DOI: 10.3745/KIPSTB.2008.15-B.6.595

한국어 발화음성에서 중점단어 탐색을 위한 기본주파수에 대한 연구

권 순 일[†] · 박 지 형[††] · 박 능 수[†††]

요 약

각 문장 별 중점단어는 발화음성을 인식하고 그 의미를 이해하는데 도움을 준다. 발화된 음성신호로부터 중점단어를 탐색할 수 있는 방법을 찾기 위한 노력의 일환으로 실험을 통하여 문장 내에서 중점단어와 그 외의 단어들의 기본주파수의 평균과 분산, 그리고 평균 에너지를 분석해 보았다. 한국어로 된 100개의 발화문장의 음성데이터를 가지고 실험을 한 결과 중점단어는 그 외의 단어들에 비해 대부분 상대적으로 높은 기본주파수의 평균값을 나타내거나 상대적으로 높은 기본주파수의 분산 값을 나타냈다. 이 연구 결과를 이용하면 한국어의 구어문장에서 운율적 특성을 알 수 있을 뿐만 아니라, 자연어 처리를 이용한 핵심어를 추출하는 데에도 도움이 될 것이다.

키워드 : 중점단어 탐색, 핵심어 추출, 기본주파수, 한국어 발화음성, 운율

A Study of Fundamental Frequency for Focused Word Spotting in Spoken Korean

Kwon, Soonil[†] · Park, Ji Hyung[††] · Park, Neungsoo[†††]

ABSTRACT

The focused word of each sentence is a help in recognizing and understanding spoken Korean. To find the method of focused word spotting at spoken speech signal, we made an analysis of the average and variance of Fundamental Frequency and the average energy extracted from a focused word and the other words in a sentence by experiments with the speech data from 100 spoken sentences. The result showed that focused words have either higher relative average F0 or higher relative variances of F0 than other words. Our findings are to make a contribution to getting prosodic characteristics of spoken Korean and keyword extraction based on natural language processing.

Keywords : Focused Word Spotting, Keyword Extraction, Fundamental Frequency, Spoken Korean, Prosody

1. 서 론

최근 다양한 모바일 기기들이 대중화되면서 여러 가지 사용자 인터페이스 방법들 중의 하나로 음성인식에 대한 관심이 증가하고 있다. 음성인식에 대한 연구는 전 세계적으로 이미 반세기 동안 수행되어 왔다. 태동기의 음성인식은 한두 개의 단어 또는 숫자정도를 인식하는 수준이었지만, 음성을 인식하고 이해하기 위한 다양한 연구 분야의 방법들이 접목되고 개발되면서 인식가능한 단어의 수와 문장의 복잡도가 점차 증가되었고, 최근에는 긴 문장에 대한 음성인식이 가능해 졌으며, 복수 개 이상의 문장이 계속되는 상황에서도 음성이 연속적으로 문자화 되는 기술에 대한 연구가 성과를 보이고 있다. 또한 다자간의 회의를 사람들 간의 중요한 교류의 도구로 보면서 회의 환경 속에서 다자간의 대화를 음성인식을 통해 문자화 하고, 이를 이해하여 그 의미를 파악할 수 있는 시스템 개발에 대한 연구가 활발히 진행되고 있다[1,2].

다자간의 협업 내지는 협의가 이루어지는 회의 환경에서 사람들 간의 대화는 그 중심이 되는 정보임에도 불구하고 저장, 관리, 검색, 및 주요 내용에 대한 자동 파악에 대한 어려움으로 인해 수기로 요약 기록되거나, 전체를 녹음 또는 녹화해 놓는 정도의 원시적인 방법이 여전히 사용되어 오고 있다. 특정인에 의해 회의록이라고 하는 기록을 작성하거나 회의 참가자 각자가 기록을 하는 방법에 있어서는 기록하는 사람이 다소간 주관적으로 요약을 할 수 있기 때문에 주요 사안의 핵심을 놓칠 수가 있다. 또한 회의내용 전체를 녹음/녹화하면, 전체 중의 일부 내용을 청취하고자 할 때 맨 처

† 정 회 원 : 한국과학기술연구원 지능인터랙션연구센터 선임연구원
†† 정 회 원 : 과학기술연합대학원대학교 HCI 및 로봇응용공학 교수
††† 종신회원 : 건국대학교 정보통신대학 컴퓨터공학부 부교수
논문접수 : 2008년 7월 25일
수 정 일 : 1차 2008년 10월 15일, 2차 2008년 11월 12일
심사완료 : 2008년 11월 13일

음부터 끝까지 모든 내용을 듣고 보는 수밖에 없을 것이다. 이런 불편을 없애고 효과적인 회의 또는 협업 진행을 위해서는 효율적인 음성정보의 저장, 관리, 검색 기능이 필요하다. 이러한 기능을 구현하기 위해서는 대화의 음성신호를 직접 분석하여 내용을 파악하고 주제어 별로 대화 내용을 적절히 분류를 해서 인덱스를 달아놓을 수 있어야 하며, 이는 추후 지난 회의 내용에 대한 검색도 효율적으로 할 수 있을 뿐만 아니라, 여러 가지 연관된 회의 자료나 기타 웹으로부터 얻어지는 자료들도 구축된 자료구조에 맞추어 저장되고 관리될 수 있게 해준다.

