한국어 음성합성기용 끊어읽기 추정기

Pause Predictor for Korean Text-to-Speech conversion

이 정 철*, 김 상 훈*, 성 굉 모**
(Jung Chul Lee*, Sang Hoon Kim*, Koeng Mo Sung**)

※본 연구는 정보통신부 출연 "HCI를 위한 음성 입출력 처리기술 개발" 과제로 수행되었습니다.

요 약

문장내 휴지구간의 위치와 길이는 합성음의 자연성을 결정짓는 주요 운율 파라미터 중 하나이다. 본 연구에서는 한국어 음성합성기의 합성음 생성에서 자연성 개선을 위해서 문장내 끊어읽기 위치 및 길이를 추정하기 위한 방법을 제안한다. 먼저 실제 발화에서 끊어읽기가 발생하는 요인을 검토하였다. 그리고 이들 요인에 부합하여 텍스트에 4단계의 끊어읽기를 표기함으로써 다량의 데이터를 확보하고 이를 이용한 NN 학습 결과와 HMM 추정기의 성능을 비교 검토한다. 현재까지의 결과로는 NN 학습의 경우 끊어읽기 없는 경우와 긴 끊어읽기의 추정에서는 우수한 예측능력을 보이지만 짧은 끊어읽기, 중간 끊어읽기의 경우는 HMM의 성능이 우수한 것으로 판명되었다. 전반적인 성능에서는 HMM이 우수하며 끊어읽기 종류에 따라 추정오차가 10~25%로서 안정적인 결과를 얻었으며 TTS에의 활용 가능성을 보였다.

ABSTRACT

Pause position and duration in a sentence is one of major factors which affect the naturalness of synthetic speech. In this paper, we propose a method to predict the pause position and duration for the enhancement in the naturalness of Korean synthetic speech. We investigated the factors of pause occurrence and gathered large text data tagged with 4-type pause symbols for the training. With these material, we tested pause prediction using a neural network and HMM method. The simulation results show HMM method predicts pause better than NN method even though NN method predicts well no break and long break. The simulation results show that prediction error is 10~25% according to the pause type. Our algorithm is very effective for the prediction of pause in a sentence, and generates a fairly natural synthetic speech.

I. 서 론

합성기를 이용하여 무제한의 텍스트를 사용자에게 낭독해주는 고품질의 음성합성 서비스를 제공하기 위해서는 합성음은 명료도와 자연성이 높아야 한다. 텍스트에서 띄어쓰기, 구둣점 등을 이용하여 독자에게 의미전달을 정확히 하고자 하는 것과 마찬가지로 발화에서는 운율을 이용하는데 운율의 표현수단은 음소 지속시간의 장단, 소리의 고저 (억양), 소리의 세기 (에너지 컨투어), 끊어읽기의 4가지 형태이다. 이중 끊어읽기와 관련된 연구는 언어처리 기술의 발전에 힘입어 많은 연구가 진행되어 발화에 대한 분석 및 합성에 응용되고 있다 [1-5]. 이들 대부분의 연구는 입력문장의 구문구조정보를 추정하고 규칙을 바탕으로 끊어읽기를 추정한다. Emerard의 경우 3단계의 끊어읽기 결정 방법을 사용하고 있다 [5]. 1단계에서는 입력문장을 breath 그룹으로 분리하며 그 경계에서 끊어읽기를 한다. 2단계에서는 한번의 날숨 (breath group) 내 단어들의 문법적 관계를 이용한 규칙을 사용하여 다시 하층구조에서의 끊어읽기를 결정한다. 마지막으로 끊어읽기 경계내 음절 수가 20음절 이상인 경우 적절한 위치에 끊어읽기를 정한다. 한국어를 대상으로 한 끊어읽기 추정방법에 대한 연구도 이와 같은 축으로 진행되어 사용되고 있다 [6-7]. 그러나 이상의 방법은 정교한 구문구조 분석기를 필요로 하고 있으며, 언어학적 지식만을 기반으로 한 것으로서 언어학적 구문구조가 발화에서의 운율구조와 일치하지 않는 단점이 있다. 이에 대한 해결방법으로 품사 추정기와 HMM방법을 이용한 확률적 추정방법이 제안되었으며 끊어읽기 유·무 2가지 유형에 대한 추정결과 높은 성공률을 보였다 [8-9]. 그러나 실제 발화를 대상으로 끊어읽기 현상에 대한 데이터를 수집하는 것이 쉽지 않은 단점이 있다.

