

음성인식·합성을 위한 한국어 운율단위 음운론의 계산적 연구: 음운단위에 따른 경계의 발견

이 찬 도[†]

요 약

성공적인 음성인식·합성 시스템을 구축하기 위해서는 음운론적 지식, 특히 운율정보의 도입이 매우 중요하다. 본 연구에서는 우선 음성인식·합성을 위한 운율음운론의 연구동향을 개관하고, 국어의 음운단위와 경계의 설정에 관한 이론적·실험적 고찰을 정리하였으며, 음운단위에 따른 경계의 자동적 발견을 위하여, 데이터를 수집하고 시스템을 구현하여 실험을 행하였다. 단순회귀 신경망을 이용하여, 2,200여 개의 문장에 있는 12,000여 개의 음운단어를 외부정보의 도움이 전혀 없이 훈련시킨 결과, 70%정도의 예측률을 보였다. 본 연구에서 사용한 방법을 다른 정보와 결합하여 사용한다면, 음운경계의 발견과 그에 따른 분절화를 정확하게 행할 수 있으리라 기대된다.

A Computational Study of Prosodic Structures of Korean for Speech Recognition and Synthesis: Predicting Phonological Boundaries

Chan-Do Lee[†]

ABSTRACT

The introduction of phonological knowledge, especially prosodic information to speech recognition and synthesis systems is very important to build successful spoken language systems. First, related works of computational phonology is overviewed and the theoretical and experimental studies of prosodic structures and boundaries in Korean are summarized. The main focus of this study is to decide which prosodic phrasing constituents are in Korean and to find the word boundaries automatically. A total of 2,200 sentences were trained on a simple recurrent network. The results show that the network was able to predict 70% of the boundaries correctly, even though it was not provided with any information other than phonetic features. This method can be combined with other useful information to predict the boundaries more correctly and to help segmentation, which are vital for the successful speech recognition and synthesis systems.

1. 머리말

인간이 가장 자연스럽게 의사소통에 사용하는 도구는 음성언어(spoken language)이며, 음성언어의 사용은 인간을 다른 동물과 구별해 주는 가장 큰 특징이다. 음성언어를 듣고 말할 수 있는 컴퓨터를 만들고자 하는 노력은 오랜 연구 초점이 되어 왔으나 많은 노력에도 불구하고 유아라도 일정 수준 발휘할 수

※ 이 논문은 1994년도 한국학술진흥재단의 공모과제 연구비와 1995년도 대전대학교 특별연구사업비에 의하여 연구되었음.

† 중신회원: 대전대학교 정보통신공학과
논문접수: 1996년 4월 1일, 심사완료: 1996년 10월 14일

있는 언어능력을 컴퓨터는 아직 충분히 갖추지 못하고 있다. 그 이유는 여러 가지를 들 수 있겠지만 그 중에서도 언어학, 특히 음운론적 이해가 부족하여 음성 언어의 근본적 구조 및 복잡다기한 현상은 도외시한 채 단순한 공학적인 접근방법만을 사용한다면 정확한 음성인식 시스템과 자연스러운 음성합성 시스템을 구현한다는 것이 불가능하기 때문이다.

현재까지의 음성인식 및 음성합성을 위한 한국어 음성언어의 연구는 고립어 단어가 위주였으며, 연속음성에 대한 연구는 미흡하다. 그러나 실제로 발화되는 연속음성을 자연스럽게 처리하기 위해서는 국어의 운율구조에 관한 연구가 필수적이다. 언어학자들 간에 공통된 의견은 단어란 분석할 수 있는 단위로서, 이것은 좀 더 작은 단위들로 이루어지며, 추상적인 단위로부터 구체적인 단위에 이르기까지 계층구조가 존재한다는 점이다. 발화의 흐름이 음운단위(phonological unit)내에서 조직됨을 강조하는 최신의 이론인 운율음운론(prosodic phonology)에 입각하여 음운단위를 설정해 보면 음절, 음운단어, 음운구, 억양구 및 발화구로 나눌 수 있다. 여기에서 각 음운단위를 설정하는 이유는 이들 음운단위들이 언어학적으로 변별력이 있는 단위로서, 발화란 의미정보별로 나뉘어 발음되며 그럴 때에 음운단위 뒤에 오는 경계로 인하여 많은 음운변동 현상이 초래되기 때문이다. 발화의 흐름은 음운단위 내에서 조직되며, 이들 음운단위들은 악센트, 리듬, 억양(intonation) 등과 관련하여 단어나 구 및 문장을 체계적으로 파싱하는데 직접적으로 사용되며, 또한 문법 및 의미의 차이를 전달하는 주요한 역할을 한다. 또한 이들과 쉼(pause), 운율변동규칙간의 관계는 연속음성을 인식하고 합성하는데 가장 중요한 단서를 제공한다. 결국 분절음과 더불어 정확한 운율정보가 첨가되어야 음성인식과 합성에서 가장 자연스럽게 합리적인 알고리즘을 만들 수 있다. 예를 들면, “아버지가 방에 들어가신다”와 “아버지 가방에 들어가신다” 이 두 문장은 악센트, 리듬, 억양 등에서 현저한 차이를 나타내므로, 이러한 운율적 정보를 이용하면 단어, 문장의 오인식을 줄일 수 있다.

