《主 題》

한국어 문자음성 변환 시스템 : 가라사대

권 철 홍* · 정 원 석* · 구 춘 모* · 김 형 순**
(*디지콤 정보통신연구소, **부산대학교 전자공학과)

■ 차 례 ■

요 약

본 논문에서는 국내 최초의 상용 한국어 무제한 음성합성 시스템인 가라사대에 관하여 기술한다. 우선, 음성합성 과정의 각 단계에 이용된 알고리즘을 설명한다. 즉, 문장의 분석을 위해서는 문장 전처리, parsing, 발음표기 변환등이 규칙에 의하여 순차적으로 수행된다. 문장 분석후에는 강세, 억양과 지속시간등의 운율을 제어하는 요소가 계산되고 음성신호는 확장된 diphone 단위의 음성신호를 연결하여 생성된다. 다음으로 가라사대의 하드웨어 및 소프트웨어의 구성에 관하여 서술한다. 범용의 디지탈 신호처리 IC를 이용하여 구현한 하드웨어와 가라사대의 소프트웨어 뿐만 아니라 PC내의 소프트웨어의 구성과 역할에 관하여 살펴본다.

Ⅰ. 서 론

임의의 문장으로부터 음성을 생성하는 무제한 음성합성 기술은 manmachine interface의 중요한 요소로 여겨져 많은 연구가 이루어져 왔다. 최근에 컴퓨터와 신호처리 기술의 발전에 힘입어 무제한 음성합성기술은 몇개의 언어에 대하여 상용의 제품이 나올 정도로 진전되었다. 우리나라에서는 대학, 국가연구소, 기업등에서 1980년 이후에 연구가 진행되어 왔으며 몇몇의 prototype이 개발되었다[1-3, 12].

본 논문에서는 국내 최초의 상용 문자음성변환 장치인 가라사대에 관하여 살펴보고자 한다. 가라사대의 하드웨어는 IBM 호환 PC의 확장 슬롯에 장착할 수 있는 카드형태로 구현되었으며, 가라사대의 소프트웨어는 가라사대 하드웨어 상에서 동작하는 문자음성변환 firmware, DOS I/O 드라이버 소프트웨어, PC 상에서 동작하는 다양한 응용 소프트웨어로 구성되어 있다. 가라사대는 임의의 문장을 받아들여 이를 합성음으로 변환하는데 이는 헤드폰이나 외부 스피커를 가라사대에 직접 접속하여 들을 수 있다.

가라사대의 주요 응용분야는 문서의 proof-reading, 전자메일의 음성변환, 컴퓨터를 이용한 교육 및 훈련, 장애인을 위한 컴퓨터등을 꼽을 수 있다. 전화선 접속기능을 가진 모드와 연결하여 사용하면 각종 정보나 메세지 등을 전화를 통하여 제공할 수도 있다. 본 논문에서는 가라사대의 전체적인 구성을 서술하고 개발중에 고려되었던 기술적인 문제에 관하여 고찰하였다.

Ⅱ. 음성합성 알고리즘

2.1 개 요

본 시스템의 구성이 그림 1에 도시되어 있다. 본 시스템은 크게 언어학적 처리부와 음성신호처리부로 구성되어 있다. 각 처리부는 몇 개의 모듈로 구성될 수 있다. 이 장에서는 각 모듈의 기능에 대해 설명한다.

2.2 언어학적 처리부

언어학적 처리부의 구성은 크게 텍스트 전처리, 발음표기 변환 알고리즘 및 문장분석과정(parsing)으로 나눌 수 있다. 텍스트 전처리 과정은 한글이 아닌 숫자, 기호, 영어, 약어등을 적절한 한글로 바꾸어 주며, 발음표기 변환 알고리즘은 입력된 한글을 사람이 발음하는 형태의 표기로 변환시킨다. 그리고 문장분석 과정은 입력문장의 구조를 파악하여 음성합성에 필요한 운율정보를 추출해 내는 기능을 갖는다.