본 연구에서는 이에 대한 대안으로서 먼저 실제 발화에서 끊어읽기가 발생하는 요인을 검토하였다. 그리고 이들

* 한국전자통신연구원
** 서울대학교 전자공학과

접수일자 : 1998년 4월 10일

요인에 부합하여 안정되게 끊어읽기를 하는 사람이 텍스트에 끊어읽기를 표기하여 다량의 데이터를 확보하였다. 이를 위해서 우리는 다음을 가정하였다. 발성자가 자신의 의사를 표현하고자 할 때 먼저 의미구조를 형성하고, 이를 전달하기에 적합한 단어들을 선택하며, 단어들의 연결이 자연스러운 구문구조를 작성한 뒤, 조음기관을 이용하여 음절단위로 발성하며 조음기관의 물리적 제약을 받게 된다. 물론 이 과정이 반드시 순차적으로 일어나는 것은 아니다. 언어학습을 통해 숙지된 의사표현과 발성자의 표현 습관 등은 반사적으로 흔히 나타나며, 특히 대화체와 같은 경우는 의미구조를 계속 생성하며 발화하는 것이 일반적이다. 그러나 이 경우라 할지라도 그 발화가 자연스럽다면 이미 상기 과정이 충분히 사전 학습된 결과라 볼 수 있다. 그러므로 끊어읽기는 문장의 의미구조, 구문구조, 단어, 조음결합 현상, 화자의 의도, 발화속도 등이 복합적으로 작용한 결과라고 할 수 있다. 이를 세부적으로 살펴보면 아래와 같다.

- 일반적으로 사람이 문장을 발성할 때 정확한 의미전달을 위해서 의미전달에 핵심이 되는 중심어를 강하게 분리 발성하여 강조를 하며, 이 외는 화자의 의도가 가미되어 강조되지 않는 일반적인 경우는 연결하여 발성한다.
- 전체 문장의 발성에서는 발성기관의 물리적 제약에 의해서 한번의 날숨으로 자연스럽게 발화할 수 있는 단위로 나누어 발성한다.
- 숨쉬기는 문장, 절, 분리도가 높은 구의 경계에서 일어난다.
- 자연스러운 숨쉬기는 그 의미적 경계가 뚜렷한 경우를 기준하여 일어난다.
- 의미적 경계의 깊이가 얕은 경우라도 한번의 날숨으로 발성하기에는 긴 발화는 짧은 숨쉬기가 breath group내의 가장 깊은 경계에서 일어나며 이때는 연속성이 유지되는 방향으로 표현된다.
- 음운환경과 사회적 관습 (표준어) 의해 기본이 결정되지만 구내 단어간 결합관계에 의해 변화된다.

본 연구에서는 의미구조, 화자의 의도, 발화속도와 같은 언어 외적인 요인은 배제하고 구문구조, 단어간 품사 결합 현상 등의 요인을 문장의 끊어읽기에 영향을 미치는 것으로 국한한다. 이 경우 문장내 어절의 품사열과 확률적 끊어읽기 분포도를 입력으로 한 NN학습 혹은 HMM을 이용하여 끊어읽기 위치 및 깊이의 추정이 가능함을 알 수 있다[10-13]. 그러므로 여기서는 이들 요인을 기준으로 하여 문장의 끊어읽기 추정기를 작성하고, 8,693 문장의 텍스트 데이타에 수작업으로 기록한 끊어읽기 정보를 대상으로 NN학습과 HMM추정결과를 비교 검토하여 음성합성기에 활용, 합성음의 자연성을 개선하고자 한다.