이러한 중요성에도 불구하고, 운율구조를 자동적으로 음성인식·합성 시스템에서 구현해 줄 수 있는 운율처리 시스템의 연구는 그리 많지 않다. 따라서

성공적인 음성인식 시스템과 음성합성 시스템의 구축을 위해서는 운율처리 시스템의 개발이 반드시 선행되어야 한다고 할 수 있다.

본 논문의 목적은 두 가지이다. 첫째, 음성인식 및 음성합성과 운율음운론 연구의 관계를 조망하므로써 이 분야의 연구에 새로운 방향을 제시하고, 둘째, 운율음운론 연구중 세부분야인 음운단위에 따른 경계의 발견을 위한 시스템을 구현하여 그 결과를 보고하는 것이다. 본 논문의 구성은 다음과 같다. 먼저 음성인식·합성을 위한 운율음운론의 연구동향을 개관하고, 음운단위와 경계간의 관계를 살펴본다. 또한 음운단위에 따른 경계를 자동적으로 찾아낼 수 있는 시스템을 설계, 구현한 후 실험을 통하여 그 결과를 분석하고, 본 연구의 기여도에 관해 논의한다.

2. 음성인식·합성과 운율음운론

컴퓨터를 이용한 음운론 연구(전산음운론), 그중 특히 운율음운론에 관한 연구는 아직도 그리 많지 않다. 특히 국내에서는 이에 대한 연구가 거의 없는 실정이다. 접근방법은 전통적 기호주의의 방법론을 쓰는 경우와 새로운 페러다임인 인공지능망을 이용하는 경우가 있는데, 종래에는 기호주의 처리방법이 주된 흐름을 형성했으나, 최근 들어서는 인공지능망을 통한 연구방법이 활기를 띠고 있다.

Cole et al[1]은 음성언어 시스템의 연구과제를 8분야로 나누어, 각 분야에 세부주제를 제시하고 있는데, 운율음운론에 관계되는 분야로는 (1) 동시조음현상, (2) 발화속도와 (3) 운율을 들고 있다. 그들은 이 분야가 음성인식과 합성에 많은 도움을 줄 수 있는 것은 확실하나, 아직은 연구가 초보단계에 머물러 있음을 지적하고 있다.

음성언어의 이해를 위한 연구로는 음운론적 파싱[2], 운율정보를 이용한 모호성 해소[3], 강세를 이용한 발화내용의 핵심 발견[4], 억양, 강세, 음운구조의 발견[5, 6, 7] 등 운율정보를 어떻게 자동적으로 찾아내어 음성인식·합성에 통합할 수 있는가에 주안점을 두고 있다.

또한 대용량 데이터베이스를 이용하여 음운구조에 따른 분절화는 어떻게 자동적으로 할 수 있는가(automatic segmentation)에 관한 연구도 활발히 진행되고

이다[8, 9, 10, 11].

국내에서의 계산음운론에 입각한 운율음운론 연구는 거의 전무하며, 그의 필요성조차도 별로 인식하지 못하고 있다고 볼 수 있다. 음성인식과 합성에 관한 연구는 1980년대 중반이후 대학교, 연구소등을 중심으로 활발히 진행되어 오고 있으나, 음성인식·합성과 관련한 음운론 연구는 근래에 들어 비로소 그 필요성이 언어학자와 음성인식·합성 연구에 종사하는 공학자들간에 인식되어 오고 있다[12, 13, 14, 15]. 그러나 아직도 어떻게 음운론의 연구가 공학적 접근방법과 조화를 이룰 수 있는지에 관한 연구는 미비한 실정이다.

