

하이브리드 자동 통역지원 시스템에 관한 연구

*임종규 · · · ***강봉균

*한국과학재단 책임관리연구원 · '국군의무사령부 실장
***인천대학교 산업공학과 겸임 · ***서울대학교 생명과학부 교수

A Study of Hybrid Automatic Interpret Support System

*Chong-Gyu Lim · · · ***Bong-Kyun Kang

*Korea Science and Engineering Foundation · 'Armed Forced Medical Command
***Industrial Engineering of Incheon University · ***Biological Sciences of Seoul National University

The previous research has been mainly focused on individual technology of voice recognition, voice synthesis, translation, and bone transmission technical. Recently, commercial models have been produced using aforementioned technologies. In this research, a new automated translation support system concept has been proposed by combining established technology of bone transmission and wireless system.

The proposed system has following three major components. First, the hybrid system consist of headset, bone transmission and other technologies will recognize user's voice. Second, computer recognized voice (using small server attached to the user) of the user will be converted into digital signal. Then it will be translated into other user's language by translation algorithm. Third, the translated language will be wirelessly transmitted to the other party. The transmitted signal will be converted into voice in the other party's computer using the hybrid system. This hybrid system will transmit the clear message regardless of the noise level in the environment or user's hearing ability. By using the network technology, communication between users can also be clearly transmitted despite the distance.

Keywords : bone conduction technical, interpreter, voice composition, speech recognition

1. 서 론

1.1 연구의 배경 및 문제의 인식

컴퓨터를 이용하여 자동 통역 분야는 음성인식, 인식된 언어의 자연어 처리, 번역 시스템, 번역된 문장을 상대방 언어로의 음성합성기술 등이 총체적으로 결합된 복합 첨단기술이다. 이러한 인간의 언어를 기계(컴퓨터)가 이해하고 생성할 수 있도록 하기 위한 연구로서 최근 인공지능기술의 발달에 따른 자연어 처리 분야가 부각되고 있으나, 통역 시스템이기에 부족함이 많다.

컴퓨터를 이용한 통역을 하기 위해서는 음성인식기술

이 필수적일 뿐만 아니라 인식된 음성이 실시간으로 통역(기계번역)이 이루어져 상대방에게 전달된 후 상대측에서 음성으로 변환(음성합성)되는 것이 필수적이라 할 수 있으나, 현재의 음성인식기술은 70% 수준으로 아직은 자동 통역을 할수 있는 기반기술이 미숙하다[4][5][25].

특히, 번역 시스템에 있어서는 각국 언어간의 문법체계가 다르고, 1:1로 대칭되는 단어나 조사 등이 없으며, 각국 언어간의 환경과 사고방식 등이 달라 번역 및 통역에 애로가 되고 있는 실정이다.

본 연구에서는 전 세계의 공통 언어로써 영어가 주류를 이루고 있다는 점에 착안하여, 한글과 문맥구조가 전혀 다른 영어 습득에 많은 시간과 비용을 투자하더라도

본인들이 원하는 수준까지 올라가기가 매우 어려운 것이 현실이라는 문제의식을 갖고, 지금까지 개별적으로 연구개발 되어 온 음성인식, 기계어 번역, 음성전달, 음성합성 및 골 전도 시스템 등의 기술지식을 통합한 새로운 개념의 통역지원 시스템을 제안하기 위한 연구를 진행시켜 나가고자 한다.

현재 자동 기계번역은 연구개발 수준이 상당히 발전되고 있으며, 음성합성은 자동차 GPS 안내 시스템 등 다양한 분야에서 응용되고 있는 실정이고, 음성인식기술은 다소간 열악한 수준에 있으며, 골 전도 연구 분야는 시중에 골 전도 듣기용 헤드셋이 판매되고 있고, 골 전도를 통하여 말하는 기술이 특허 출원되어 있다[15].

그러나 양자간 혹은 다자간 한 장소에서 또는 원격지에서 자동 통역을 할 수 있도록 하기 위해서는 화자의 헤드폰에 부착된 마이크 및 골 전도를 통하여 2중화된 음성을 획득하고, 대화자의 몸에 부착된 강력한 초소형화된 컴퓨터를 이용하여 기계 번역되어, 그 결과가 상대방 컴퓨터에게 무선으로 전달되어 그 곳에서 음성으로 합성되어 대화자의 귀(이어폰 혹은 골 전도 시스템)로 전달되는 통합 시스템이 필요하다.

본 연구의 목적은 지금까지 알려진 개별적인 기술들을 통합하여 골 전도 기술과 무선기술을 첨가한 새로운 개념의 통역지원 시스템을 제안하는 데에 있다.

1.2 연구의 범위 및 기존 연구와의 차별성

지금까지 개별적으로 연구되어져 온 음성인식기술, 음성 합성기술, 번역기술, 무선전달기술, 골 전도 시스템 등을 새로운 통역지원 시스템이라는 하나의 통합목표에 연결시켜 연동하는 방법까지로 한다.

기존의 방식들은 각각의 개별적인 기술개발을 위한 연구에 주안점을 두어 왔으며, C-Star 컨소시엄 (일본 ATR 연구소, 미국 카아네기 멜론대학, 이탈리아, 독일, 프랑스, 한국 ETRI 등 참여)에서는 대화체간 휴대폰을 이용하여 다국어간 자동 통역 노력에 주안점을 두어 왔다[26][27].