II. 학습에 사용된 데이터

본 연구에서 사용한 텍스트 데이터는 58개의 품사set을 이용하여 형태소 단위로 태깅된 8,693 문장으로 구성되어 있다. 전체 단어의 수는 114,199개이며 각 단어에 대한 형태소 분리 및 품사 할당은 실에서 보유중인 품사 추정기의 결과를 토대로 수작업으로 보완한 것이다. 그리고 이 텍스트 데이터에 끊어읽기 정보를 수작업으로 어절의 경계에 끊어읽기의 정도를 4단계로 나누어 1인이 태깅하였다. 태깅의 기준은 긴 숨쉼(long break, LB)이면 /// 끊어읽기 표시를, 중간 숨쉼(medium break, MB)이면 //, 짧은 숨쉼(short break, SB)이면 /, 숨쉼 없이 빨리 읽혀지는 부분(no break, NB)은 슬래쉬를 생략한다. 수작업으로 태깅한 결과 NB는 81,129개, SB는14,565개, MB는 7,437개, LB는 11,068개가 발견되었다. 그러나 실제 텍스트상에서는 끊어읽기를 하는데 정확한 기준이 없고, 사람의 주관적인 판단에 의해 행하여지기 때문에 일관성을 유지하는 못하는 문제점이 있다.

표 1. 조사, 어미의 품사별 끊어읽기 종류와 빈도

품 사	LB	MB	SB	NB
독립어	103	31	30	39
의존명사	115	347	530	2196
주격조사	110	653	1384	4144
목적격조사	31	282	1098	6815
관형격조사	2	13	189	4480
부사격조사	105	660	1603	5592
접속격조사	3	47	304	1070
공동격조사	0	5	20	177
호격조사	14	4	1	0
통용보조사	69	321	639	2016
특수보조사	647	1553	1602	2462
대등적 연결어미	354	829	982	853
종속적 연결어미	736	2174	2413	3106
보조적 연결어미	0	4	62	3748
인용표시 어미	6	52	136	387
관형사형 전성어미	17	168	1142	10973
종결어미	8529	34	74	30
명사파생 접미사	13	10	38	849

또 문형, 문장의 길이, 의미구조, 구문구조, 화자의 의도, 발화속도에 따라 끊어읽기가 달라지므로 실제 발화에서 나타나는 끊어읽기와 수작업 결과가 일치하지 않는 단점이 있지만 이들 문제를 통계적 분석방법으로 해소하고자 한다. 표 1에 대표적 품사별 끊어읽기 종류에 따른 빈도를 나타내었다. 빈도를 고려하면 호격조사, 독립어, 종결형 어미 뒤에는 LB가 될 확률이 높으며, 기타의 경우는 NB가 될 확률이 높은 것을 알 수 있다. 그러나 비록 NB의 확률이 높다고 해도 SB, MB, LB가 상대적으로 발생빈도가 낮을 뿐이지 그 중요도가 낮다고 볼 수는 없다. 즉 문맥을 고려하지 않고 어절의 단일 품사로 끊어읽기 추정을 하면 정확도가 낮다는 것을 알 수 있다. 특히 독립어, 비단위성 의존명사, 주격조사, 목적격 조사, 특수보조사, 통용보조사, 명사파생 접미사, 대등적 연결어미,

종속적 연결어미 등은 문맥의존도가 높음을 보이고 있다.

표 2에 대표적인 품사 bigram에 따른 끊어읽기의 유형과 빈도수의 일예를 나타내었다. 표에서 알 수 있듯이 단일 어절의 품사정보로 끊어읽기 유형을 예측하는 것보다는 bigram을 사용하는 것이 더 좋은 결과를 얻을 수 있을 것이라는 것을 짐작할 수 있다. 그러나 여전히 판정의 모호성이 해결되지 않음을 알 수 있다.