3. 음운단위와 경계

운율음운론에서는 문장을 이루고 있는 분절음들의 연쇄체가 여러 음운단위간의 계층적 구조로 되어 있다고 간주한다. 운율음운론에서 운율구조는 음운론적 요소와 함께 통사론적, 의미론적 정보가 모두 포함되어 있는 구조이다. 이러한 음운단위는 경계를 수반하며, 이들은 음운규칙과 관련하여 중요한 역할을 한다.

국어의 음운단위와 경계의 설정에 대한 이론적·실험적 고찰은 본 연구자가 속해 있는 연구팀에서 수행한 연구보고서[15]에 잘 정리되어 있는데, 이를 본 연구에서 행한 시스템 구현과 직결되는 부분만 발췌 요약하여 보면 다음과 같다.

국어의 운율구조를 구성하는 주요한 음운단위로는 음절(syllable), 음운단어(phonological word), 음운구(phonological phrase), 억양구(intonational phrase)와 발화구(utterance)가 있다.

음운단어는 단어와 그 단어에 결합되는 조사와 어미까지 포함한다. 복합어는 그 내부에 아무런 경계가 나타나지 않을 경우에 전체를 하나의 음운단어로 간주한다. 또한 불완전 명사들은 앞에 나오는 수식어와 함께 하나의 음운단어를 형성한다.

음운구는 음운단어들의 연결체이며, 그 통사적 정보는 주부, 술부, 수식어와 피수식어, 문장부사, 부사절, 등위구문 등과 같은 문법적 최대투영이다. 초점(focus)이 주어지는 경우, 초점을 받는 단어는 음운구를 형성한다. 또한 복합어의 경우, 통사적, 의미적 정보를 반영하는 경계가 나타날 때 하나의 음운구로 간

주할 수 있다.

억양구는 음운구들의 연속체로서, 원형문장과 그것과 외적 관계를 맺고 있는 모든 요소들이고, 주제어, 호칭어, 감탄사 등 구조적으로 문장에 바로 붙지 못하는 음운구들은 그 자체가 억양구가 된다. 또한 삽입절, 부가의문문, 비제한적 관계절도 별개의 억양구를 형성한다.

발화구는 억양구의 연속체로서 문장 전체를 한 단위로 한다.

경계는 연속적인 발화의 흐름을 끊어서 일정한 의미정보별로 나누어주는 기능을 하는 운율 요소로서, 경계를 결정해주는 운율요소로는 쉼, 음조변화(pitch change), 경계선 길이연장(preboundary lengthening)이 포함된다. 이 세 가지 요소들은 복합적으로 상호작용 하여 음성의 차이를 가져온다. 경계는 주로 쉼에 의해 쉽게 인지되지만, 음조변화와 그에 따른 길이증가를 수반하기도 하며, 청각적으로 인지할 수 없는 정도의 쉼이 있는 경계에서는 음조와 길이가 경계를 구분하는 중요한 요소로 작용한다. 길이증가는 음조가 올라갈 경우 나타나기도 하는 특징으로, 이 둘 두 요소는 특히 서로 연관되어 비례적 작용을 보인다. 경계는 이러한 음성특질로 실현되어 나타나는 현상이며, 음운경계는 형태·통사론을 반영한 음운단어에 쉼, 음조, 길이와 같은 운율요소가 작용하는 현상이다.

각 음운단위에 나타나는 경계패턴과 상기한 운율요소간의 관계를 구명하기 위하여 유형별로 녹음된 자료를 음성학적으로 분석하여 그 자료의 경계에 나타나는 음성적 특성을 통해 경계의 패턴을 정하여 보면, 다음과 같다. 음운단어 경계에서는 음조변화가 수반되지만, 장음화와 쉼은 미약하다. 음운구 경계에는 장음화가 중요요인으로 작용하며 쉼과 음조변화는 선택적이다. 억양구 경계에는 음운단어나 음운구 경계와 뚜렷이 구별되는 쉼이 의무적으로 나타나며, 그 외는 선택적으로 올 수 있다. 발화구 경계는 위 세가지 운율요소가 다 나타나는 하지만 통사적으로 매우 분명한 문장단위이므로, 별도의 음성학적 정리는 필요없는 것으로 보인다.