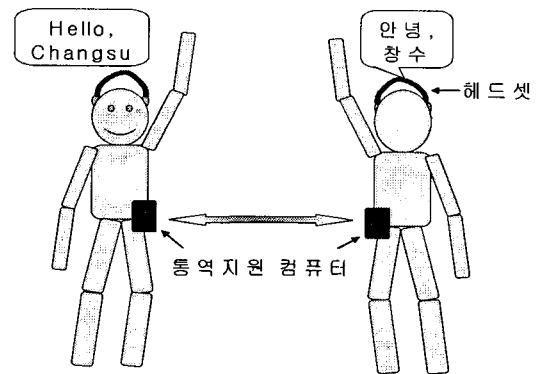
그러나 본 연구에서는 골 전도 시스템과 상대방에게 무선을 통한 전달이라는 새로운 아이디어를 접목시킴으로써 중간의 매질인 공기 중에서 음질이 약화되는 일이 없이 원격지의 대화도 통역된다는 장점을 가진다.

부가적으로는 음성인식에 있어서 골 전도 및 입 마이크 등 하이브리드(hybrid) 시스템을 통하여 음성을 인식함으로써 시끄러운 환경에서도 음성의 인식과 전달이 가능해져 노인 또는 청각 장애인에게 특히 많은 도움을 줄 수 있다는 장점이 있다.

2. 자동통역지원시스템의 통합메커니즘

2.1 하이브리드 시스템

<그림 1>은 상대방과의 대화를 하기위한 과정을 통역지원 컴퓨터를 통해 서로간의 헤드셋으로 음성을 송수신 하는 과정을 나타낸 것이다.



<그림 1> 통역지원 컴퓨터를 통한 송수신 경우

이를 구체적으로 살펴보면 화자(speaker)가 대화하고자 하는 언어를 폐의 호흡을 동력원으로 하여 성대를 통해 연구개 및 경구개를 경유한 호흡이 구강 내에서 혀의 위치와 치아 및 입술의 접촉되는 형태에 따라 언어가 생성되며 이 생성된 언어는 외부의 공기 중으로 발산된다.

이 언어는 자신의 귀인 외이도로 전달되어 골 전도 시스템에 의해 구사된 언어를 인식하게 되며, 인식된 신호는 아날로그 신호로 자기 자신의 몸에 부착된 컴퓨터에 전달되게 된다.

전달된 아날로그 신호는 음성인식기술에 의해 디지털 신호인 0과 1로서 재구성되며 변경된 신호는 알파벳으로 변환시킨 후 단어로 조립되며, 단어의 품사를 결정하고 동사구·!사구의 구조를 판정하여 주어와 술어의 관계를 추출한다. 주어진 문장을 문법규칙에 맞추어 구문 해석을 한다[17][18][19][20][21][22][23].

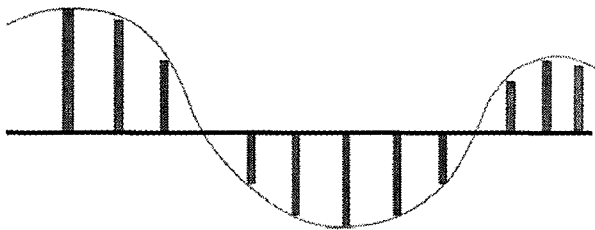
다음 단계는 구문해석으로 분석된 글의 구조를 분석하여, 뜻이 타당한지의 여부를 여기서 판단한다. 마지막으로 문맥해석을 하고, 이를 다시 해당국 언어로 번역한다. 이 때에 두개의 언어 사이에서는 각 단어끼리의 사전에 의하여 단어들어 지원된다. 지원된 글의 의미가 타당한지를 검사한 후 그 결과를 디지털신호로 재전환하여 상대방 컴퓨터에 무선으로 전달되면 그곳에서 음성합성기술에 의해 음성으로 변환되어 대화자의 귀로 전달되는 아이디어다[10][11][24].

음성인식(음성획득)시에도 일반적인 마이크를 설치하

여 음성인식을 하거나, 공기를 통하지 않고 골 전도 시스템을 통한 인식으로 2중화함으로써, 평상시 시끄러운 환경 하에서도 골 전도 인식으로 어느 환경에서나 음성 인식과 전달이 가능한 하이브리드 시스템을 제안한 것이다.

2.2 아날로그 신호의 디지털 변환

인간의 음성은 <그림 2>와 같이 연속적이고 이산적인 아날로그 신호이다. 컴퓨터에서 이를 이용하기 위해서는 컴퓨터가 이해할 수 있는 디지털 값으로 변환하여야 한다.



<그림 2> 아날로그 신호의 주기적인 샘플링 예

위 <그림 2>에서 파형의 일정한 시간마다 일정한 간격에서 수직적인 값(높이)을 취하여 디지털 값으로 변환하는 샘플링을 실시한다. 샘플링 속도는 (식 1)과 같은 나이퀴스트(Nyquist) 기준을 통해 계산된다[12].

$$F_s > 2 \times BW \dots\dots\dots (1)$$

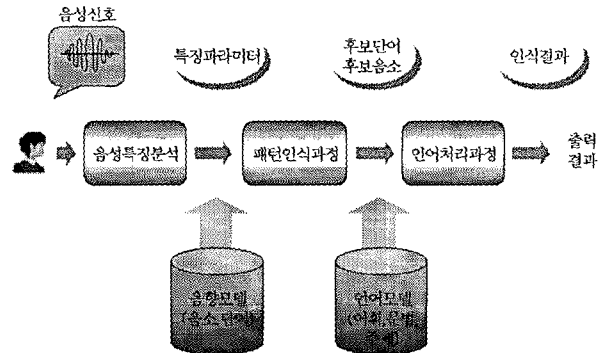
F_s = sampling frequency
 BW = bandwidth of input analog voice signal

나이퀴스트 기준은 샘플링 되는 최대 주기를 나타내는 수의 두배 이상으로 샘플링을 취하면 음성을 복원할 수 있다. 거의 모든 아날로그 음성신호가 4khz의 대역폭 범위 내에 들어가기 때문에, 샘플링 속도 8khz은 대부분의 음성통신에 대해 충분하다고 간주한다.