사람들 간의 대화 또는 회의의 내용이나 주제를 파악하기 위해서는 단순한 음성인식이 아닌 음성을 이해하는 시스템에 대한 연구가 필요하다. 음성을 이해하기 위해서 대체로 음성인식을 통해 음성신호를 문자화 하고, 이후 자연어 처리를 통해 그 의미를 파악하는 순으로 시스템이 구성되어 있다. 하지만 자연어 처리의 대상은 대체로 신문이나 책, 논문 등 각종 문자화 되어있는 정보, 다시 말해 문자로만 이루어진 정보 문장들이다. 반면 사람들 간의 대화와 같은 발화된 구어의 경우 그 의미를 이해하는데 있어서는 언어적인 요소 못지않게 비언어적인 요소를 활용된다고 한다. 비언어적인 요소란 대화 시에 사용되는 얼굴표정, 몸동작, 말소리의 크기, 말의 속도 등을 의미 하며, 사람들은 이러한 요소들을 복합적으로 활용하여 상대방 대화의 정확한 의미를 이해하게 된다 [1,2,7,10,14]. 또한 사람들 간의 대화에 있어서는 문자와 화자의 발성을 통한 음성신호가 합쳐져 있기 때문에 문자가 지니고 있는 의미 정보와 함께 발성을 하면서 포함된 화자의 의도가 포함되어 있다. 하지만 음성이 음성인식과정을 통해 문자화 되고나면 이러한 화자의 의도가 담긴 정보가 사라지게 된다. 화자가 발화시에 표현하는 정보 중의 하나는 중점단어(Focused Word) 이다. 이는 화자가 한 문장 안에서 가장 강조하고자 한 단어이면서 전체적인 대화의 흐름 및 토픽을 추출하는데 도움을 줄 수 있는 부가정보가 될 수 있다. 이 논문에서는 중점단어를 찾기 위해 필요한 비언어적인 요소들 중에서 음성신호로부터 추출되는 운율적 특징요소들을 분석해 보려한다. 이번 연구에서 실시하는 실험적 분석을 통해 얻게 되는 결과를 이용하여 한 문장 안에서 말하는 사람이 가장 강조하고 하려고 한 단어가 무엇인지, 즉 중점단어(Focused Word)가 어느 것인지를 찾아낼 수 있다면, 이는 사람들 간의 대화로부터 핵심어 및 주제를 추출하는데 도움이 될 것이다 [3-6].

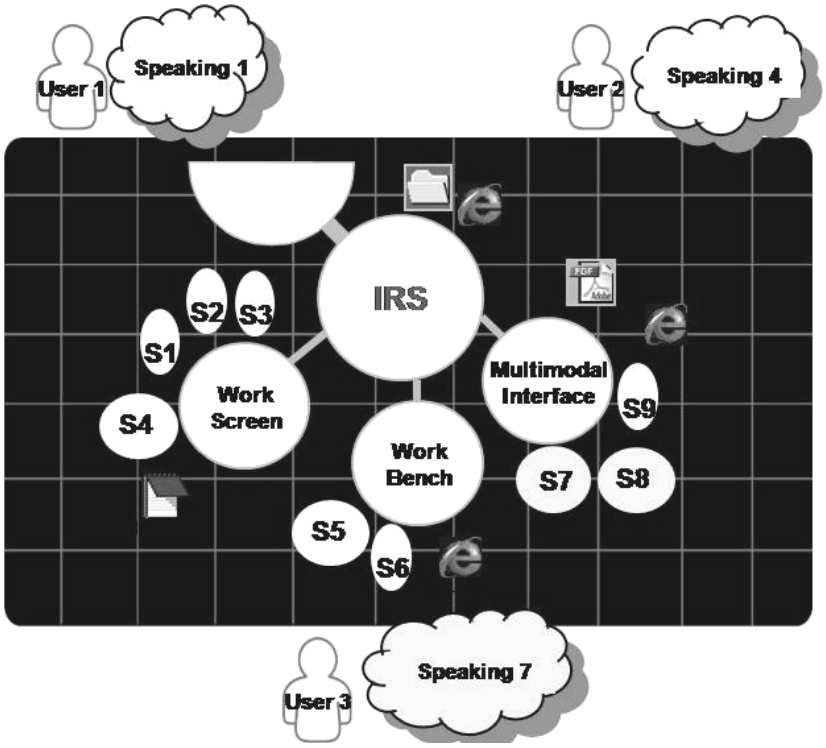
본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 발화음성신호와 관련된 주요연구들에 대해 소개하고 이 연구와 비교해 본다. 제 3장은 연구개발의 대상인 토픽추적 시스템에 대한 설명에 이어, 운율적 특징들에 대하여 설명하고, 운율적 특징들 중 기본주파수와 에너지의 한국어 발화음성의 중점단어탐색에 있어서 그 의미를 설명한다. 제 4장은 본 논문에서 제시한 핵심어와 비핵심어 구분하기 위한 기본주파수와 에너지의 변화 추적 및 분석 실험의 구성에 관하여 기술한 후, 제 5장에서 실험 결과에 대해 자세히 분석하고 유용성을 제시한다. 제 6장은 본 논문의 결론을 내린다.