음운경계를 찾을 수 있는 또 다른 방법은 한국어에 나타나는 음소결합 제약규칙(phonotactics)을 이용하는 것이다. 즉, 각 음소들이 다른 음소들과 어떻게 결

합하는가를 조사하여, 빈도수가 적거나 없는 음소결합은 일단 경계의 후보로 정하는 계량 언어학적 방법이다. 예를 들면, 장애음 뒤에 오는 장애음들은 모두 된소리가 되는 한국어의 특수성으로 말미암아 장애음 종성 /ㅂ/, /ㄷ/, /ㄱ/ 뒤에는 여린 소리 /ㅂ/, /ㄷ/, /ㄱ/, /ㅅ/가 결합되지 못하므로¹⁾, “한국사람[한국사람]”은 “한국”과 “사람”의 두 단어로 이루어져 있음을 발견해낼 수 있다. 이러한 방법을 컴퓨터에 도입하여 음운경계를 찾기 위해서는 한국어 전부에 대한 음소결합법칙을 규칙화한 생성시스템(production system)을 이용하거나, 대용량 데이터베이스를 사용하여 스스로 학습하도록 하는 방법을 택할 수 있다.

4. 음운단위에 따른 경계의 발견

제2장에서 개관한 바와 같이 음성인식·합성을 위한 운율음운론 연구의 주요점은 (1)음운단위 경계에 따른 분절화는 어떻게 자동적으로 할 수 있는가 하는 것과 (2) 강세나 발화속도 등과 같은 운율정보를 어떻게 자동적으로 찾아내어 시스템에 통합하는가 하는 데에 두고 있다. 그런데 경계를 먼저 찾아낸 후에야 그에 따른 음운규칙의 적용여부를 효율적으로 처리할 수 있으므로, 본 연구에서는 음운단위, 그 중 음운단위에 따른 경계를 자동적으로 찾아낼 수 있는 시스템을 구현하고자 한다. Kaye[16]도 음운론의 가장 중요한 기능 중 하나는 파싱을 하는 과정에서 경계를 찾아내어 분절화 하는 것이라고 주장하였고, Gaskell[11]도 단어인식을 위해서는 분절화가 선행되어야 함을 지적하고 있다.

자동 분절화에 관한 연구는 제3장에서도 제시한 바와 같이 실제 발화된 음성신호의 쉼, F₀, 길이등의 정보를 이용하여 음운구조를 찾아내는 방법[5, 6, 7]과 음소자질을 이용하여 이진수로 나타낸 음소열을 입력자료로 하여 스스로 음소결합 제약규칙을 인지하고, 이 정보를 이용하여 경계를 발견하는 방법[8, 9, 10, 11] 등이 사용되고 있는데, 본 연구에서는 후자를 택하였다.

경계를 찾아내는데 효과적인 방법으로는 다음과 같은 알고리즘을 적용하였다. 대용량의 음성을 음소

열로 바꾸어 연속적으로 입력하면서 다음에 올 음소를 예측하도록 한다. 경계에서는 아무래도 예측율이 저조할 것이므로, 이 정보를 이용하여 분절화를 행한다. 이 방법은 경계에 관한 정보는 전혀 가지고 있지 않은 시스템이 시간축을 따라 변화하는 입력자료에 대한 일반성(generalization)을 표현하는 과정에서 통계를 통해 스스로 음운단어별로 음소결합 제약규칙을 찾아냄으로서 경계를 예측하도록 하는 비통제 학습(unsupervised learning)의 일종이다.

4.1. 데이터

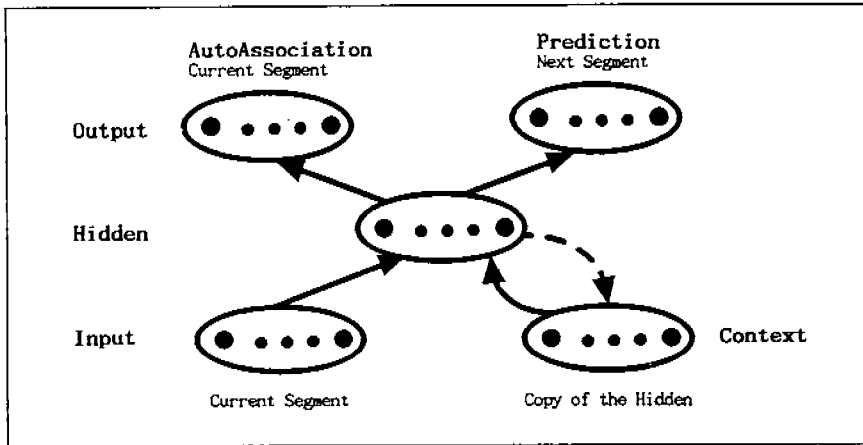
음운단어의 경계를 자동적으로 찾아내기 위해서는 대용량 데이터가 필요하다. 데이터로는 책에 쓰여진 텍스트(written text)는 실제 발화된 음성(spoken language)에 비해 아무래도 현장감이 결여되고 실제 대화상과는 많이 상이할 수가 있기 때문에 실제 대화를 녹음한 자료를 사용하였다[17]. 이 자료는 2,200여 개의 문장에, 12,000여 개의 음운단어를 포함하고 있다.