2.3 음성인식 및 음성합성

2.3.1 음성인식

음성인식이란 사람 음성의 음파를 분석하여 음소단위로 인식하는 기술을 말한다. 즉, 음성인식(speech recognition)은 컴퓨터가 음향학적인 음성신호(acoustic speech signal)를 텍스트로 사상(mapping) 시키는 과정이다. 이를 그림으로 나타내 보면 <그림 3>과 같다.



<그림 3> 음성인식 과정

음성인식은 일반적으로 마이크나 전화를 통하여 얻어진 음향학적 신호를 단어나 단어집합 또는 문장으로 변환하는 과정으로 이루어진다.

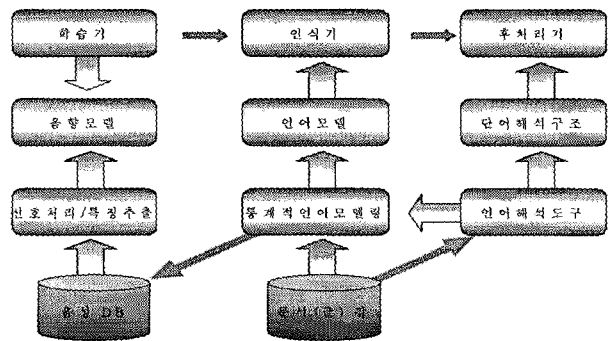
인식된 결과는 명령이나 제어, 데이터 입력, 문서 준비 등의 응용 분야에서 최종 결과로 사용될 수 있으며, 음성이해와 같은 분야에는 언어처리과정의 입력으로 사용될 수 있다.

그 결과 음성인식기술은 인간과 컴퓨터 간의 자연스러운 의사소통을 가능하게 해주며, 인간에게 보다 밀착시킴으로써 인간의 생활을 보다 풍요롭게 만드는데 필수적인 최첨단기술이다.

음성인식은 크게 3가지 방식으로 분류되는데 사용자에 따라 화자 종속방식(speaker dependent system), 화자 독립방식(speaker independent system), 화자 적응방식(speaker adaptive system) 이 있다.

또한 발음방식에 따라 고립단어인식, 핵심어인식, 연결단어인식, 연속음성인식, 대화음성인식으로 나눌 수 있다. 인식대상 어휘의 수에 따라서 소용량 인식기술, 중용량 인식기술, 대용량 인식기술로 나눌 수 있다.

인식단위에 따라서 단어단위로 인식하는 단어 인식기술, 음소단위로 인식하는 가변어 인식기술로 나눌 수 있다.



<그림 4> 음성인식 시스템 구성도

음성인식기술은 음성명령 컴퓨터는 물론 무인 전화번호 안내, 음성명령 주문형 비디오, 각종 음성안내 시스템, 가전제품을 비롯해 자동차 항법장치(car navigation system), 홈오토메이션, 음성다이얼링 등 이용영역이 무한하다.

<그림 4>는 음성인식 시스템 구성도를 제시한 것이다. 음성인식을 위해서는 음성인식(speech recognition), 해석(analysis), 이해(understanding) 등 많은 단계에 대한 시스템적 접근이 필요하다.

전형적으로 음성인식은 음성의 디지털 샘플링에서 시작되며, 이 과정은 <그림 3>에서 설명한 바와 같다.

다음 단계는 음향 신호처리(acoustic signal processing)이다. 대개의 기술은 스펙트럼 분석(spectral analysis)기술을 포함하고 있다. 예를 들면 선형 예측 부호화 기술인 LPC(linear predictive coding), 음성인식시스템 기술인 MFCC(mel frequency cepstral coefficients), 코클리어 모델링 기술인 CM(cochlea modeling)[6] 등이다.

다음 단계는 음소의 인식(recognition of phonemes)으로서 대부분의 시스템들은 인식과정에 부가하여 어느 정도의 언어에 대한 지식(knowledge of the language)을 이용한다. 몇몇 시스템은 음성을 '이해' 해서 그들이 인식한 단어를 화자가 말하려는 의미로 표현되도록 변환시킨다.

여기에는 음소들의 그룹과 단어들이 포함된다. 이 단계에서 사용되는 기술들에는 동적 계획법 매칭법인 DTW (dynamic time-warping method), 음성인식기술 알고리즘인 HMM (hidden markov model), 신경회로망(neural network) 복합기술, 전문가 시스템(expert system) 등이다.

HMM은 이중 통계적 모델로서, 기본이 되는 음소열의 생성과 프레임 단위의 표면적 음향학적인 표현을 확률로서 제시한다. 프레임 단위의 점수를 예측 하는데 있어 신경망이 사용되기도 하며, HMM 시스템과 결합되어 혼합 모델로서 사용되기도 한다.

단어에서의 유사성은 인식 시스템의 성능에 직접적인 영향을 미친다. 즉, 음향학적 신호와 관련된 많은 변이로 인하여 음성인식이 어렵게 된다. 음성인식의 문제점을 야기하는 요소들은 많지만, 각각의 단어가 구성되는 최소 단위인 음소의 음향학적인 표현은 문맥에 종속된다[1][6][7][14].