2. 관련연구

운율적 특징(Prosodic Feature)을 이용한 연구는 그동안 음성정보처리와 자연어 처리 분야, 그리고 음성학 분야에서 꾸준히 있어왔다. D. Wang과 S. Narayanan은 대화체 영어에서 말의 빠르기를 이용해 문장들 간의 구분을 위한 연구를 했는데 [13,14], 영어 발화문장의 피치정보의 연속성만을 측정하였고, 이를 통해 언어나 문자정보 없이 문장의 시작과 끝을 구분할 수 있다는 가능성을 보여주었다. E. S. Kim과 B. Scassellati는 운율적 특징 정보를 로봇을 학습시키는데에 활용하였는데 [7], 영어운율정보 중 각 문장전체의 피치 평균값과 에너지 변화량을 이용하여 로봇과의 대화자가 찬성하는지 반대하는지를 판단하고, 이를 통해 로봇을 훈련시키려는 시도를 하였다. K. Sönmez 등은 동적인 운율적 특징을 모델링하여 화자인증에 이용하는 연구를 수행했는데 [10], 사람마다의 영어발화 문장전체의 기본주파수 변화를 패턴화하고 통계적 모델링하여 화자를 인증하는 연구결과를 보여주었다. S.-A. Jun, H.-J. Lee, H.-S. Kim 등은 운율적 특징을 이용하여 한국어의 음성학적 관점에서 초점 투사(Focus Projection)에 관한 연구를 계속하고 있다 [3,5,8]. 이 논문에서는 기존의 초점 투사 연구와 관련하여 음성신호처리 관점에서 접근을 하고 있고, 한국어 발화음성을 대상으로 하고 있다. 또한, 문장단위의 상대적 중점단어를 찾아내기 위해 운율적 특징들의 패턴을 음소/음절 단위가 아닌 단어 단위로 관찰하고 분석하고 있는데, 그 이유는 한국어가 다른 외국어들에 비해 음소 내지는 음절 별 강세의 정도가 뚜렷하지 못하여 측정하기 힘들 뿐만 아니라 패턴을 찾아내기도 힘들다. 이와 더불어 한 문장 안에서 화자가 강조하고자 하는 의미를 담고 있는 최소한의 단위는 음절보다 큰 단어 단위 이상으로 추측되고 있으며, 이번 연구 결과를 자연어 처리결과로 얻어질 수 있는 문장 별 핵심어와 비교 분석을 하고자 하는 목적이 있다.

3. 한국어 발화음성의 중점단어 탐색

(그림 1)은 다자간 협업 또는 회의를 위한 핵심어 및 주제어 추적을 통하여 참가자들에게 회의 진행 상황과 각종 관련 자료를 효율적으로 제공해 주는 시스템의 예를 도시한 것이다. 세 명이 회의에 참여하고 있고, 가운데에는 테이블탑 시스템이 갖추어져 있어, 참여자들이 공유하는 디스플레이가 지원되고 있다. 회의 내용에 있어서 예를 들자면, 세 명의 참가자들이 '지능형 반응공간 (Intelligent Responsive Space)' 이라는 주제를 놓고 아이디어 회의를 한다고 했을 때, 회의의 흐름에 따라 각각 하위레벨의 주제어들이 달라질 수 있는데, 그 예가 'Work Screen', 'Work Bench',



(그림 1) 회의 환경에서의 토픽추적 시스템 일례

'Multimodal Interface' 등이다. 즉, 회의 중 대화의 흐름이 'Work Screen' 이었다가 중간에 'Work Bench' 라는 주제로 이야기가 전환되고, 마지막에는 이러한 것들과 사람간의 인터페이스 측면에서 'Multimodal Interface'에 관해 대화를 하게 되는 시나리오이다. 이러한 회의 과정에서 대화의 내용을 분석하고, 주제어들의 변화를 감지하며, 이들을 추출하여 상관관계를 구조화 하는 것이 전체 시스템의 가장 핵심이 되는 기술이다.

회의 중 발화된 대화내용에 대한 분석 및 분류를 하기 위해서는 기본적으로 음성인식과 자연어처리의 기술이 필요하다. 음성인식을 통하여 음성신호를 문자화 해주고, 자연어처리 방법을 이용하여 문자화 된 문장들로부터 핵심어, 주제어, 그리고 문맥을 파악하게 된다. 음성인식시스템에서는 입력 음성신호를 사전적 정보와 언어모델, 음성모델 등을 이용하여 문자화 해준다. 이때 출력되는 정보는 음성신호에 포함되어 있는 언어정보내지는 문자정보만을 지니고 있게 된다. 이 문자중심의 언어정보로 부터 형태소분석, 구문분석, 의미분석 등을 통해 핵심어와 주제어를 추적하게 된다. 하지만 이러한 언어정보만으로는 부족한 면이 있다. 음성 신호는 언어가 물리적 진동으로 표현된 것으로 언어적 의미를 담고 있는 동시에 음향적 특징들 (Acoustic Features)도 함께 가지고 있기 때문에 문자화된 언어정보 이외의 정보를 포함하고 있다.

음향적 특징들 중에는 운율적 특징들이 있는데, 이것들을 적절히 활용하면 발언의 목적 또는 종류 (질문, 진술, 소리지름 등)를 알아내고, 애매한 문장의미를 이해하고, 음성을 인식하는데 도움이 된다. 운율적 특징들로는 기본 주파수, 에너지, 묵음 길이, 음절의 지속시간 등이 있다 [1,2,13,14]. 묵음 길이는 문장이나 절의 구분 점과 눌변(Disfluency), 즉 말 더듬을 찾아내는데 유용하게 쓰인다 [13]. 음절의 지속시간은 그 음절을 포함하고 있는 단어가 강조가 되었는지를 알아내는데 활용된다. 사람은 말을 할 때, 어떤 단어를 강조하다보면 무의식적으로 그 단어의 특정 음절을 길게 발음하게 된다는 원리를 활용한 것이다 [14]. 또한, 에너지와 기본주파수의 여러 통계치(최댓값, 최솟값, 평균, 분산 등)를 추적하면 특정문장 안에서 어떤 단어가 강조되었는지 구분이 가능하다 [12,14]. 예를 들자면, 방 안의 책상위에 파란색 색연필, 빨간색 유성펜이 놓여 있다. 화자가 도화지에 파란색 색칠을 하기위해 자신 보다 책상에 가까이 있는 친구에게 파란색 색연필을 가져다 달라고 부탁한다고 가정해 보자. 화자는 상대방에게 '파란색 색연필 좀 가져다 줘' 라고 말할 때, 자연스럽게 '파란색' 이라는 단어를 가장 강조하여 말하게 된다. 왜냐하면 자신의 의도를 전달하는데 있어서 '파란색'이 가장 중요한 정보라고 생각되기 때문이다. 다시 말해, 책상 위에는 한 개의 색연필과 한 개의 유성펜이 놓여 있는데, 그 중에서 '파란색'이란 단어가 가장 중요하면서 정확하게 전달되어야 한다는 생각이 발화 시에 반영이 되는 것 이다. 하지만 색깔보다는 지우개로 잘 지워질 수 있도록 무언가를 그리려고 해서 색연필을 사용하고자 한다면, '색연필' 이라는 단어를 더 강조하게 될 것이다. 그런데, 만약 앞뒤 문맥이 충분하지 않은 상황이라면 언어적인 정보만으로 이러한 의도 내지는 상황에 대해 정확히 인지하기는 어려울 것이다. 음성신호를 음성인식기가 인식하고 나면, '파란색 색연필 좀 가져다 줘' 라는 글자만 남아있기 때문에 발화시의