표 2. 품사 bigram에 따른 끊어읽기 type별 빈도수

선행 품사	후행 품사	LB	MB	SB	NB
대등적 연결어미	관형사적 연결어미	31	71	102	187
대등적 연결어미	보통명사	64	146	144	0
종속적 연결어미	보통명사	99	325	360	0
종속적 연결어미	관형사형 전성어미	0	176	0	635
주격조사	일반부사	19	111	186	0
부사격조사	관형사형 전성어미	0	69	182	1514

이상의 관찰 결과 끊어읽기 유형을 결정짓는데 있어서 품사정보에 따른 규칙성이 보이지만 표층에 보이는 단일 혹은 bigram를 이용하는 것은 부적절함을 알 수 있다. 이에 다수 어절에 대한 품사열 정보와 끊어읽기 유형을 지식기반으로 한 NN학습과 발생 및 천이 확률 모델을 이용한 HMM 방식의 끊어읽기 유형을 추정을 시도하였다.

III. HMM 방식의 끊어읽기 추정

문장의 끊어읽기는 아래의 과정으로 추정한다.

- 문장내 각 어절별 품사를 추정한다. 여기서 사용하는 한국어 품사 추정기는 복수후보 형태소 분석결과와 통계적 확률을 기반으로 한 HMM방식을 이용하여 문장 단위의 최적 품사열을 추정하는 것으로 정확도는 90% 이상이다. 품사 추정기의 출력은 어절내 형태소별 분리 결과와 형태소별 품사 추정값이다. 문장의 끊어읽기에서는 문장내 어절의 문법적 기능을 중심으로 구현한다.
- 어절과 어절사이에 끊어읽기가 존재한다고 가정한다. 끊어읽기 유형은 단순하게는 NB (no break), B (break)의 2가지로 둘 수 있으며, NB, SB, MB, LB의 4가지 혹은 그 이상으로 둘 수 있다. 테이블에 저장된 각 끊어읽기 유형에 대한 품사열의 발생확률을 이용하여 어절별, 상태별 확률값을 할당한다. 여기서 사용할 품사열은 대상 어절과 대상 어절의 직전, 직후 어절 1개씩으로 구성되며, 구둣점도 하나의 품사로 간주한다. 확률값은 식 1과 같다.

$$P(c_{k-1}, c_k, c_{k+1} \mid j_k) \tag{1}$$

여기서 C_k는 k번째 어절의 품사 정보이고, j_k는 k번째 어절을 뒤따르는 끊어읽기 유형을 나타낸다.

- 끊어읽기 유형의 천이 확률값으로 n-gram을 사용한다. 천이확률은 식 2와 같다.

$$P(j_k) = P(j_k \mid j_{k-1}, j_{k-2}, \cdots, j_{k-N+1}) \tag{2}$$

- 끊어읽기 유형별 발생확률은 식 3의 Bayes' rule을 이용하여 구한다.

$$P(j \mid C) = \frac{P(C \mid j)P(j)}{P(C)} \tag{3}$$

- 문장 전체에 대한 끊어읽기 유형의 선정은 Viterbi 탐색 방법을 이용하여 최적 끊어읽기 type열을 선택한다.

3.1 어절 끝 형태소의 품사로 구성된 품사열 Trigram을 이용한 추정기

어절의 형태소 분석결과중 어절의 문법적 기능의 대표값인 마지막 형태소의 품사값을 사용한다. Target 어절의 뒤에 위치하는 끊어읽기 유형을 추정하는데 직전, 직후 어절의 품사를 이용하여 연속된 3개의 품사열을 이용한다. 아래에 예를 보이고 있다.