2.3.2 음성합성

음성합성이란 기계적인 장치나 전자회로 또는 컴퓨터 모의를 이용하여 자동으로 음성의 파형을 생성해 내는 것으로써, 음성합성에 대한 연구는 다른 음성에 관련된 기술들 보다 가장 먼저 연구된 기술이다.

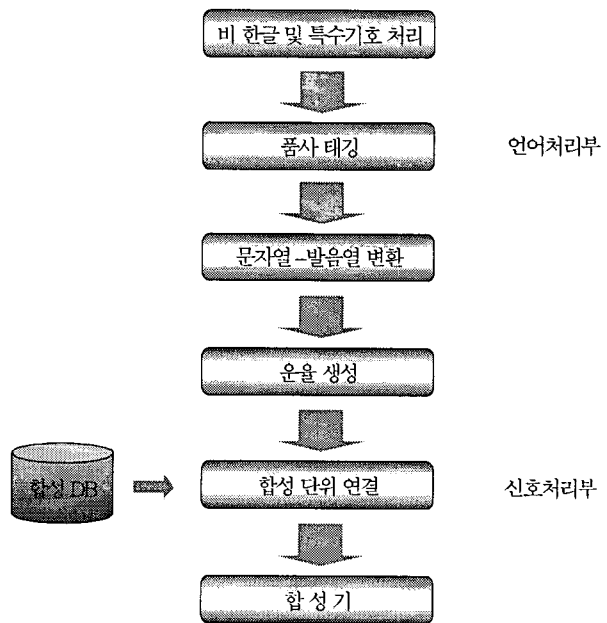
초기의 음성합성에 대한 연구는 대부분 기계적 또는

전자회로를 이용하여 인간의 발성기관을 모의하는 것이었다. 인간의 발성기관을 모델링 하는 것은 아직까지도 음성합성 연구에 궁극적인 목표로 남아 있지만, 컴퓨터의 연산속도 및 기억용량이 급속히 발전하면서 음성합성에 대한 연구는 단순히 인간의 발성기관 모델링에 그치지 않고 문서처리기술을 포함한 문서·음성 변환기술로 확장되었다.

음성합성(speech synthesis) 프로그램은 문서를 입력(written input) 하여 자동적으로 생성되는 합성음성으로 변환하여 출력(spoken output) 하는 것을 말한다. 그래서 음성합성을 텍스트 음성 변환 즉, TTS (text-to-speech conversion) 라고 불리어 지기도 한다.

이러한 음성합성기술은 컴퓨터가 인식할 수 있는 문자를 사람이 이해할 수 있는 자연어로 바꾸어 주는 기술로서 컴퓨터가 이해하는 텍스트(text)를 사람의 목소리(남성, 여성, 노인, 어린이 등)로 표현하는 문자·음성 변환 기술이다.

TTS 기술은 대화속도 처리, 글자간 간격 처리, 높낮이 처리, 단어사전처리, 단어구술처리, 문자구술 처리 등의 기술들을 기반으로 하고 있다. 음성합성기술은 언어처리부와 신호처리부로 나눌 수 있다. 언어처리부는 비 한글 및 특수기호를 처리하는 부분, 품사 태깅(tagging) 부분, 문자열·발음열 변환 부분으로 나누어진다. 신호처리부는 운율생성 부분, 합성 데이터베이스를 통한 합성단위 연결 부분, 합성기 부분으로 나누어진다. 지금까지 설명한 음성합성의 과정을 <그림 5>와 같다[8].



<그림 5> 음성합성 과정도

2.3.3 음성이해

자동 음성이해(speech understanding)는 컴퓨터가 음향 음성신호(acoustic speech signal)를 인식한 다음에 이를 음성으로부터의 추상개념을 표현한 의미(abstract meaning)로 사상(mapping)시키는 과정이다.

이것은 한마디 한마디를 정확히 인식하는 것만이 아니라, 발음되는 문장 전체의 의미를 이해하려는 시도를 음성이해라 부른다. 이를 위해서는 인식된 음성을 분석하여 의미가 파악되어지고, 필요로 하는 형태로 전환하여야 하는데 이 기술이 자연어 처리기술이다.

음성이해는 기본적으로는 ‘음성인식+자연언어 이해’인데, 음성인식기술이 아직 미숙하기 때문에 이 이해(understanding)는 단순한 자연언어의 이해에 비해 수준이 떨어진다. 음성이해는 보통 다음과 같은 순서로 행하여진다.

- ① 입력된 음성신호를 일정 길이의 단 구간으로 구분한다.
- ② 각각의 단 구간을 음향 분석하여 특정 파라미터(주파수 스펙트럼과 음성파워 등)를 추출한다.
- ③ 음소단위로 분할(segmentation)하고 가 분할단위(segment)의 음소인식을 행한다. 여기에서 음소개별이 만들어진다.
- ④ 단어사전과 비교하여 단어를 예측한다.
- ⑤ 구문해석을 하여 예측단어가 정확한가를 검증한다.
- ⑥ 의미적으로 타당한가를 조사한다. 만약 정확하지 않으면 구문해석을 다시 실시한다.

2.3.4 기계번역

최근 인공지능기술의 발달로 자연어 처리 분야가 발전되고 있는데 이는 인간의 언어를 기계(컴퓨터)가 이해하고 생성할 수 있도록 하기 위한 연구로서 컴퓨터를 사용하여 한국어나 영어 등의 자연언어를 다른 자연언어로 자동적으로 번역하는 것을 말하며, 이때 자연언어로의 번역은 단지 단어 수준에서 외국어로 바꾸어 놓는 것만으로는 어떤 의미인지 알 수가 없기 때문에 글의 뜻을 문맥에서 이해하여 그 나라 언어의 독특한 표현으로 바꾸어야 뜻이 통하는 문장으로 번역될 수 있다.