상황에서 파란색이 중점이 되는 정보인지 색연필이 중요정보인지는 알 수가 없다. 그러므로 발화된 음성신호로부터 정확한 정보를 얻기 위해서는 음성인식과 자연어처리 외에 음성신호처리를 통해 특정한 음향적 특성요소들을 찾아내고, 분석하는 것이 필요하다. 음성신호로부터 추출되는 음향적 특성을 분석하게 되면 화자에 의해 가장 강조된 단어를 알아낼 수 있고, 이를 자연어처리에서 활용한다면 기존의 자연어처리만 사용할 때보다 더 정확하게 핵심정보를 파악할 수 있게 된다.

음성은 언어에 밀접하게 연관되어 있어서 연구에 대상이 되는 언어에 따라 다른 방법이 필요할 수 있다. 기본적인 원리에 대해서는 일반적으로 영어를 대상으로 연구가 되어 왔다. 게다가 한국어는 세계적으로 주류로 인정받지 못하고 있어서 이를 대상으로 하는 연구가 아직 미흡한 실정이다. 한국어 발화음성은 영어의 발화음성과 비교해 볼 때 운율적 특성에 있어서 다음과 같은 다른 점들이 있다. 대체로 억양이나 어조(Intonation)의 운율적 단위가 영어의 음절단위의 강세구(Accentual Phrase)와 Intonation Phrase의 중간인 Intermediate Phrase를 형성한다 [4]. 또한, 영어에 비해 강세 받는 음절과 인접음절간의 고처차이가 상당히 적고, 중점단어는 피치의 범위가 넓어지고, 첫 번째 음절이 길어지며, 그 뒤들 따르는 단어의 피치의 범위는 급격히 줄어드는 경향이 있다 [3-5]. 그런데 이러한 특성들이 주로 기본주파수와 에너지와 깊은 관련이 있어 보인다. 게다가 중점단어와 그 외의 단어에 대한 구분이 강세나 억양과 밀접한 관계가 있을 것으로 짐작된다. 그래서 기본주파수와 에너지에 대해 한국어 발화음성신호들은 어떠한 패턴을 보여주는지를 면밀히 관찰해 보았더니 다음과 같은 특징을 발견할 수 있었다. 중점단어는 그 외의 단어들에 비해 기본주파수의 평균값이 상대적으로 높았지만 그렇지 않은 경우가 종종 있었다. 즉 중점단어 기본주파수의 평균값이 그 외의 단어의 것과 별 차이가 없거나 오히려 낮기도 했다. 하지만 이러한 경우들에 있어서는 대체로 중점단어 기본주파수 분산이 그 외의 단어에 비해 높았다. 이에 비해 에너지는 어떤 일관된 패턴을 찾기 힘들었다. 우리는 이러한 점들에 착안하여 한국어의 발화음성에 있어서 중점단어는 그 외의 단어들과 기본주파수와 에너지의 변화에 있어서 개별적으로 또는 상호작용하는 가운데 어떻게 다른지 실험을 통해 그 패턴을 분석하고 앞선 관찰결과를 검증해 보려한다.

4. 실 험

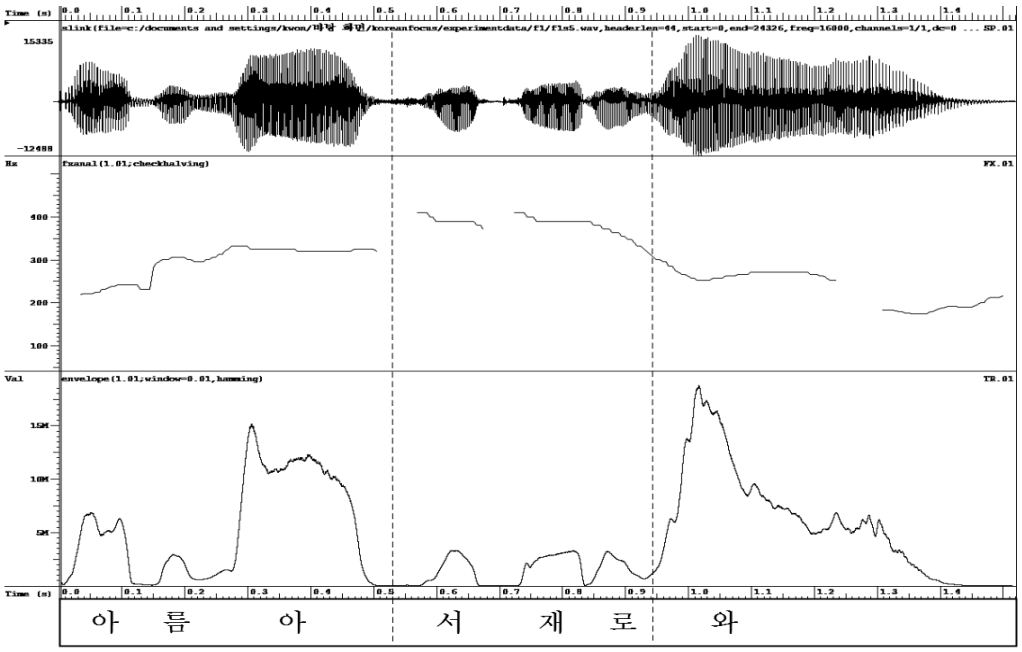
이 논문에서는 한국 구어에서 중점단어를 찾아내는 실험을 하기 위해 한국전자통신연구원에서 구축한 표준형 한국어 언어/음성 데이터모음들 중 핵심어 검출용 음성인식을 위한 데이터모음 중의 일부를 사용하였다. 음성 데이터는 대화체 발화음성으로 구성되어있고, 여성 5명과 남성 5명이 각각 10개 문장, 총 100개의 문장이며, 내용은 정보검색에 관련된 것으로 3개 내지 4개의 단어로 구성된 문장들을 골랐다. 문장들은 평서문이나 명령문들이며, 예를 들어 '아름아 거실로 와', '엄마가 좋아하는 프로그램이야' 등이 있다. 녹음환경은 사무실환경으로 화자와 마이크간 거리는 0.5 미터이고, 마이크는 Shennheiser MD431-II 이다. 샘플링 주파수는 16kHz이고, 데이터 형식은 16bit Linear PCM 으로 되어 있다.