(예 1.1)

- 입력문장 : 꼭 필요한 회의가 시의적절하게 개최되는 것으로 평가된다.
- 끊어읽기 : 꼭 필요한 회의가// 시의적절하게 개최되는 것으로/ 평가된다.
- 품사추정 : mag etm jcs ecs etm jca ef punc
- 학습형태 : mag etm jcs[MB] ecs etm jca[SB] ef punc [LB]
- 상태 발생확률 : "mag etm jcs" 품사열의 경우 etm 뒤에 NB, SB, MB 혹은 LB의 끊어읽기가 된다면 이때 "mag etm jcs"의 품사열이 발생할 확률.
- 상태 천이확률 : 선행되는 5개의 juncture 열에서 target state가 NB, SB, MB 혹은 LB의 끊어읽기로 천이할 확률

한국어 품사셋과 4종류의 끊어읽기 유형을 이용하여 끊어읽기 유형별 tri-gram품사열을 조사한 결과 18,048 종류가 발견되었고, hexa-gram 천이확률은1,800 종류가 되었다. 이를 이용한 끊어읽기 추정 결과에 따른 confusion matrix를 표 3에 나타내었다.

표 3. HMM-1 끊어읽기 추정결과 confusion matrix

실제값 \ 추정값	NB	SB	MB	LB
NB	82%	11%	6%	1%
SB	22%	54%	23%	1%
MB	10%	23%	64%	3%
LB	2%	4%	16%	78%

아래에 입력문장과 추정된 결과의 예를 2개 나타내었다.

(예 2.1)

- 입력문장 : 이같이 중국의 오염물질은 동북아 지역의 심각한 환경파괴 요인으로 등장하고 있다.
- 끊어읽기 : 이같이// 중국의 오염물질은// 동북아 지역의 심각한 환경파괴 요인으로/ 등장하고 있다.
- 추정결과 : jca[SB] jcm jxt[MB] nq jcm etm ncn jca

[SB] ecx ef punc[LB]

→ 아같이/ 중국의 오염물질은// 동북아 지역의 심각한 환경파괴 요인으로/ 등장하고 있다.

(예 2.2)

- 입력문장 : 일본에도 심각한 영향을 미칠 중국의 대기 오염을 개선하는데 일본의 협력은 필수적인 것으로 보인다.
- 끊어읽기 : 일본에도 심각한 영향을 미칠/ 중국의 대기 오염을 개선하는데// 일본의 협력은/ 필수적인 것으로 보인다.
- 추정결과 : jxc[B] etm jco etm[SB] jcm jco[SB] etm nbn[MB] jcm jxt[MB] etm jca[SB] ef punc [LB]

→ 일본에도/ 심각한 영향을 미칠/ 중국의 대기오염을/ 개선하는데// 일본의 협력은// 필수적인 것으로 보인다.

예 2.1은 끊어읽기 추정결과가 수작업 결과와 일치함을 보여주는 것으로 품사열을 이용한 끊어읽기 성공예를 보인다. 예 2.2는 추정의 문제점을 보여주는 것으로 빈번한 끊어읽기와 부적절한 끊어읽기를 보여주고 있다.

확률기반 HMM 추정방법이 자연성, 일관성을 향상시킨다고 볼 수 있다. 그러나 각 어절의 의미구조, 화자의 의도, 발화속도와 같은 언어 외적인 요인이 배제되는 제한점이 있고 빈번한 끊어읽기와 부적절한 끊어읽기의 길이 현상이 생긴다. 그리고 어절에 대한 품사의 할당이 어절의 끝에 위치하는 형태소의 품사로 이루어지므로 앞, 뒤 어절간의 문법적 관련 정보를 놓칠 가능성이 높은 단점이 있다.

3.2 어절 앞, 끝 형태소의 품사로 구성된 품사열 hexagram을 이용한 추정기

어절 끝 형태소의 품사로 구성된 품사열 Trigram을 이용한 추정기의 경우 어절에 대한 품사의 할당이 어절의 끝에 위치하는 형태소의 품사로 이루어지므로 앞, 뒤 어절간의 문법적 관련 정보를 놓칠 가능성이 높은 단점이 있다. 예를 아래에 보인다.