기계번역은 분석-변환-합성과 같은 과정을 거쳐 이루어지는데, 이 과정을 보면 다음과 같다.

- ① : 입력언어의 구조를 분석하여, 중간적 구분표현을 만들어 낸다.
- ② : 중간표현에 번역을 할당하거나 의미론에서 추출된 표현을 목표로 언어의 해당표현으로 바꾸어 놓는다.
- ③ : 변환된 중간표현을 목표언어의 구문에 맞게 정렬시키거나, 필요한 어미변화를 부여한다.

여기서 번역되는 국어의 문장을 입력언어, 번역할 상대국어의 문장을 목표언어라 한다. 분석은 다음과 같은 순서로 진행된다.

- ① 형태해석 : -장을 단어로 분해한다.
- ② 구문해석 : |진 단계에서 분해된 단어의 품사를 결정하고 동사구 및 명사구의 구조를 판정하여 주어와 술어관계를 추출한다. 주어인 문장을 문법규칙에 맞추어 구조를 길정하는 작업이다.
- ③ 의미해석 : -장 즉, 기호열(symbol string)에 표현된 단어가 의미하는 개념과 문장이 표현하고자 하는 ‘개념들 간의 관계’를 규명하는 것이다. 즉, 글이 지닌 뜻을 컴퓨터로 어떻게 표현하느냐는 것이 여기서의 과제이다. 구문해석으로 분석된 글의 구조를 분석하여, 뜻이 타당한지의 여부를 여기서 판단한다. 의미해석의 어려움은 크게 3가지가 있다. 첫 번째는 인간의 정신활동에 대한 표현의 한계에서 오는 어려움이며, 두 번째는 계산 모델(computational model) 구현의 어려움이고, 세 번째는 현실 사회의 지식(real world knowledge) 구축의 어려움 등이 제시되고 있다.
- ④ 문맥해석 또는 어용론해석 : |어진 글의 문맥에서 글의 뜻에 맞는 다의어의 해석, 생략어의 보조, 대명사가 무엇을 가리키느냐 하는 것과 글의 주제나 의도 등을 추정한다.

3. 골 전도 시스템에 의한 음성인식

3.1 음의 생성원리

통역은 대화자 상호간에 대화를 통하여 주고받는 형태이므로 우선 말이 만들어지는 원리에 대하여 살펴볼 필요가 있다. 상대방에게 어떤 말을 하고자 할 때는 두뇌에서 사고하여 입을 통하여 대화를 하게 된다.

기관으로는 두뇌, 신경계, 폐, 기관, 후두(성대 포함), 인두(목구멍), 코, 턱, 그리고 입(연구개, 정구개, 이, 혀, 입술 등을 포함)이 있다. 이 기관들은 모두 폐에서 입술까지 연장되어 있는 복잡한 모양의 ‘관’을 형성한다.

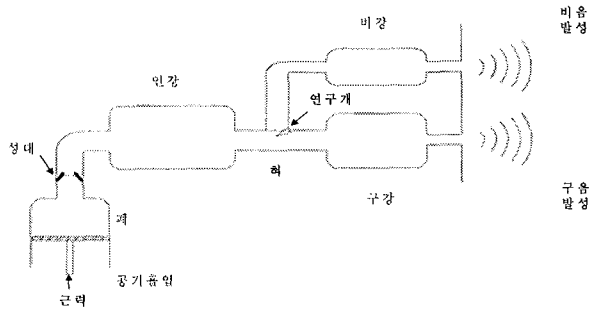
언어사고시 관련되는 두뇌의 오른쪽 피질은 음악적 능력, 감정적 반응, 시각적 형태를 인식하는데 중요하다고 알려져 있으며, 언어적 능력은 주로 뇌의 왼편에 주로 집중되어 있다.

소리가 만들어지기 위해서는 혀와 입술의 고도로 숙련된 움직임을 필요로 하며, 이는 복잡한 활동으로서 두뇌에서 통제한다. 소리는 폐로 들어 선 숨을 내릴 때 발성에 대한 에너지원이 되어 성대가 진동하면서 소리가

만들어 진다.

성대의 진동 속도를 증가시키므로써 주파수는 점점 더 높아지고, 음향적 특성은 성도의 모양에 따른다. 말하는 동안 우리는 계속해서 혀와 입술, 다른 조음기관 등을 움직여서 성도의 모양을 바꾼다.

이러한 움직임들은 성도의 음향적 특성을 바꿈으로써 다른 음성을 생성할 수 있게 한다. 다른 음성을 생성하기 위해 성도의 모양을 조정하는 과정을 조음이라고 부른다.



<그림 6> 음성생성 모형도

<그림 6>은 음성생성 모형도를 제시한 것이다. 폐에서 나온 기류는 음성과 생성을 위한 에너지를 제공해 주고, 성대는 소리로 바꾸고, 성도의 모양을 변경하여, 조음기관들은 이 단조로운 음을 구별할 수 있는 음성으로 변환시킨다. 이렇게 생성된 소리는 혀, 입술, 인후 등의 구강 내의 구조에 따라 우리가 인식할 수 있는 다양한 음색의 목소리를 형성한다. 자연어(모국어) 가운데 일부 언어는 숨을 들이키면서 발성하는 소리를 이용한다. 통상 한국어, 영어, 일본어 등 각국의 언어들은 보통 숨을 내쉬면서 발성된다[9].