기본주파수와 에너지는 개별적인 분절음뿐만 아니라 음절유형에 따라 달라질 수 있으므로 [15], 여러 음절이 결합된 낱말 내지 단어 단위로 평균값을 측정한다. 기본주파수를 측정하기 위해서 Speech Filing System[11]이라는 소프트웨어를 이용하였데, 이 소프트웨어는 University College London에서 음성학과 언어학을 연구하기위해 만들었으며, 음성신호를 입력해 주면 에너지부터 포먼트 등에 이르는 각종 음성신호분석 데이터 값들을 추출해 줄 뿐만 아니라 이를 그래프로 보여준다. 이를 이용하여 음성신호로부터 Autocorrelation 방식을 이용하여 기본주파수를 예측하였다. 이러한 방법은 B. Secrest 와 G. Doddington가 제안했던 Pitch Tracking Algorithm이 활용되었다 [9,11]. 에너지 또한 Speech Filing System을 이용하였는데, Rectangular window를 씌워 얻어진 샘플 값들을 제곱하여 합하는 방법으로 구해졌다. 그리고 기본주파수와 에너지의 상댓값들은 각 문장 전체의 기본주파수 평균값과 에너지 평균값과의 차이를 구한 값들이다.

각 문장별 단어의 구분은 사람에 의해 이루어졌고, 각 단어별 기본주파수의 평균과 분산, 그리고 에너지의 평균값을 측정하였다. 각 문장마다 중점단어를 정하는데 있어서 실제 사람이 발화 문장들을 들어보고 인식되는 결과를 이용하기 위해, 10명의 사람들이 듣고 각 문장 별 상대적으로 화자에 의해 가장 강조되었다고 여겨지는 단어를 정하고, 다수가 선택한 결과를 바탕으로 최종 중점단어로 선정하였다. 일단 각 문장별로 반드시 한 단어가 중점단어라는 전제하에 선정된 것이라 상대적으로 강조의 정도가 덜 두르러진 경우도 있을 수 있다.

5. 분 석

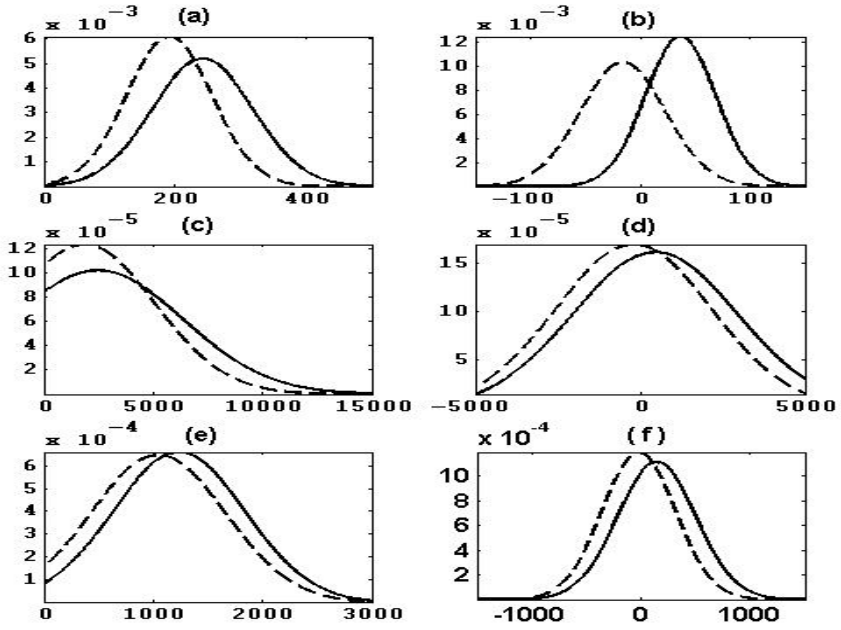
(그림 2)는 이번 실험의 대상인 발화된 100개의 문장들 중의 하나의 예를 Speech Filing System을 이용하여 얻어진 기본주파수 평균, 기본주파수의 분산, 에너지 평균의 데이터 값을 토대로 그래프로 표현 된 모습을 보여준 것이다. 이 문장은 세 개의 단어로 구성되어 있고, 성인 여성의 음성이다. 첫 번째 열의 그래프는 시간 축에 따른 PCM 샘플 값들로 표현된 음성신호를 보여주는 것이고, 두 번째 그래프는 이 음성신호의 기본주파수를 추적한 결과를 보여주는 것으로, 첫 번째 단어는 낮은 주파수에서 높은 주파수 대역으로 올라갔고, 두 번째 단어는 상대적으로 높은 기본주파수를 보였으며, 세 번째 단어는 높은 주파수에서 낮은 주파수 대역으로 다시 내려오는 추세를 보였다. 세 번째 그래프는 에너지를 측정한 것인데, 한 눈에도 첫 번째와 세 번째 단어에 비해 두 번째 단어의 에너지가 대체로 작은 것을 볼 수