(예 3.1)

- 입력문장 : 나는// 그가 일을 함에 있어서// 아주 만족함을/ 느낀다.
- 추정결과 : jxt jcs [MB] jco [MB] jca ecs [SB] mag jco ef punc [LB]

→ 나는// 그가 일을/ 함에 있어서// 아주 만족함을/ 느낀다.

예 3.1에서 /일을 함에/의 경우 목적격조사와 서술어 /하다/는 긴밀하게 묶일 확률이 높은데 어절의 끝 형태소 품사만 사용할 경우 jco+jca의 결합관계만 살피게 되므로 오류를 피하기 어렵다. 이에 대한 보완 방법으로 어절 앞,끝 형태소의 품사로 구성된 품사열 hexagram을 이용한 끊어읽기 유형별 발생확률을 이용한 추정기를 생각할 수 있다. 한국어 품사셋과 4종류의 끊어읽기 유형을 이용하여 끊어읽기 유형별 hexagram 품사열을 조사한 결과 39,347 종류가 발견되었다. hexa-gram 천이확률은 동일한 것을 사용한다.

이를 이용한 끊어읽기 추정 결과에 따른 confusion matrix를 표 4에 나타내었다. 3-2방법이 어절의 끝형태소의 품사열을 이용하는 3-1방법에 비해 추정성능이 7%~22% 향상되었다. 즉 어절간 결합관계를 활용할 수 있는 구조를 사용함으로써 추정의 정확도가 향상됨을 알 수 있다.

표 4. HMM-II 끊어읽기 추정결과 confusion matrix

실제값 \ 추정값	NB	SB	MB	LB
NB	89%	8%	3%	0%
SB	8%	75%	16%	1%
MB	2%	11%	86%	1%
LB	0%	2%	7%	90%

IV. NN 학습 방식의 끊어읽기 추정

HMM 방식을 이용한 끊어읽기의 추정은 상태 발생확률과 천이확률을 이용하여 문장내 끊어읽기의 다양한 후보들에 대해 확률적으로 가장 높은 값을 보이는 경우를 선택한다. 만일 발생 빈도에 있어서 확률적으로 분해도가 낮은 (비슷한 확률분포) 경우에는 기존의 연속 3어절 품사열을 다수 어절로 확장하여 문맥의 상관도를 살필 수 있는 범위를 넓힘으로써 분해도를 높일 수 있다. 그러나 이 경우 학습 데이터가 제한되어 있어서 발생빈도가 한쪽으로 치우치는 small sample 문제에 민감한 경향을 보인다. 또한 탐색 공간이 팽창하여 수행에 필요한 계산적 부하가 증폭되는 단점이 있다.

Unknown case, 즉 학습에 사용되는 데이터에 없는 품사열이나 천이가 발생할 경우, 높은 대가를 지불하는 조건으로 후보에 등록시킴으로써 다소 배타적 입장을 보인다. 그러나 학습 데이타에 없는 품사열의 다양한 조합이 가능하며 실제 문장에서 이러한 조합이 발생한다. 이 경우 주어진 조건 이외에는 유추의 가능성이 부족하다.

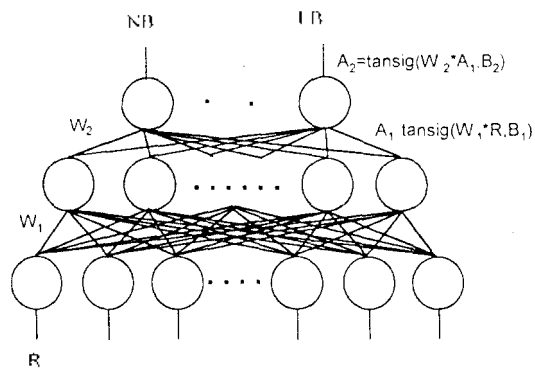

그림 1. NN학습기의 구조

이상의 관점에서 이들 문제점을 보완하기 위하여 신경망 학습방법을 고려해 볼 수 있다. Netalk와 같은 발음변환기, F0 contour 추정기 등에서 NN을 이용하여 성공한 예를 참조하여 품사열을 입력으로 하고 끊어읽기 유형을 출력으로 하는 끊어읽기 추정기의 가능성이 있다고 예측할 수 있다. 특히 앞의 HMM추정기의 결과에서 품사열로부터 끊어읽기 추정의 결과가 성공적이므로 규칙성이 있음을 짐작할 수 있고, 이 경우NN의 학습이 성공할 확률이 높기 때문이다. 그림 1에 본 연구에서 사용한 NN학습기의 구조를 나타내었다.