3.2 골 전도 방식 및 효과

3.2.1 골 전도 방식

지하철이 흠에 들어올 때에는 아주 시끄러운 환경인 상태가 되며, 이 때에는 휴대폰으로 통화를 하고자 하면 전혀 들리지 않는다. 즉, 높은 레벨의 소음이 존재하는 공간에서는 기존의 마이크 시스템과 이어폰으로는 뚜렷한 송·신이 부족 어려워진다.

사람이 자연의 소리를 들을 수 있는 방법은 크게 두 가지로 공기전도(air conduction) 방식과 골 전도(bone conduction) 방식이 있다.

첫째, 공기전도란 소리가 고막을 통해 내이(inner ear)로 전달되는 방식이다.

둘째, 골 전도란 두개골(cranial bone)을 통한 소리가

달팽이관에 전달된 후 청각신경을 거쳐 뇌로 전달되는 것을 말한다.

사람의 귀 구조는 소리를 전달하는 전음계와 소리를 지각하는 감음계로 나누어 볼 수 있다. 전음계는 소리를 약 40db 정도(약 100배) 증폭시키는 역할을 하기 때문에 전음계에 장애가 있으면 기도청각으로는 소리를 들을 수 없는데, 이러한 문제를 해결한 것이 골도 청각기술이다. 즉, 전음계 장애인의 경우에 골도 청각을 이용하면 귓바퀴 주위의 뼈를 진동시켜 감음계 내이에 소리를 전달해 주기 때문에 소리를 들을 수 있는 것이다.

이러한 원리를 이용한 것이 골 전도 진동자(bone conduction transducer)이다. 진동자란 전기적인 신호를 진동의 신호로 바꾸는 변환기를 말하며, 골 전도 원리를 이용해 최상의 소리를 들을 수 있는 위치는 <그림 7>과 같이 귓바퀴 주위이며, 이에 따라 고안된 제품이 골 전도 헤드셋 헤드폰이다.



<그림 7> 골 전도시 잘 들리는 위치와 골 전도 헤드폰

3.2.2 골 전도의 효과

기존의 공기전도 헤드셋은 반드시 귀에 정확히 부착해야만 재 기능을 발휘할 수 있다. 따라서 장시간 사용 시에는 청각이 크게 손상된다.

반면 골 전도 헤드셋은 두개골의 진동을 통해 소리를 들기 때문에 장시간 사용에도 청각에 무리가 전혀 없으며, 헤드셋이 귀를 덮지 않아 착용감이 편하며, 헤드폰이 귀를 덮지 않아 헤드셋을 통한 소리와 외부의 소리를 동시에 들을 수 있다. 따라서 주위의 위협요소로부터 신속하게 대처할 수 있다는 장점이 있다.

소음의 세기를 데시벨(db)로 나타내며, 골 전도 사용시 90db 이상의 소음환경에서도 명료한 송·신이 가능하다. 참고로 속삭임 20db, 회화 60db, 교통량이 많은 장소 80db, 주행 중인 지하철 100db, 귀에 아픔을 느끼는 소리 140db, 비행중인 제트기 160db이 된다.

3.3 골 전도 음성인식

시끄러운 환경에서 마이크를 통해 입에서 나는 소리

를 입력하는 경우 소음까지 같이 유입되어 상대방에게 어떤 내용인지 의사전달이 어렵게 된다. 그러나 콜 전도 시스템을 활용 시 극도의 소음환경에서도 송수신이 가능하게 된다. 이는 주변에 존재하는 소음이 마이크를 통해 입력되어 송신자의 음성신호와 함께 전달되는 것을 최소화시켜 음성신호 대 주변소음의 비율을 증대시켜 줄 수 있는 양 방향 통화시스템이다.

통상 헤드셋을 끼고서 양방향 통화를 하는 경우에 헤드셋의 양쪽 귀마개 내부에 스피커가 설치되어 있어서 양쪽 귀로는 수신된 신호를 듣고, 마이크는 사용자의 입 앞에 설치되어 있어서 입으로부터 방출되는 음성신호를 포착하도록 되어 있다. 그러나 주변에 높은 소음이 존재하는 장소 또는 환경 하에서는 헤드셋을 사용하고 통화시 입 바로 앞에 설치된 마이크에 음성신호와 함께 주변 소음이 입력되기 때문에 음성신호와 함께 주변 소음이 같이 증폭되어 상대방에게 전달되므로 상대방의 스피커에는 음성신호와 소음이 동시에 방출되는 단점이 있었다. 따라서 소음이 심각한 환경에서 양방향 통화를 하는 경우에는 상호간에 어떠한 이야기를 하는지 명확히 알기 어려우며, 계속적으로 재확인을 해야 하는 불편이 따랐다.

사람이 말을 하면 성대가 진동을 하며, 성대의 진동은 입과 귀를 연결하는 유스타키오관을 통하여 자신의 고막을 진동시킨다. 이렇게 자신의 고막이 진동하게 되면 자신이 현재 어떠한 말을 하고 있는지를 스스로 확인하게 되는데, 이때 진동되는 고막은 자신이 말하는 음성신호를 외이도를 통하여 밖으로 방출하게 된다.

물론 이때 외이도를 통해 방출되는 음성신호의 크기는 그다지 크지 않기 때문에 평소에 인식하기는 힘들지만, 포착된 신호의 적절한 증폭이 이루어질 경우 충분히 청취할 수 있다.

특히 고막은 유스타키오관의 한쪽 끝에 붙어 있는 매우 얇은 피부로 된 막으로서, 질량이 작으면서 비교적 넓은 면적을 갖고 있기 때문에 중이 내의 공기 진동에너지로 진동을 유발하기에 충분하며, 유발된 고막의 진동은 관(pipe) 형의 외이도를 통해 손실 없이 방출된다.