2.2.1 텍스트 전처리

본 시스템은 입력으로 한글 뿐만 아니라 숫자, 기호, 영어, 약어등 컴퓨터에서 쓰일 수 있는 모든 문자를 받아들인다. 그런데 숫자, 기호, 영어, 약어 등 한글이 아닌 문자를 음성으로 변환하기 위해서는 한글로 바꾸어 주는 과정이 필요하다. 예를 들어 '1.22'는 '일점 이이'로 '91. 1. 22'는 '구십 일년 일월 이십이일'로 변환해야 한다. 즉 같은 숫자의 나열인 '1.22'도 문맥에 따라 다르게 변환된다.

이와 같이 숫자를 한글로 적절히 변환하기 위해서는 주어진 입력 문장의 구문 분석은 물론 의미 분석까지 해야 하는데, 그것은 매우 어려운 일이므로 여기서는 단지 주위의 상황으로 그 숫자, 기호, 영어, 약어가 내포하고 있는 의미를 추정하여 적절한 한글로 변환하도록 하였다.

본 절에서는 이와 같이 숫자, 기호, 영어, 약어 등 한글이 아닌 문자를 한글로 바꾸어 주는 text 전처리 과정에 대해 다룬다.

(1) 숫자의 처리

숫자는 다음 세 가지중 한가지 방식으로 변환된다.

1) 12 → 십 이

2) 12 → 일십

3) 12 → 열 둘

이 세 가지 변환 방식중 1 을 기본으로 하고, 몇 가지 특수한 경우에 2 또는 3 이 적용된다.

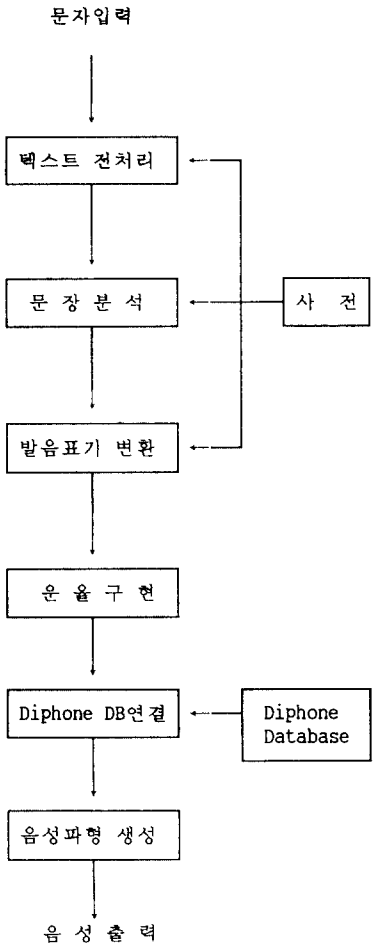


그림 1. 음성합성시스템 구성도

(2) 기호의 처리

기호에는 문장부호, 수학기호, 특수기호 등이 있는데, 같은 기호라도 그것이 차지하고 있는 위치에 따라 의미가 다르다. 예를 들면 점표(.)는 문장의 끝을 나타내는 종지부로 쓰이기도 하고, 소숫점을 표시하는데도 사용되고, 나위일을 나타내는 데도 쓰인다. 여기서는 그 기호의 주위 상황으로 의미를 추정하여 적절히 변환을 한다. 이때 문장부호는 변환을 하지 않고 다음 단계에서 억양을 제어하는데 필요한 정보로 이용된다. 그리고 수학기호나 특수기호는 적절한 한글로 변환하거나 묵음으로 처리한다.

(3) 영어의 처리

영어 문자열의 입력 문장에서 의미하는 바는 여러

가지가 있을 수 있다. 의미가 있는 영어 단어일 수도 있고, 약어일 수도 있고, 알파벳을 나타낼 수도 있고, 단위로 쓰일 수도 있다. 우선 이 네가지중에 어디에 속하는지 적절하게 판단한 후에 각각의 경우에 알맞게 처리한다.