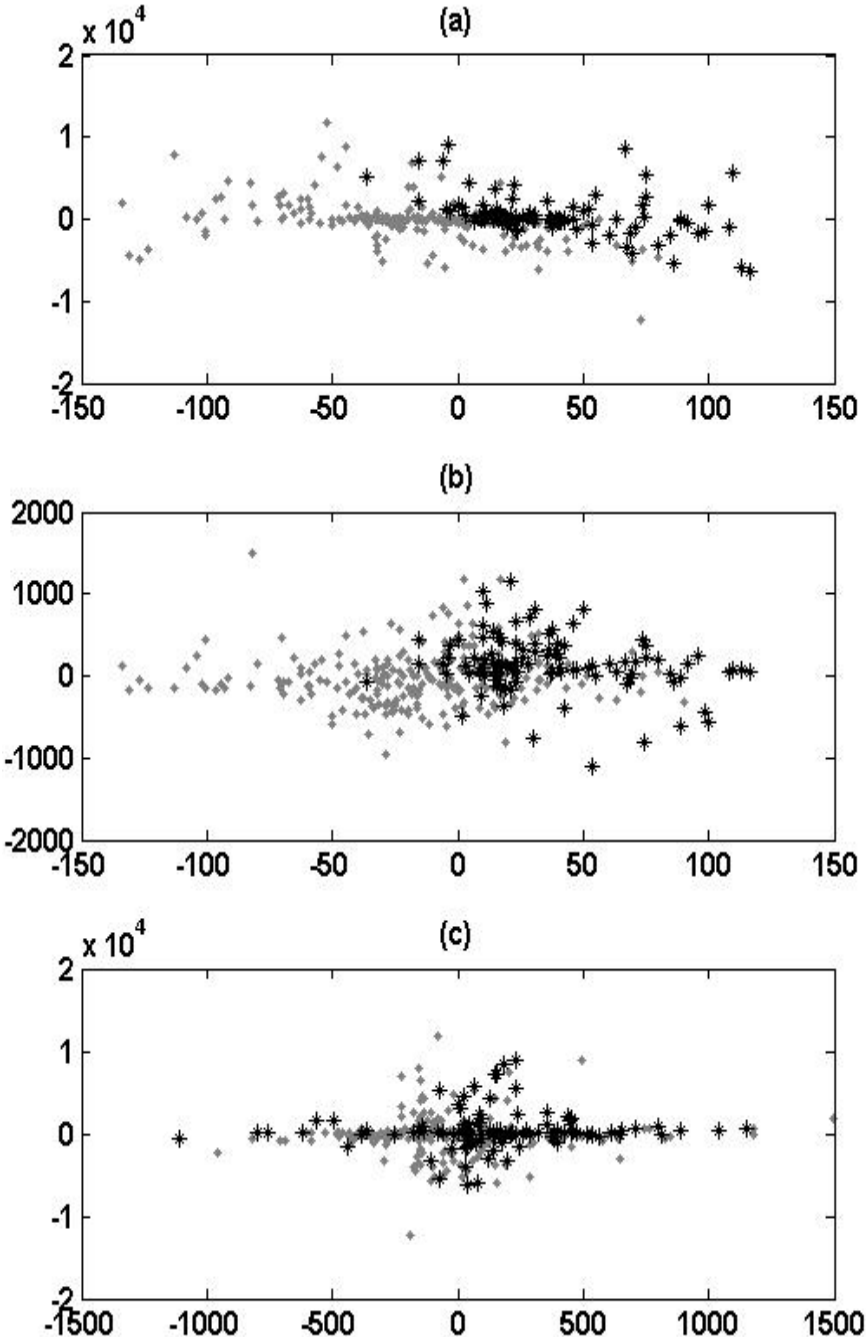
(그림 2) "아름아 서재로 와"라는 음성관련 그래프들; 공통으로 x축은 시간이고, y축은 첫 번째 그래프는 Linear PCM 음성샘플 값, 두 번째 그래프는 기본주파수, 세 번째 그래프는 Linear Scale의 에너지 값임

있다. 사람에 의해 미리 수행된 분석에서 두 번째 단어가 가장 강조된 단어로 판단되었다. 즉 '서재로'라는 단어가 화자에게 있어서 가장 중요한 정보로 여겨졌다는 의미이다. 이를 유추해서 생각해 보자면, 아름이가 현관에 서 있는데, 안에서 화자가 거실이나 침실이 아닌 서재로 오라고 말한 것으로 볼 수 있다. 그래프 상으로도 기본주파수나 에너지와 중점단어의 상관관계가 있을 것 같다는 추측을 해볼 수 있었지만, 이러한 상황을 염두하고 음향적 특성들을 분석함에 있어서 그래프가 아닌 각 데이터 값들의 통계적 분석을 통하여 발화된 문장에서는 에너지 값이 중점단어와의 상관관계가 있는지는 더 자세한 분석위한 실험 결과는 (그림 3)과 (그림 4)에 나타나 있다.