4.1 NN 학습방법 I

NN을 학습하는데 있어서 그 목적은 입력된 품사열 정보를 이용하여 target 어절 뒤에 올 juncture type을 예측하는데 있다. 이 목적을 위해서 방법 1에서는 연속되는 5어절의 품사열을 입력으로 하며 target 어절은 중간에 위치한다. 즉 target 어절을 중심으로 선행 2어절, 후행 2어절의 품사 정보를 입력으로 한다. 각 어절에 대한 품사정보의 입력은 58개의 입력 벡터중 해당되는 위치가 ON, 그 외는 OFF되는 sparse 방법을 사용한다. 그리고 부가적 입력조건으로 문장내 어절수, 문장내 위치, 어절별 음절수를 입력 데이터로 사용하였다. 출력은 NB, SB, MB, LB 4가지의 출력이 되며 학습시 target 어절 뒤 끊어읽기 유형에 해당되는 출력이 ON, 그 외에는 OFF 되도록 학습한다. 예측시 판정의 기준은 4개의 출력중 가장 높은 값을 보이는 것을 ON으로, 그 외는 OFF로 판정한다.

학습에 사용된 데이터를 테스트 데이터로 사용하여 추정성능 실험을 한 결과 5,000번의 반복학습에서 구한 confusion matrix는 표 5와 같다.

표 5. NN-I 끊어읽기 추정결과 confusion matrix

실제값 \ 추정값	NB	SB	MB	LB
NB	98%	1%	1%	0%
SB	71%	21%	8%	0%
MB	45%	23%	31%	1%
LB	7%	4%	9%	80%

표 5의 결과를 보면 NB의 경우 98%의 정확도를 보이고, LB의 경우는 HMM방식과 비슷하여 희망적이지만, SB와 MB의 경우는 NB로 오판하는 경우가 많으며 성공률은 20% 내외로서 학습이 실패함을 알 수 있다 특히 LB의 경우는 대다수의 경우가 문장의 끝으로 특정한 품사열의 패턴이 형성되므로 학습의 성능이 높아지는 경향이 있다.

4.2 NN 학습방법 II

방법 NN-I에서 사용한 입력 데이터의 차원은 298 (58*5+5+2+1) 이지만 품사조합의 경우 최대 58**5 개의 패턴을 생각할 수 있다. 그러나 실제 학습에 입력되는 데이터는 최대 80,000개, 최소 7,000개의 학습데이타가 있을 뿐으로 small sample 문제가 심각함을 알 수 있다. 따라서 입력 데이터의 통합 분류가 필요하다. 방법 2에서는 각 어절에 대한 품사정보의 입력을 품사별 대그룹 9종류 그리고 그룹내 품사에 대해서는 무작위로 4bit를 할당하여 13개의 입력벡터 중 해당되는 위치가 ON, 그 외는 OFF 되는 방법을 사용한다.

학습에 사용된 데이터를 테스트 데이터로 사용하여 추정성능 실험을 한 결과 5,000번의 반복학습에서 구한 confusion matrix는 표 6와 같다.

표 6. NN-II 끊어읽기 추정결과 confusion matrix

실제값 \ 추정값	NB	SB	MB	LB
NB	98%	1%	1%	0%
SB	67%	25%	7%	0%
MB	42%	28%	30%	0%
LB	6%	5%	9%	79%

NN 학습방법I의 결과에 비해 성능이 3~4% 정도 향상되지만 여전히 HMM에 비해서 성능에 큰 차이를 보이고 있다.