영어 약어는 미리 사전에 수록해 놓았다가 약어라고 판단되면 사전에서 찾아 변환하며, 한 문자로 된 단어나 자음만으로 구성된 단어는 알파벳으로 변환하며, 단위를 나타내는 경우에는 이 역시 미리 구성된 사전에서 적절한 단위를 찾아 변환하며, 마지막으로 의미가 있는 영어 단어로 판단될 경우에는 적절한 한글로 변환해 준다.

영어 단어를 한글로 변환하는 경우, 사전을 이용하게 되면 데이타베이스의 크기가 방대해질 뿐만 아니라 사전에 수록되어 있지 않은 단어, 복수형, 활용형 등을 융통성있게 처리할 수 없으므로 규칙을 이용하여 변환할 수 있는 알고리즘을 사용했다. 이 알고리즘에서는 300여 가지의 규칙을 갖고 단어의 첫 글자로부터 차례로 규칙을 적용하여 영어 음소열을 변환해 준다.

2.2.2 발음표기 변환 알고리즘

발음표기 변환 알고리즘은 한국어 음성합성시스템에서 음운학적인면과 음성학적인 면의 차이를 해결해 준다. 즉 합성음의 명료성을 위하여 한국어의 음성학적 발음 특성을 토대로 발음규칙을 설정하며, 이를 이용하여 입력된 한글문장을 정확한 발음표기로 바꾸어 준다. 따라서 발음표기 변환 알고리즘에서는 입력문장이 한국어 정서법에 의한 띄어 쓰기 원칙을 준수하고 있다는 가정하에, 뚜렷하며 편안한 발음을 위하여 그리고 음소가 결합할때 발생하는 음운변동 현상을 처리한다. 이때 발음표기변환 알고리즘에 적용된 음운변동 현상은 니은 첨가, 히읗 탈락, 경음화 현상, 격음화 현상, 종성법칙, 자음접변, 연음법칙, 구개음화, 절음법칙 등이 있다.

2.2.3 문장분석 과정

보다 자연스러운 합성음을 만들기 위해서는 입력 문장의 구조와 형태를 분석할 필요가 있다. 문장분석 과정은 입력 문장의 구조와 형태에 따라 적절한 운율을 부여함으로써 보다 자연스러운 합성음을 내도록 하는데 그 목적이 있다.

이러한 운율 정보를 사용자가 약속된 형태로 미리 입력 문장에 포함시킨다면 문장분석(또는 parsing)을 생략할 수 있으나, 이는 특별한 응용분야에만 가능한 방법이며 일반적으로 입력 문장에 대한 문장분석 과정을 통해 필요한 운율 정보를 추출해야 한다. 문장분석부(parser)는 요구된 입력들을 원하는 목적에 맞게 분석하는 시스템을 말하며 보통 문법적 분석과 의미적 분석을 동시에 행한다. 이에는 입력 문장의 문법이 일률적으로 정해진 형식어 parser(formal language parser)와 입력의 문법에 제한이 없는 자연어 parser (natural language parser)의 두가지가 있다. 본 과제를 통해 개발한 음성합성 시스템을 위한 parser는 필연적으로 자연어 parser이며, 시스템의 목적상 문자의 운율 정보만을 추출해 내면 되기 때문에 의미적 분석은 생략하고 문법적, 형태적 분석만을 수행하도록 구성되었다. 여기서는 parser를 통해 입력 한글 문장을 분석하고 결과로써 구나 절의 경계나 또는 어절내의 조사의 유무 등에 관한 정보를 추출한다. 이와 같은 단순한 목적만을 수행하기 위해서도 상당히 많은 사전을 필요로 하며 또 자연어 처리의 본래의 어려움에서 기인한 여러가지 문제점으로 완벽한 parser를 구현하기는 어렵다.

따라서 본 시스템에 구현된 parser는 완벽한 parser를 목표로 하지않고 가급적 오류가 적은 확실한 운율 정보(이러한 정보는 국문법적으로 틀리는 경우도 있으나 본 음성합성 시스템의 운율 정보로써 적절하다고 판단되면 이를 받아들였다.)를 추출해 내는 것을 목표로 했을 뿐 아니라, 합성 시스템의 특성상 방대한 양의 사전을 가질 수는 없으므로 비교적 적은 사전을 토대로 꼭 필요한 정보만을 추출해 낼 수 있는 간단한 parser로 구현되었다.