(그림 3)에서는 기본주파수 평균, 기본주파수의 분산, 에너지 평균값들이 중점단어와 그 외의 단어를 구분하는데 얼



(그림 3) 각 운율적 특징 별 중점단어와 비중점단어의 확률밀도함수(PDF) 비교 (실선은 중점단어의 PDF 이고, 점선은 비중점단어의 PDF임.): (a)기본주파수 평균의 절댓값 비교, (b)기본주파수 평균의 상댓값 비교, (c)기본주파수의 분산의 절댓값 비교, (d) 기본주파수 분산의 상댓값 비교, (e)에너지 평균의 절댓값 비교, (f)에너지 평균의 상댓값 비교



(그림 4) 중점단어(검정색*)와 그 외의 단어(회색 ·)의 운율적 특징 비교: (a)x축은 기본주파수 평균의 상댓값, y축은 기본주파수 분산의 상댓값, (b)x축은 기본주파수 평균의 상댓값, y축은 에너지 평균의 상댓값, (c)x축은 에너지 평균의 상댓값 비교, y축은 기본주파수 분산의 상댓값

마나 역할을 할 수 있는지 알아보기 위해 각각의 가우시안 분포 기반의 확률밀도함수(Probability Density Function, PDF)를 구하여 비교해 보았다. 왼쪽 열인 (a), (c), (e)는 각각의 절댓값을 가지고 비교를 한 반면, 오른쪽 열인 (b), (d), (f)에서는 각 문장 별 절댓값들의 평균값과의 차이를 구한 상대적인 값으로 도시해 보았다. 절댓값과 상댓값을 비교해 본 이유는 사람의 성별뿐만 아니라 각 개인별로도 발화시 주변상황이나 문장내용, 그리고 발화습성 등에 따른 기본주파수 및 에너지 값의 변이성이 중점단어와 그 외의 단어를 구분하는데 있어서 어느 정도 영향을 주는지를 가늠해 보기 위함이었다. 비교결과 전체적으로 절댓값보다는 상댓값이 중점단어와 그 외의 단어를 조금 더 구별시켜주는 경향을 나타냈다. 이는 사람마다 기본주파수의 영역과 발화강도가 차이가 날 수 있고, 문장마다도 약간 차이를 보일 수도 있기 때문에 절댓값들의 전체적인 분포가 넓어질 수 있기 때문이다. 상댓값으로 구하여진 분포들만 비교해 보더

라도 기본주파수의 평균에 있어서는 중점단어와 그 외의 단어 간에 큰 차이를 보여주었다. 하지만 기본주파수의 분산과 에너지는 그 차이가 다소 미미했다. 이 실험의 결과를 볼 때, 세 가지 특성 중 한가지만을 사용한다면 기본주파수의 단어별 평균의 상댓값으로 중점단어와 그 외의 단어들을 구분하는 것이 가장 좋은 결과를 가져다 줄 것으로 분석할 수 있다. 이 실험 결과를 바탕으로 기본주파수의 평균, 분산, 그리고 에너지 평균의 상댓값들을 짝을 이루어 활용할 경우 중점단어 탐색에 도움을 줄 수 있는지를 추가적으로 실험해 보았다.

(그림 4)는 기본주파수 평균, 기본주파수의 분산, 에너지 평균들의 서로간의 시너지 효과에 관한 비교 및 분석을 하기위해 수행했던 실험의 결과이다. 앞의 실험결과에서도 알 수 있었듯이 에너지의 평균은 중점단어와 그 외의 단어를 구분하는데 큰 역할을 못하고 있다. 그래서 (그림 4) (b)와 (c)에서는 중점단어와 그 외의 단어의 운율적 특징들이 서로 많이 겹치는 모습을 보이고 있다. 또한 (그림 3)의 (c)와 (d)에서 볼 수 있듯이 기본 주파수의 분산도 중점단어와 그 외의 단어를 구분하는데 도움이 많이 되지 못했다. 하지만 (그림 4)의 (a)에서는 중점단어와 그 외의 단어가 상당히 구분이 되는 것을 볼 수 있다. 즉 기본주파수의 평균과 분산은 서로 어느 정도의 연관성을 가지며 상호 보완적인 역할을 한다고 볼 수 있다. 이는 (그림 4)의 (a)를 보면 알 수 있는데, x축 값이 -50에서 50사이에서 중점단어와 그 외의 단어가 겹쳐지는데, 이 부분에서 y축의 0보다 큰 값들은 주로 중점단어로부터 측정된 것이었고, 0근처나 0보다 작은 쪽은 그 외의 단어들로부터 측정된 것들이었다. 다시 말해 중점단어는 그 외의 단어들에 비해 기본주파수의 평균값이 상대적으로 높았지만, 중점단어 기본주파수의 평균값이 그 외의 단어의 것과 별 차이가 없거나 오히려 낮은 경우에는 대체로 중점단어 기본주파수 분산이 그 외의 단어에 비해 높다는 것을 알 수 있었다.

지금까지의 실험 결과로 부터 유추해 볼 때, 기본주파수 평균의 상댓값과 기본주파수 분산의 상댓값을 서로 연관 지어 분석하면 중점단어와 그 외의 단어를 구분 지을 수 있다. 복수개의 중점단어가 존재할 경우와 약하게 강조되어 그 외의 단어와의 구분이 어려운 경우 등이 중점단어에 대한 구분을 어렵게 하는 원인이 된다고 여겨지는데 이는 앞으로 흥미로운 연구주제가 될 것이다.