그러나 본 연구를 위해서는 실제 음성자료 그 자체보다는 전사(transcription)된 텍스트를 음소자질[18, 19]에 따라 이진수로 변환한 데이터를 사용하였다. 물론 이렇게 하게 되면 음성언어가 제공하는 중요한 운율정보인 악센트, 억양, 쉼, 길이 등의 정보는 사용할 수 없다는 제약이 있지만, 음소자질만을 갖고도 음소결합제약 법칙은 충분히 찾을 수가 있으며, 녹음자료의 디지털화, 전처리 등이 필요하지 않게 되어 시스템이 비교적 간단해지고 실험을 준비하고 진행하는데 드리는 노력이 줄어든다는 장점이 있다. 사실 음향적 실마리는 매우 모호하며, 음성인지는 이 정보가 비음향적·조음적, 언어학적, 의미적, 상황적·실마리와 결합할 때 비로소 가능하다[20].

4.2. 시스템

모델을 선정하는 과정에서 가장 중요하게 고려한 요소는 언어란 시간성을 가지고 있으므로, 시스템은 이 시간성을 간직할 수 있는 단기기억(short-term memory)을 가져야 한다는 점이었다. 여러 종류의 인공신경망을 검토한 결과 Elman[21]의 단순회귀 신경망(simple recurrent network)을 기초로 한 모델을 개발했다. 단

1) “한국 방송 공사[한국방:송공사]”와 같은 예외도 있으나, 이 규칙의 적용을 받지 않는 예외는 합성명사로 한정되고 있다.



(그림 1) 경계발견을 위한 시스템의 구성
 (Fig. 1) Architecture of the system for boundary prediction

순회귀 신경망은 음성인식, 자연어처리 등 시간성(temporal context)을 처리해야 하는 분야에 응용되어 놀라운 성과를 거두고 있으며[22, 23, 24], Kremer[25]에 의해 단순회귀 신경망의 계산능력은 유한상태기계(finite state machine)와 대등함이 증명되었다. 본 연구에 사용한 모델은 제한된 피드백(feedback)을 갖는 3계층 인공신경망으로써, 전 단계 은닉층(hidden layer)의 내용을 문맥부(context units)에 복사하여 현 단계의 입력값과 함께 은닉층에 그 값을 전달한다. 따라서 문맥부에 현 음소 이전에 어떤 음소들이 나타났는지 하는 정보를 기억할 수 있으므로, 음소간에 연관성을 처리할 수 있으며, 따라서 여러 종류의 다른 음운현상을 설명할 수 있다.

경계발견을 위해 사용된 시스템은 (그림 1)과 같다. 이 시스템이 하는 일은 현재 입력되는 음소(current segment)를 출력층에 복사(autoassociation)하고, 또한 다음에 올 음소(next segment)를 예측(prediction)하는 것이다. 입력값을 출력층에 복사하도록 하는 이유는 그렇게 함으로써 다양한 입력값을 은닉층에서 구별해 내는데 도움을 주기 때문이다. 일정한 학습시간이 경과한 후에 예측된 음소와 실제 다음에 오는 음소간의 오류율을 측정하여 그 값이 일정기준(문턱값; threshold)을 초과할 경우에 그 부분을 경계의 후보로 택하게 된다[10]. 입력데이터의 훈련을 위해서는 역전파(backpropagation) 알고리즘을 사용한다. (그림 1)에

서 실선으로 표시된 화살표는 하위 계층의 각 노드가 상위 계층의 각 노드에 1대다(one-to-many)로 연결되어 학습이 이루어지는 연결무게(connection weights)를 나타내며, 점선으로 표시된 화살표는 은닉층의 각 노드가 문맥부의 각 노드에 1대1(one-to-one)로 학습이 없이 그대로 복사되는 연결무게를 나타낸다.