다음은 parser를 통한 문장분석의 결과 예를 보여준다.

입력 : "나는 매우 아름다운 꽃을 좋아한다."

결과 : 나는 : 주격조사
매우 : 부사
아름다운 : 형용사
꽃을 : 목적격 조사가 있는데 앞에 수식하는 말이 길기 때문에 뒤에 휴지기를 넣는다.
좋아한다 : 종결형 어미가 있으므로 다음에 휴지기를 넣는다.

2.3 음성신호 처리부

음성합성 시스템에서의 음성신호 처리부는 언어학적 처리부에서 넘어온 운율 정보와 음성 데이타베이

스를 이용하여 음성 파형을 생성시키는 기능을 갖는다. 음성신호 처리부는 다시 운율 조절 과정 및 음성 파형 생성 과정으로 나누어지며, 음성파형 생성을 위해서는 음성의 기본 단위들에 대한 정보를 저장해 놓은 음성 데이타베이스를 구축해야 한다. 가변사대를 1명의 성우성의 기본 단위로는 diphone이 사용되었다. 다음 세절을 통해 이들 운율조절 과정과 음성파형 생성과정, 그리고 음성 데이타베이스 구축 과정에 대해 상세히 설명하기로 한다.

2.3.1 운율조절 과정

합성음의 성능을 평가하는 두가지 기준은 합성음의 명료성(intelligibility)과 자연스러움(naturalness)이라고 할 수 있다. 이 중, 음의 자연스러움을 결정하는 요소로 크게 어절 단위의 악센트와 구나 문장단위의 억양으로 나누어 볼 수 있다. 한국어의 악센트에 대해서는 학자에 따라 많은 학설이 있으나 대체로 음의 지속시간(duration), 음의 고저(pitch), 음의 세기(intensity)의 복합적 현상으로 설명할 수 있다. 문장 단위의 억양은 문장의 형태나 구나 절의 경계에 대한 정보를 포함하고 있다고 알려져 있으며, 이는 음의 기본 주파수의 변화에 따라 조절된다.

이러한 악센트와 억양은 음의 자연스러움과 밀접한 관계가 있을 뿐 아니라, 화자의 감정, 의도 등 음의 부가정보를 제공한다. 또한 실제로 음의 자연스러움과 음의 명료성 사이에는 상호 보완적 관계가 있다고 알려져 있으므로 그 중요성이 한층 부각된다고 할 수 있다.

화자의 감정이나 의도 등의 정보를 명확하고 자연스럽게 전달하기 위해서는 문장의 형태나 구의 경계 등의 의미적, 문법적 정보를 추출해 낼 필요가 있으며, 이러한 정보들은 앞장에서 설명한 언어학적 처리부를 통해 얻어진다. 따라서 언어학적 처리부에서의 분석의 정확도가 음의 자연스러움에 큰 영향을 미치게 된다. 본 절에서는 음의 자연스러움에 관련된 악센트와 억양에 의해 조절되는 한국어 운율에 관하여 살펴보고자 한다.

악센트를 조절하는 요소로서 먼저 지속시간에 대한 규칙에 대해 살펴보도록 한다. 앞에서 설명한 것처럼 음소의 지속시간은 악센트에 영향을 주어 자연스러움을 결정할 뿐 아니라 발화 속도를 결정할 수 있으므로 중요한 요소중의 하나이다. 또한 자음의 경우 그 지속시간이 그 자음의 음가에 영향을 미친다고 생각되므로 그 중요성은 더욱 커진다고 할 수 있다. 음소의 지속시간은 심리적, 의미론적 요인에 따라서도 변화할 수 있으나 이러한 정보는 추출하기가 쉽지 않으므로 본 연구에서는 구와 절의 경계, 단어내의 음절 위치, 한 음소의 앞뒤 등의 형태론적 요인만을 생각하여 각 음소의 지속시간을 조절하였다.