6. 결 론

본 논문에서는 한국어 발화음성에 있어서 화자의 초점이 맞추어져 있는 문장 내의 중점단어를 구분해 내기 위해 유용하게 사용될 수 있는 운율적 특징들을 분석하고 패턴을 찾아보았다. 기본 주파수의 단어별 평균과 분산의 상댓값을 계산하여 이를 동시에 활용하는 것이 중점단어와 그 외의 단어를 구분 짓는 데에 유용한 것으로 나타났다. 이와 같은 패턴분석에 관한 연구 결과는 나아가 패턴인식을 위한 기초연구라는데 의의가 있다. 이를 기반으로 하면 통계적인 접근이나 기존의 패턴인식의 방법들을 사용해서 발화된 한국어 대화음성으로부터 직접 중점단어를 구분해 냄으로써 화자의 의도를 파악하는데 도움이 됨과 동시에 핵심어의 단서를 찾아내는데 큰 역할을 할 수 있다. 또한 자연어처리를 하여 추출된 핵심어들과 이번 논문의 결과로 얻어진 중점단어 간의 비교를 통하여, 서로 간에 어떠한 상관관계를 갖고 있는지에 대한 연구가 진행 중에 있다. 이러한 후속 연구를 통해 보다 다양한 경험적 검증을 할 수 있으며, 더 나아가 음성이해 분야에 많은 도움이 될 것 이다.

참 고 문 헌

[1] S. Ananthakrishnan and S. Narayanan, “Automatic Prosody Labeling using Acoustic, Lexical, and Syntactic Evidence,” IEEE Transactions on Speech, Audio and Language Processing, 16(1), pp.216-228, Jan., 2008.

[2] D. Baron, E. Shriberg and A. Stolcke, “Automatic punctuation and disfluency detection in multi-party meetings using prosodic and lexical cues,” In Proc. of International Conference on Spoken Language Processing (ICSLP), pp. 949-952, 2002.

[3] S.-A. Jun and H.-J. Lee, “Phonetic and phonological markers of contrastive focus in Korean,” In Proc. International Conference on Spoken Language Processing (ICSLP), pp.1295-1298, 1998.

[4] S.-A. Jun, “Intonational Phonology of Seoul Korean Revisited,” Japanese-Korean Linguistics 14 , Stanford: CSLI [Also printed in UCLA Working Papers in Phonetics, #104, pp.14-25, 2005], 2006.

[5] S.-A. Jun and H.-S. Kim, “VP Focus and Narrow Focus in Korean,” In Proc. of ICPhS, Saarbruecken, Germany, 2007.

[6] S. Kang and S. Speer, “Prosody and clause boundaries in Korean,” In Proc. of International conference on Speech Prosody, pp.419-422, 2002.

[7] E.-S. Kim and B. Scassellati, “Learning to refine behavior using prosodic feedback,” In Proc. of IEEE 6th International Conference on Development and Learning, pp.205-210, 2007.

[8] H.-S. Kim, S.-A. Jun, H.-J. Lee, and J.-B. Kim, “Argument Structure and Focus Projection in Korean,” In Proc. of International conference on Speech Prosody, Dresden, Germany, 2006.

[9] B. Secrest and G. Doddington, “An integrated pitch tracking algorithm for speech systems,” In Proc. of International Conference on Acoustics, Speech, and Signal Processing, pp.1352-1355, Apr., 1983.

[10] K. Sönmez, E. Shriberg, L. Heck, and M. Weintraub, “Modeling Dynamic Prosodic Variation for Speaker Verifi-

cation," In Proc. of International Conference on Spoken Language Processing, Sydney, Australia, Vol.7, pp.3189-3192, 1998.

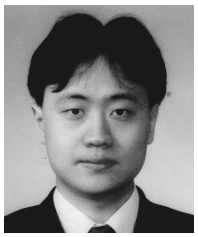
[11] Speech Filing System [Online]. Available: http://www.phon.ucl.ac.uk/resource/sfs

[12] F. Tamburini, "Automatic prosodic prominence detection in speech using acoustic features: an unsupervised system," In Proc. of Eurospeech, pp.129-132, 2003.

[13] D. Wang and S. Narayanan, "A multi-pass linear fold algorithm for sentence boundary detection using prosodic cues," In Proc. of International Conference on Acoustics, Speech, and Signal Processing, pp.525-528, May, 2004.

[14] D. Wang and S. Narayanan, "An Acoustic Measure For Word Prominence In Spontaneous Speech," IEEE Transactions on Speech, Audio and Language Processing, 15(2), pp.690-701, Feb., 2007.

[15] 구희산, "영어와 한국어 낱말 운율의 음성학적 연구", 응용언어학, 제8호, pp.123-140, 1995년 2월.



권 순 일

e-mail : soonil@kist.re.kr

1998년 연세대학교 전자공학과(학사)

2000년 미국 University of Southern California 전기공학과(공학석사)

2005년 미국 University of Southern California 전기공학과(공학박사)

2005년~2006년 삼성전자 정보통신총괄 통신연구소 책임연구원

2006년~현 재 한국과학기술연구원(KIST) 지능인터랙션연구센터 선임연구원

관심분야 : 음성인식, 화자인식, 음성합성, 음성/오디오 신호처리, HCI, HRI 등

박 지 형

e-mail : jhpark@kist.re.kr

1979 서울대학교 기계설계학과(공학사)

1981 서울대학교 기계설계학과(공학석사)

1993 서울대학교 기계설계학과(공학박사)

1981~현 재 한국과학기술연구원 지능인터랙션연구센터 센터장

2004~현 재 과학기술연합대학원대학교 HCI 및 로봇응용공학 교수

관심분야 : Interactive Tabletop Computing, Cognitive Human Robot Interaction, Reality Mining

박 능 수

e-mail : neungsoo@konkuk.ac.kr

1991년 연세대학교 전기공학과(학사)

1993년 연세대학교 대학원 전기공학과(석사)

2002년 미국 University of Southern California, 전기공학과(컴퓨터공학)(공학박사)

2002년~2003년 삼성전자 책임연구원

2003년~2007년 건국대학교 정보통신대학 컴퓨터공학부 조교수

2007년~현 재 건국대학교 정보통신대학 컴퓨터공학부 부교수

관심분야 : 컴퓨터구조, 임베디드 시스템, 고성능 병렬시스템, 멀티미디어 컴퓨팅, 컴퓨터보안 등