V. 결 론

본 논문에서는 끊어읽기 위치를 추정하기 위한 NN 학습 결과와 HMM 추정기의 성능을 비교 검토하였다. 현재까지의 결과로는 NN 학습의 경우 no break와 long break의 추정에서는 우수한 예측능력을 보이지만 short break, medium break의 경우는 HMM의 성능이 우수한 것으로 판명되었다. 전반적인 성능에서는 HMM이 우수하며 안정적인 결과를 얻었다. 그러나 NN에 대한 입력조건 변화나 다량의 학습 데이터를 사용할 경우 예측능력이 개선될 여지는 있다고 본다.

이상의 결과를 토대로 음성합성기에 끊어읽기 추정기를 이용하여 합성음을 생성한 결과 합성음의 자연성이 향상됨을 청취실험을 통하여 알 수 있었다.

앞으로 체계적인 청취실험을 통하여 향상도에 대한 객관적 수치와 개선방향에 대한 연구를 진행하고자 한다.

참 고 문 헌

1. D.H. Klatt, "Review of text-to-speech conversion for English," J. Acoust. Soc. Am. 82(3), pp.737-792, 1987. 9.
2. J. Allen, M.S. Hunnicutt, D. Klatt, From text to speech: The MITalk system, Cambridge University Press, 1987.
3. D.Hirst, "Structures and Categories in Prosodic Representation," in *Prosody: Models and Measurements*, Springer-Verlag, pp.93-109, 1983.
4. D.Hirst, "Prediction of prosody: An overview," in *Talking Machines: Theories, Models, and Designs*, North-Holland, pp. 199-204, 1992.

5. F. Emerard, L. Mortamet, A. Cozannet, "Prosodic processing in a text-to-speech synthesis system using a database and learning procedures," in *Talking Machines: Theories, Models, and Designs*, North-Holland, pp.225-254, 1992.
6. 김세린, 외 4인, "한국어 문장-음성 변환 시스템에서의 운율 처리," KSCSP'96 13권 1호, pp.415-418, 1996.
7. 김정수, 이혜정, "언어정보 및 통계 데이터를 이용한 한국어 운율 생성," KSCSP'96 13권 1호, pp.227-231, 1996.
8. A.W. Black and P. Taylor, The festival speech synthesis system: system documentation. Technical report H A.W. Black and P. Taylor, "Assigning phrase breaks from part-of-speech sequences," in Eurospeech'97 Proceedings, vol.2, pp.995-998, 1997.
9. CRC/TR-83, Human Communication Research Centre, University of Edinburgh, 1997.
10. J.C. Lee, and Koeng-mo Sung, "Improvement of synthesised speech intonation with stylisation and neural network learning," in Electronics Letters Vol. 33, No. 19, pp.1600-1601, 1997.
11. J.C.Lee, S.H.Kim, Minsoo Hahn, "Intonation Processing for TTS Using Stylization and Neural Network Learning Method," in *Proc. ICSLP'96*, pp. 1377-1380, 1996.
12. 이정철, 이영직, "음성합성시에서의 한국어 내화체 운율 구현," '96 음향학회 학술대회 논문집, 제 15권1호, pp.103-106, 1996.11.
13. H. Oh, Y.J. Lee, "A modified error function to improve the error back-propagation algorithm for multilayer perceptrons," ETRI J., vol.17, pp.11-22, 1995.

▲이 정 철(Jung-Chul Lee)
1984년 2월 : 서울대학교 전자공학과(학사)
1988년 2월 : 서울대학교 전자공학과(석사)
1990년 3월~현재 : 서울대학교 전자공학과 박사과정
1985년 9월~현재 : 한국전자통신연구원 선임연구원

▲김 상 훈(SangHoon Kim)
한국음향학회지 제16권 1호 참조

▲성 굉 모(Koeng-Mo Sung)
한국음향학회지 제16권 3호 참조