4.3 프로그램 작성 및 구현

제한된 모델을 C언어를 사용하여, 워크스테이션 상에서 구현하였다. 또한 utility프로그램으로 한글 조합형으로 된 문장을 음소로 나누어서 음성자질에 의한 이진수로 바꾸어 주는 프로그램, 입출력 데이터를 생성하는 프로그램 등도 작성하였다.

5. 실험

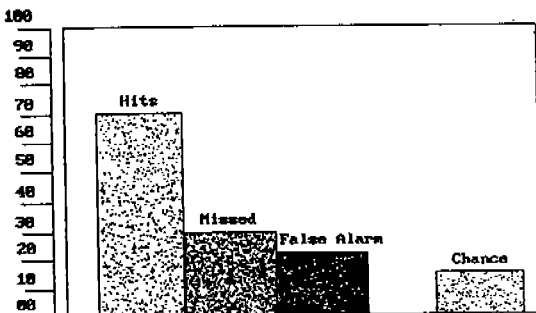
5.1 방법

구현된 프로그램을 이용하여, 실험을 하였다. (2,200 개중 사용이 불가능한 9개를 제외한) 실험에 사용된 2,191개의 문장은 중복된 단어를 포함하여 12,025개의 음운단어에, 총 69,466개의 음소를 포함하는데, 그중 1,972개의 문장, 총 10,965개의 음운단어를 사용하여 시스템을 훈련하였다. 구두점이 포함되지 않는 6개의 문장은 띄어쓰기 없이 일련의 음소열로 바뀌어 각 음소는 경음(tense), 기식음(aspirated), 유성음(voiced)

등의 13개 음소자질에 따라 코딩된 후, 한 번에 한 음소씩 입력되었다. 입력된 음소는 전단계의 은닉층값(문장초인 경우에는 미지수를 표시하는 0.5)인 문맥부의 값과 함께 은닉층을 거쳐, 출력층에서 입력음소의 값을 재생(reproduction)하고, 다음에 올 음소를 예측하도록 훈련되었다. 물론 아무리 오래 훈련을 하더라도 다음의 음소를 정확히 예측해 내는 것은 불가능하다. 왜냐하면 어떤 음소 다음에 올 수 있는 음소의 수는 여럿이기 때문이다. 단지 본 연구에서는 오류율이 단어내인 경우와 경계 바로 앞인 경우와는 크게 차이가 있을 것이라는 가설을 입증하여 단어경계를 자동적으로 찾아내기 위하여 이러한 실험방법을 고안하였다. 시스템의 입력층과 출력층은 각각 음소자질을 나타내는 13개의 노드로 구성되어 있으며, 은닉층과 문맥부에는 100개씩의 노드를 사용하였다. 각 문장은 학습률 0.25, 모멘텀(momentum) 0.9로 역전파 알고리즘을 이용하여 200번 훈련하였으며, 그 후에 각 음소에 대해 오류율(RMS error)을 측정하였고, 다음에 예측해내야 하는 음소가 단어의 일부분인지, 또는 경계인지로 나누어 각각의 경우 평균 오류율을 구하였다.

5.2 결 과

훈련에 사용된 63,312개의 음소 중 단어 내에 해당하는 54,319개의 경우 평균 오류율은 1.15이었으며, 8,993개의 경계의 경우 2.19였다. 두 값의 절반에 해당하는 1.67을 문턱값으로 하여 정확도를 측정한 결과는(그림 2)와 같다.



(그림 2) 경계의 예측률 및 우연치

(Fig. 2) The prediction of word boundaries by the model and by chance

그림에서 Hits로 표시된 부분은 경계를 정확히 예측한 경우, Missed는 경계를 예측해 내는데 실패한 경우, 그리고 False Alarm은 경계가 아닌데, 경계로 예측한 경우이다. 또한 Chance는 우연히 맞출 수 있는 총 음소 수에 대한 경계의 비율이다.

5.3 논 의

70%라는 정확도는 그 자체로서는 매우 우수한 결과라고 말할 수 없을지 모르나, 우연치(14%)와 비교한다면, 그 것도 음소자질 이외에는 다른 정보가 전혀 주어지지 않은 상태에서의 예측율인 것을 감안한다면, 매우 우수한 수치이다.