이와 같이 외이도를 통해 밖으로 방출되는 고막의 진동을 포착하여 음성신호를 송출하게 된다.

4. 자동 통역지원 시스템의 유용성 및 발전방향

4.1 하이브리드를 활용한 자동 통역지원 시스템의 유용성

현재 한국인들은 외국어 습득에 무척 애로를 겪고 있

으며, 특히, 글로벌화에서도 영어 사용에 불편을 느껴 정신적 경제적으로 많은 피해를 입고 있을 뿐만 아니라 국가적 피해 또한 막대하다고 해도 과언이 아닐 것이다.

앞에서 살펴본 바와 같이 본 연구에서 제안하는 자동 통역지원 시스템은 기존의 기술을 종합한 새로운 형태의 통역지원 시스템으로서, 각각의 기술 발전을 고려해 볼 때 현재의 기술수준은 어느 정도의 통역이 가능한 기술 수준이라고 할 수 있다. 특히 대화자간에 대화라는 데이터를 아날로그로부터 디지털로 변환 하고, 이를 다시 번역하여 상대방에게 무선을 통한 전달이라고 하는 새로운 아이디어를 접목함으로써, 중간의 매질인 공기 중에서 음성이 약화되는 일이 없이 원격지의 통화에서도 통역된다는 장점을 가진다.

또한 부가적으로 음성인식 시에도 콜 전도와 입 마이크 등 하이브리드 시스템을 통해 인식함으로써 아주 시끄러운 환경 하에서도 음성인식이 가능하며 음성전달 또한 가능해져 노인이나 청각 장애인에게도 많은 도움이 될 것으로 본다.

본 연구결과가 현실화 될 때에는 많은 세계인들은 어려운 외국어 습득 없이 컴퓨터의 도움으로 인간간의 관계가 더욱 돈독해져 글로벌화가 앞당겨 진척될 수 있을 것으로 본다.

4.2 자동 통역지원 시스템의 과제와 향후 발전방향

자동 통역기술이란 언어장벽을 허물어 서로 다른 언어를 사용하는 사람들 간에 대화가 가능하게 하는 기술이다. 이는 통역자 없이 컴퓨터를 활용한 자동 통역 시스템으로서 이를 위한 현재의 과제와 향후의 발전방향을 종합적으로 살펴보면 다음과 같다.

첫째는 콜 전도 시스템 분야로서, 콜 전도 시스템은 현재 듣는 용도로서 콜 전도 헤드셋이 개발되어 있으며, 콜 전도 마이크 형태는 이어폰 형태로 일본에서 개발되어 현재 시판되고 있으며, 특정 용도의 콜 전도 마이크는 특허로 등록되어져 시험개발 중에 있다.

둘째는 음성인식 분야로서, 현재의 음성인식기술 수준은 음성인식 성능에 좌우된다. 음성인식 성능은 인식대상 어휘의 수, 언어모델의 복잡도, 배경 잡음 및 채널왜곡, 발화자유도(낭독음성과 대화음성의 차이), 음성신호의 샘플링 주파수, 그리고 훈련에 사용된 음성 데이터베이스의 규모 등 많은 요인에 의해 영향을 받기 때문에 현재의 음성인식기술 수준이 어느 정도인지를 간단한 수치로 표현하는 것은 다소 어렵다.

현재 발표된 논문에 따르면 연속 음성인식기술은 95% 수준 이하이며, 2천개의 단어 내지 3천개의 단어로 이루어진 대화체인 경우 70% 수준 정도라고 발표되어지고

있다. 그러나 음성인식기술의 알고리즘 개선, 인간공학 기술 발전으로 인해 더욱 더 개선되어질 전망이다.

셋째는 자연어처리(번역) 분야로서, 음성인식기술에 의해 대화자의 음성이 인식되면 해당국 언어 텍스트로 치환되어지고 다시 1대1의 사상(mapping) 시스템을 통하여 문법과 단어사전에 의거 해당국 언어인 글(텍스트)로 변화되어 진다.

한·! 자연어처리 분야는 시중에 나와 있는 소프트웨어를 활용할 경우 만족도의 측면에서 현재 대략 90% 이상의 번역율을 보이고 있으며, 한·! 자연어처리 분야는 현재 대략 70% 정도로 아직 초보적 수준이다. 다국어간 통역 측면에서는 중간언어를 통한 번역을 실시하고 있으나 초보적 수준이다. 이를 위한 국제 통역 공동연구 컨소시엄인 C-Star(Consortium for Speech Translation Advance Research)에서 연구 중에 있으며, 우리나라는 한국전자통신연구원(ETRI)이 C-Star에 참여하고 있다.

넷째는 음성합성 분야로서, 1980년대 후반 음성합성기술이 출현된 이후 음성합성기술도 급속적인 발전을 거듭해 왔다. 음성합성 시스템의 성능은 합성음의 명료도와 자연성으로 표현될 수 있다. 지난 수십 년간의 음성합성 연구는 1차적으로 합성음(컴퓨터 기계음)의 명료성을 높이는 데 주력해 왔으며, 그 결과 명료성 측면에서는 사용자들이 어느 정도 큰 불만 없이도 사용할 수 있는 수준에 이르렀으며, 데이터 튜닝 기술에 의해 더욱 진보될 것으로 보인다.