1) 한 어절내의 음절 수와 한 음절당 발음 길이는 반비례한다. 즉, 다섯개의 음절로 이루어진 어절내의 음절들의 한 음절당 발음 길이는 세개의 음절로 이루어진 어절내의 음절당 발음 길이보다 짧다.

2) 모음의 길이는 뒤에 오는 음소에만 영향을 받는다. 앞에 오는 음소에 의해서는 영향을 받지 않는 것으로 가정한다.

3) 악센트가 있는 음절은 강세 발음된다.

4) 구나 절, 문장의 마지막 음절에서는 강세 발음된다.

5) 어절의 마지막 음절에서는 강세 발음하게 한다.

6) 모음 뒤에 파음이 올 때에는 앞의 모음의 길이를 줄인다.

7) 모음으로 끝나는 어절에서는 그 마지막 모음을 강세 한다.

다음으로 음의 세기(intensity)에 대해 살펴보면, 음의 세기에 대해서는 정확히 알려진 바가 없고 또 악센트에 영향을 미치는 요소 중 가장 영향이 적다고 추측되므로, 단독의 악센트가 있는 음절의 세기를 조금 더 증가시키고, 문장 형태에 따라 평서문인 경우 문장이 진행됨에 따라 그 세기를 서서히 감소시키고, 의문문과 같은 경우는 그 세기를 감소 시키다가 문미에 와서 다시 약간 증가시키는 등의 비교적 간단한 방법을 이용해 구현하였다. 합성음의 세기에 대한 보다 섬세한 조절을 위해서는 향후 이에 대한 심도있는 연구가 계속 진행되어야 할 것이다.

마지막으로, 억양에 관해 살펴보기로 하자. 우선 문장 전체의 기본 주파수 패턴을 살펴보면, 문장 끝으로 갈수록 점차 하강하는 형태가 거의 모든 언어에 공통된 현상이다. 그리고 pause가 포함된 문장의 경우, 기본 주파수 패턴은 문장의 첫 부분에서 pause가 있는 곳까지 하강하며 pause가 있는 곳에서 기본주파수가 급격히 상승한 후 다시 점차 하강, 의문문에서는 상승 하강, 예/아니오의 대답을 요구하는 의문문에서는 상승, 그리고 감탄문에서는 상승 하강의 형태를 갖고 있다. 그러나 이와 같은 사실이 모든 경우에 일률적으로 적용되는 것은 아니고, 더 구나 불완전한 문장 등에 대해서는 연구된 바가 부족하므로 앞으로 계속되어야 할 것이다. 본 연구에서는 시간의 함수인 기본주파

수 패턴을 어절 단위의 악센트와 문장 단위의 억양으로 구분하여 그 각각을 stepwise command에 대한 2차 선형 시스템의 출력으로 모델링하여 구현하였다[10].

2.3.2 음성파형 생성

음성 합성을 위해 고려할 수 있는 음성 부호화 방식은 크게 세가지 부류로 나눌 수 있다. 그중 첫째는 파형 부호화 방식으로서, 음성 파형을 PCM, ADPCM 등으로 부호화하여 컴퓨터에 저장한 뒤, 합성할 문장에 필요한 데이타베이스를 꺼내어 연결시켜 음성 파형을 만들어내는 방법이다. 두번째 방식은 vocoding 방식으로서, 사람의 발성기관을 모델링하여 음성을 합성하는 방식이다. 마지막 세번째는 articulatory synthesis 방식으로서, 사람이 발음할때 발음기관이 변화하는 모양을 X-ray 분석을 통해 관찰함으로써 발음기관의 모양을 모델링하여 음성을 합성하는 방법이다. 이상에서 설명한 세가지 음성부호화 방식중에서, 무제한 어휘를 대상으로하는 Text-to-Speech합성 시스템에는 vocoding 방식이 가장 널리 사용되고 있다. Vocoding 방식에 의한 음성합성 방법에는 formant synthesis와 LPC(Linear Predictive Coding) synthesis를 들 수 있다. 본 시스템은 LPC 합성방식에 기초를 두고 있다.

