 기능이 트리는 앞에 오는 내용어에 따라서 건너뛰어질 수도 있다.

## 3. 실험결과

### 3.1. 실험 환경

대어휘 인식을 위한 연속음성 인식기의 성능을 테스트하기 위하여 만적 5,500어절 규모의 대화체

여행계획 태스크의 음성 데이터베이스를 사용하였다. 훈련용 음성데이터는 819 대화, 6901 문장이었으며, 테스트용 음성데이터는 55 대화 283 문장이었다. 어휘외단어는 없는 경우에 대하여 실험하였다. 테스트 셋의 perplexity 는 어절 단위의 경우에는 95, 확장 의사형태소 단위로는 36 이었다. 인식용 발음사전의 엔트리 개수는 어절 단위에서 5,501 개, 의사형태소 단위에서는 3,157 개이었다.

본 연구에서는 어절 단위 인식기와의 성능비교와 인식단위의 변경에도 무관하도록 어절인식률을 성능 지수로 사용하였으며, 대치, 삽입, 삭제 오류를 모두 고려하였다.

### 3.2. 인식결과

그림 4 에 인식결과의 일부를 나타내었다. 의사형태소 단위 인식 결과는 후처리하기 전의 결과이다.

입력문장: 네 네명이 갈려고 하는데요 호텔 방을 예약할려고 하거든요
원래 의사형태소: +lip+/f 는데/ecs 애/ncn 명/xsn 이/jp 가/px 르려고/ecx 하/xsv 는데/ecs 어/ep 호텔/ncn 방/nbu 을/jis 예약/ncn 갖/pvg 르려고/ecx 드/pvg 씨내요/ef
확장 의사형태소: +huh+/f 내용/ncn 에/jca 갈려고/px_ecx 하는데요/xsv_ecs_jxf 호텔/ncn 방/ncn 을/jco 예약/ncpa 하려고/px_ecs 거든요/ef_jxf
어절: +huh+ 네명이 갈려고 하는데요 호텔 방을 예약할려고 하거든요

그림 4. 인식결과 예문

의사형태소 단위 탐색기의 단어간 천이시 log likelihood ( $L$ )는 다음과 같이 계산하였다.

$$L = (L_a + zL_{LM} + p) + z_{pos}L_{LM,pos} + p_{evp}$$

여기서 $L_a$ 는 음향모델에 의한 값이며, $z$ 는 의사형태소단어 단위 언어모델의 가중치, $z_{pos}$ 는 품사언어모델 값의 가중치, $p$ 는 단어삽입 벌점, $p_{evp}$ 는 어절규칙에 의한 벌점이다. 원래의 의사형태소 단위 인식기의 성능은 상당히 저조하여 더 이상 인식률을 측정하지는 않았다. 표 1 은 확장의사형태소 단위 인식기의 $z$ 와 $p$ 의 변화에 따른 인식률을 나타낸 것이며, 표 2 는 최대성능을 보인 $z=28$ 과 $p=5$ 에 대하여 $z_{pos}$ 의 변화에 따른 인식률을 나타낸 것이다. 표 2 로부터 확장의사형태소 인식기의 인식률은 최대 58.4%로 나타났다. 이 태스크에 대하여 어절 단위 인식기의 인식률은 79.6%이었다. 확장의사형태소 단위 인식률은 어절 단위 인식기의 인식률보다 상당히 낮으나, 확장의사형태소 단위의 인식기는 인식어휘의 증가에 따라서 어휘외단어가 비교적 적고, 대어휘 인식을 위해서는 필수적인 인식단위이다.

표 1. $z$ 와 $p$ 에 따른 확장의사형태소 단위 인식기의 어절인식률(%)

| $z$ | 16 | 20 | 24 | 28 | 32 |
|---|---|---|---|---|---|
| $p=5$ | 55.9 | 55.7 | 56 | 55.3 | 53.6 |
| $p=10$ | 55.5 | 55.8 | 55.9 | 55.1 | 53.6 |
| $p=15$ | 55.6 | 55.3 | 55.6 | 54.8 | 53.2 |

표 2. $z_{pos}$ 에 따른 확장의사형태소 단위 인식기의 어절인식률(%)

| $z_{pos}$ | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
|---|---|---|---|---|---|---|
| 어절 인식률 | 55.3 | 57.0 | 57.7 | 57.6 | 58.1 | 58.4 |

최대 성능을 나타내는 확장의사형태소 단위 인식기에 어절 규칙을 적용한 결과, $p_{evp}=10$ 일때 60.0%의 어절인식률을 얻었다. 확장의사형태소를 단어로 취급할 경우에 단어인식률은 66.4%이었다. 이 결과로 볼 때, 어절규칙에 의한 성능향상은 아직 연구의 여지가 있다.

## 4. 맺는말

대어휘 연속음성인식을 목표로 의사형태소를 언어학적 지식을 이용하여 연결함으로써 향상된 인식률을 나타내는 확장의사형태소 단위의 초벌 인식기를 개발하였다. 대화체 여행계획 영역의 음성데이타베이스를 사용하여 실험한 결과 60.0%의 어절인식률을 나타내었다.

이를 바탕으로 낭독체 수만단어 규모의 어휘를 갖는 연속음성인식 엔진을 개발할 예정이다. 인식률 향상을 위하여 인식단위의 결합을 언어학적 지식에 의존하지 않는 방법을 연구할 계획이다.

### 감사의 글

이 연구는 정보통신부의 지원에 의해 이루어진 결과물입니다.

### 참고문헌

[1] H. Hermansky, "Perceptual linear prediction (PLP) analysis for speech," *J. Acoust., Soc. America*, vol. 87, pp. 1738-1752, 1990.

[2] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley-Interscience, New York, 1973.

[3] S. M. Katz, "Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 35, 1987.