예측률을 높이기 위하여, 여러 가지 다른 값의 문턱값을 사용해 보았다. 그런데, 문턱값을 낮추면, 경계에 대한 정확도는 올라가나 false alarm도 또한 상승하며, 문턱값을 내리면, false alarm의 비율은 낮아지나, 경계에 대한 예측률은 떨어졌다.

훈련 결과 얻어진 문턱값을 시스템이 전혀 훈련하지 않은 시험용 데이터(문장수=219, 단어수=1,060, 음소수=6,154)에도 적용해 보았는데, 그 때에도 경계에 대한 예측률은 거의 비슷했다(0.702 대 0.704). 이는 훈련을 통해 스스로 학습한 음소결합 제약규칙이 국어의 어느 문장이나 적용될 수 있음을 보여주는 것이며, 여기에 본 연구에서 사용한 방법의 타당성이 있다고 하겠다. 이것이 가능한 것은 전술한 바와 같이 경계에 관한 정보는 전혀 갖고 있지 않은 시스템이 시간축을 따라 변화하는 입력자료에 대한 일반성을 은닉층에 표현하는 과정에서 통계를 통해 스스로 음운단어별로 음소결합 제약규칙을 찾아내고, 그 정보를 이용하여 경계를 예측할 수 있었음에 기인하며, 제안한 단순회귀 신경망이 이러한 종류의 과제(task)에 적합함을 입증하는 것이라 하겠다.

한가지 특기할 점은 단어처음에서의 오류율과 그 외의 오류율을 비교했을 때, 단어 초에서의 오류율이 훨씬 낮다(0.78 대 1.38)는 것이다. 이는 일단 단어의 처음인 것을 알면, 그 다음에 오는 음소는 단어절 단어의 경우 외에는 경계일 수가 없으며, 현재 입력된 음소(초성)가 자음이면, 다음에 오는 음소(중성)는 반드시 모음이며, 현재 입력된 음소가 모음이면, 다음에 오는 음소는 대개 자음임을 통계적으로 쉽게 알아낼 수 있었기 때문인 것으로 판단된다.

6. 결 론

본 연구에서는 음성인식·합성을 위한 운율음운론의 연구동향을 개관하고, 국어의 음운단위와 경계의 설정에 관한 이론적·실험적 고찰을 정리하였으며, 음운단위에 따른 경계의 자동적 발견을 위하여 데이터를 수집하고 시스템을 구현하여 실험을 행하였다.

실험방법으로는 대량의 데이터를 단순회귀 신경망에 입력한 후, 시스템이 스스로 음소결합 제약법칙을 학습하도록 하여 음운단어의 경계를 찾도록 하였다. 이러한 방법은 어린아이가 처음 언어를 배울 때, 사용하는 방법과 유사하다고 할 수 있겠다. 예를 들면, /ㄱ/+/-/+/ㄷ/+/ㅁ/라 하는 음소열이 주어졌을 때, 이 때까지 학습한 결과로 "ㄷ"이 종성에 올 수 없음을 알기 때문에, "ㄱ/ㅁ"와 같이 경계를 설정하지, "ㄱㅁ/에"로는 하지 않을 것이다.

본 연구에서는 음성언어를 사용하기는 하였지만, 실제 자료(raw speech data)보다는 전사된 자료를 사용하였다. 실제자료를 사용하여 같은 방법의 실험을 한다하더라도, 예측율은 감소하지 않으리라 기대된다.

본 연구의 결과를 그대로 음성인식·합성에 적용하기에는 정확도가 다소 떨어진다. 그러나 이러한 방법이 다른 정보(실제 음성자료에서 추출한 쉘, F_0 , 길이 등을 이용한 억양, 동시조음, 장음화, 리듬 등)와 함께 사용된다면, 음성인식·합성시에 음운단어 경계를 찾아내어 정확한 분절화를 행할 수 있으리라 여겨진다.

본 연구는 계산음운론 연구의 본질에 대한 새로운 지평을 제시하고, 음성인식 및 음성합성과 운율음운론 연구의 사이에 보다 명료한 관계를 보여주며, 운율음운론 연구, 음성인식, 그리고 음성합성의 연구에 새로운 방향을 제시하는데 일조를 하였을 것으로 기대된다.