일반적인 LPC 해석 방식은 음성신호가 10~20 msec 정도의 짧은 구간에서는 stationary 하다는 가정하에, 이 구간의 음성신호에 창함수를 취하여 LPC 파라미터를 구하며, 따라서 매 10~20 msec 마다 파라미터를 update한다. 그런데 이러한 LPC 해석 방식은 음성신호가 quasi-stationary하다는 가정에 기초를 두기 때문에, 음성신호의 특성이 빨리 변하는 transient 부분은 정확히 모델링하기가 어렵다. 따라서 여기서는 기존 LPC 방식의 문제점을 해결하기 위하여 resursive least square(RLS) 알고리즘을 사용했다. 이 알고리즘은 수렴 속도가 빠르고 파라미터 추적 특성이 우수할 뿐만 아니라, Log-area-ratio(LAR) 파라미터의 frequency sensitivity 특성이 비교적 일정한 장점을 갖고 있다. 그런데, 이 방식은 매 음성 샘플마다 파라미터들이 update되는 pointwise processing을 하므로 기존의 blockwise LPC 방식에 비해 저장해야 할 데이타 양이 많기 때문에 음성합성에 적용하기 위해서는 적절한 데이타 감축을 수행해야 한다.

2.3.3 음성 데이타베이스 구축

무제한 음성합성 시스템은 새로운 단어나 문장을 합성하기 위하여 이전에 추출해 놓은 음성 단편을 연결하는 방식을 취한다. Diphone을 데이타베이스의 기본 단위로 삼아 음성을 합성하는 경우에 중요한 문제는, 이러한 diphone의 시작점과 끝점, 즉 경계를 최적으로 위치시키는 것이다. 추출해 놓은 diphone을 연결하여 단어나 문장을 구성할 때 이 diphone들은 추출해 낸 환경과는 다른 순서로 연결된다. 따라서 연결 부위에 불연속이 발생하게 되어 음질의 명확도가 떨어지게 된다. 본 절에서는 diphone을 추출하는 과정에서 이와 같은 불연속성을 최소화하고, diphone을 연결하는 과정에서 부각적인 부상을 하지 않도록 최적인 diphone 경계를 위치시키는 방법에 대해 다룬다.

우선 diphone의 종류와 각 종류에 해당하는 diphone의 수, 그리고 diphone 추출을 위한 참고 단어에 대하여 살펴보자. Diphone은 크게 7가지로 분류했는데 각 경우에 해당되는 diphone에 대해 살펴보자[11]. 데이타베이스에 저장해 놓은 diphone은 모두 1373개이다.

첫번째 종류의 diphone은 '묵음+초성+중성'으로 구성되어 있으며, 이 경우는 대체적으로 단어의 첫음절에 해당된다. 예를 들어 '아기'에서 '묵음+ㅏ'가, '남자'에서 '묵음+ㄴ+ㅏ'가 한 diphone이다. 이때 묵음의 길이는 초성이 시작되기 전 10msec 정도이다. 이 경우에 산술적으로 계산한 diphone의 수는 399개이나 실제적으로 존재하는 diphone의 수는 259개이다.

두번째는 '중성+종성+묵음'으로 구성된 diphone이고, 이 경우는 대체적으로 단어의 끝음절에 해당되며, 묵음의 길이는 종성이 끝난 뒤 45 msec 정도이다. 예를 들어 '아내'에서 'ㅐ+묵음'이, '서울'에서 'ㅜ+ㄹ+묵음'이 한 diphone을 이룬다. 이 경우에 해당되는 diphone의 수를 조사해 보면, 중성에는 21개의 모음이 오고, 종성에는 7개의 대표 자음(ㄱ, ㄷ, ㅂ, ㄴ, ㄹ, ㅁ, ㅇ)과 종성이 없는 경우가 있으므로 총 168개의 diphone이 있으나, 이 경우에도 실제적으로 존재하는 diphone은 147개이다.