합성음의 자연성도 코퍼스 기반의 합성방식이 등장함에 따라서 상당한 발전이 되었으며, 그 결과 낭독음성 형태에서는 상용서비스가 가능해져 일기예보, 전화번호 안내, 각종뉴스, 교통정보 등 다양한 형태의 서비스들이 시행되고 있다. 그러나 대화음성의 합성은 아직 자연성 측면에서 많은 개선이 필요하며, 합성음에 감정정보를 포함시키거나 특정한 화자의 음색으로 변화하는 기술은 아직 초보적인 수준이라고 하겠다[3].

5. 결 론

오늘날 급속한 항공의 발달에 따라 대부분 하루 이내에 전 세계 어느 곳이든 갈 수 있게 되었으며, 또한 인터넷의 발달로 지구 반대편의 사람들과 화상통신을 할 수 있게 됨으로써, 글로벌화의 시대에 살고 있다.

그러나 글로벌화의 선결과제 중의 하나인 전 세계의 공통 언어는 현재 영어가 주류를 이루고 있지만, 한국인들이 모국어로 사용하는 한글은 영어와 문맥구조가 관이하게 달라 영어 습득에 많은 시간과 자금을 투자하더라도 대다수의 한국인들이 원하는 기대수준까지 올라가

기가 어렵다고 하는 현실적 문제점을 안고 있다. 본 연구결과 주요 골자는 다음과 같다.

첫째로, 글 전도 시스템과 일반적인 입으로의 대화를 위한 하이브리드 시스템을 제시하였다.

둘째로, 제시된 통역지원 시스템은 서로 대화 시 공기의 매질을 이용하지 않고 무선기술을 이용함으로써 바로 앞의 상대나 원거리에서의 대화도 가능하고 또한 네트워크를 통한 원격 화상통신에서 이용할 수가 있다.

셋째로, 상대방에 무선을 통한 전달이라는 새로운 아이디어를 접목함으로써 중간의 매질인 공기 중에서 약화되는 일이 없이 원격지의 통화에서도 통역된다는 장점을 가진다.

마지막으로, 음성인식 시에 글 전도와 입 마이크 등 하이브리드 시스템을 통해 인식함으로써 아주 시끄러운 환경 하에서도 음성인식이 가능하고, 음성전달 또한 가능해져 노인이나 청각 장애인에게 많은 도움을 줄 수 있다.

참고문헌

- [1] 강동규, 한민수, “고품질 다채널 한국어 음성합성 시스템 개발동향”, 전자공학회지 제30권 7호, 2003. 7.
- [2] 길이만, “인간의 청각 시스템에 기초한 음성 전처리기의 모델링”, KAIST 뇌과학연구센터,
- [3] 김형순, “음성정보처리기술의 현황과 전망”, 전자공학회지 제30권 7호, 2003. 7.
- [4] 김형순, “음성인식기술”, search net-신기술동향
- [5] 김희린, “음성인식기술개요 및 향후과제”, 대한전자공학회지, 제28권 제5호, pp. 580-586, 2001. 5.
- [6] 박준, 이양직, 양재우, “대화체 음성언어 번역 시스템 개발”, 제15회 음성통신 및 신호처리워크샵, 한국음향학회, pp. 320-323, 1998.
- [7] 유계원, “한국어 음성인식을 위한 음운 규칙에 대한 연구”, 한글학회, pp. 41-87, vol 260, 2003. 6.
- [8] 이정철, “음성합성기술개요 및 향후과제”, 대한전자공학회지, 제29권 제12호, pp. 1491-1497, 2002. 12.
- [9] 정재호, “생체측정 신호처리 시스템에 관한 연구”, IITA, 2002. 8.
- [10] 정한민, 이근배, 최원석, 민경구, 서정연, “관계형 데이터베이스 상에서 다국어 질의응답 시스템”, 제13회 한글 및 한국정보처리학술대회 논문 발표집, pp. 530-537, 동국대학교, 2001. 10.
- [11] A.L. Gorin et al., “Automated natural spoken dialog”, IEEE Computer, pp. 51-56, 2002. 4.
- [12] H.J. Keiser, “Mathematical Logic and Computability”, McGraw-Hill Companies, 1996.

- [13] Maebara, “기호논리입문”, 일본평론사, 1990.
- [14] Ueno, “지식공학입문”, 오음사, 1993.
- [15] 특허 제0366065호[헤드셋]
- [16] [http : www-903.ibm.com/kr/software/embedded/](http://www-903.ibm.com/kr/software/embedded/)
- [17] [http : infocom.chonan.ac.kr/~limhs](http://infocom.chonan.ac.kr/~limhs)
- [18] [http :](http://www.kci.go.kr) : 기계어 정의수록
- [19] [http : www.voicetech.co.kr](http://www.voicetech.co.kr)
- [20] [http : www.aistudy.co.kr/physiology/phona-tion_denes.html](http://www.aistudy.co.kr/physiology/phona-tion_denes.html)
- [21] [http : www.aistudy.co.kr/branch/natural/natu-ral_language_processing.html](http://www.aistudy.co.kr/branch/natural/natu-ral_language_processing.html)
- [22] [http : www.ktechno.co.kr/rd_news/news08/index.html](http://www.ktechno.co.kr/rd_news/news08/index.html)
- [23] [http : idec.hanyang.ac.kr/new/speak/2000_11/sound.html](http://idec.hanyang.ac.kr/new/speak/2000_11/sound.html)
- [24] [http : www.neuronetism.com/sub09.html](http://www.neuronetism.com/sub09.html)
- [25] [http : newton.hanyang.ac.kr/~hongsjun/main.html](http://newton.hanyang.ac.kr/~hongsjun/main.html)
- [26] [http : www.c-star.org](http://www.c-star.org)
- [27] [http : voice.etri.re.kr](http://voice.etri.re.kr)