참 고 문 헌

- [1] Cole, R. et al., "The Challenge of Spoken Language Systems: Research Directions for the Nineties," IEEE Trans. Speech and Audio Processing, Vol. 3, No. 1, pp. 1-21, 1995.
- [2] Church, K., Phonological Parsing in Speech Recognition, Kluwer Academic Pub, 1987.
- [3] Price, P. et al., "The use of prosody in syntactic disambiguation," J. Acoust. Soc. Amer., Vol. 90, pp. 2956-2970, 1991.
- [4] Chen, F. and M. Withgott, "The use of emphasis to automatically summarize a spoken discourse," Proc. 1992 ICASSP, pp. I. 229-232, 1992.
- [5] Wightman, C. and M. Ostendorf, "Automatic recognition of prosodic phrases," Proc. IEEE ICASSP, pp. 321-324, 1991.
- [6] Wightman, C. and M. Ostendorf, "Automatic recognition of intonational features," Proc. IEEE ICASSP, 1992.
- [7] Wightman, C. and M. Ostendorf, "Automatic labeling of prosodic patterns," IEEE Trans. Speech and Audio Proc., Vol. 2, No. 4, pp. 469-481, 1994.
- [8] Cartwright, T. and M. Brent, "Segmenting speech without a lexicon: the roles of phonotactics and speech source," Proc. 1st mtg, ACL SIG Comp. Phonology, pp. 83-90, 1994.
- [9] Brent, M., A. Gafos, and T. Cartwright, "Phonotactics and the lexicon: beyond bootstrapping," Unpublished ms, 1994.
- [10] Schillcock, R. et al, "Connectionist Modelling of Phonological Space," Ellison and Scobbie (eds), Computational Phonology, Edinburgh Working Papers in Cognitive Science 8, pp. 179-195, 1993.
- [11] Gaskell, M., Spoken Word Recognition: A combined computational and experimental approach, Ph. D. Dissertation, University of London, 1994.
- [12] Lee, C. -D., Learning to Perceive and Produce Words in Connectionist Networks, (Technical Report 334), Department of Computer Science, Indiana University, 1991.
- [13] 이 찬도, "인공신경망을 이용한 한국어 형태음운 현상 연구," 정보처리논문지, 제2권, 제2호, pp. 215-228, 1995.
- [14] 정 국 의, "음성인식/합성을 위한 국어의 음성-음운론적 특성 연구," 한국음향학회지, 제13권, 제6호, pp. 31-44, 1994.

- [15] 정 국, 음운단위 음운론 및 음운통계에 관한 연구, 한국전자통신연구소, 1993.
- [16] Kaye, J., Phonology: a Cognitive View, Lawrence Erlbaum Associates, 1989.
- [17] 정 국 외, 한국어 특질에 관한 연구: 자동 통역전화 시스템 구현을 위한 음운 및 문법구조 연구, 한국통신 연구개발단, 1993.
- [18] 허웅, 국어음운학 (중판), 샘문화사, 1989.
- [19] Chomsky, N. and Halle, M., The Sound Pattern of English, Harper and Row, 1968.
- [20] 고 도홍, 구 회산, 김 기호, 양 병근 공역, 음성언어의 이해, 한신문화사, 1995.
- [21] Elman, J., "Finding structure in time," Cognitive Science, 14, 179-211, 1990.
- [22] Elman, J., "Structured representations and connectionist models," Proceedings of the Eleventh Annual Conference of the Cognitive Science Society, pp. 17-25, 1989.
- [23] Gasser, M. and Lee, C. -D., "Networks that learn about phonological feature persistence," Connection Science, No. 2, pp. 265-278, 1990.
- [24] Hare, M., "The role of similarity in Hungarian vowel harmony: a connectionist account," Connection Science, No. 2, pp. 123-150, 1990.
- [25] Kremer, S. C., "On the Computational Power of Elman-Style Recurrent Networks," IEEE Transactions on Neural Networks, Vol. 6, No. 4, pp. 1000-1004, 1995.



이 찬 도

- 1975년 서울대학교 사범대학 독
일어과(문학사)
- 1984년 미국 Arizona State Univer-
sity 독일어과(M.A.)
- 1987년 미국 Indiana University
전산과학과(M.S.)
- 1991년 미국 Indiana University
전산과학과(Ph.D.)

1991년~1992년 KAIST 인공지능 연구센터 선임연구원
1992년~현재 대전대학교 정보통신공학과 부교수
1994년~1996년 정보처리학회지 편집위원
관심분야: 인공지능, 음성언어 처리, 인공 신경망, 유
전 알고리즘 등