세번째 종류는 '중성+중성'으로 구성된 diphone이고, 이는 모음끼리 연결되는 경우에 해당되는데, 예를 들어 '오이'에서 'ㅗ+ㅣ'가 한 diphone이다. 이 경우에 diphone의 수는 모두 441개 중에서 308개만 실제적으로 존재한다.

네번째 종류의 diphone은 '중성+종성(ㄴ, ㄹ, ㅁ, ㅇ)'으로 구성되어 있다. 두번째 종류의 diphone과의 차이는 두번째 종류는 단어의 끝음절에 오는 diphone이고, 네번째 종류는 한 단어내에서 바로 뒤에 초성이 연속하여 나타나는 경우이다. 예를 들어 '찬미'에서 'ㅏ+ㄴ'이, '상사'에서 'ㅏ+ㅇ'이 한 diphone을 이룬

다. 그런데 이 경우에 diphone의 수는 다음에 오는 초성에 따라 세가지씩 존재하여 모두 165개가 된다.

다섯번째는 '종성(ㄴ, ㄹ, ㅁ, ㅇ)+초성(ㄴ, ㄹ, ㅁ, ㅇ, ㅎ)+중성'으로 구성된 diphone이고, 예를 들어 '찬미'에서 'ㄴ+ㅁ+ㅣ'가 한 diphone이다. 이 형태는 diphone이라고 볼 수 없으나 음성의 여러 음향학적 특성을 고려하여 한 부류의 diphone으로 구분지었다. 이 경우에 해당되는 diphone의 수는 한 음절의 종성과 다음 음절의 초성 사이에 존재하는 음소 결합의 제약으로 인하여 420개에서 155개로 줄어든다.

여섯번째 종류는 '종성+초성(ㄱ, ㄷ, ㅂ, ㄴ, ㄹ, ㅁ, ㅇ, ㅅ, ㅈ, ㅎ)'으로 구성된 diphone이고, 이 경우는 음절과 음절이 연결될 때 앞 음절의 종성 모음과 다음 음절의 초성이 연속하여 나타나는 경우이다. 예를들어 '아기'에서 'ㅏ+ㄱ'이, '공이'에서 'ㅗ+ㅇ'이 한 diphone을 이룬다. 초성이 'ㄱ, ㄷ, ㅂ'인 경우에는 이들이 모음 사이에서 유성음화 되므로 변이음을 데이타베이스에 갖게 되는 장점이 있다. 이 경우에 diphone의 수는 179개이다.

일곱번째 종류의 diphone은 '초성(ㄱ, ㄷ, ㅂ, ㄴ, ㄹ, ㅁ, ㅇ, ㅅ, ㅈ, ㅎ)+중성'으로 구성되어 있는데, 이 경우는 여섯번째 종류에 연속해서 발생한다. 예를 들어 '아기'에서 'ㅏ+ㄱ'이 여섯번째 종류로, 'ㄱ+ㅣ'가 일곱번째 종류의 diphone이다. 이 경우에 diphone의 수는 160개이다.

Diphone의 경계를 추출해 내어 diphone 데이타베이스를 구축하는 방법은 다음과 같다. Diphone의 경계는 Kaeslin[9]이 제안한 알고리즘을 이용하여 선정했다. 우선 4회 발음한 참고 단어들에 diphone의 경계가 될 수 있는 시작 frame과 끝 frame을 선정한다. 그리고 나서 centroid vector를 구하는데, 이를 위한 초기값은 각 참고단어의 기준 frame과 거리가 가장 가까운 frame으로 정한다. 마지막으로 cost 함수를 구하는데, 이 함수는 음성신호의 stationarity와 centroid vector와의 거리가 포함되어 있다. Diphone의 경계는 이 cost 함수가 최소인 frame이 되나, 불연속성이 최소가 되도록 diphone의 경계를 조정하는 것이 필요하다.

Ⅲ. 시스템 구현

3.1 하드웨어 구성

가라사대는 PC의 plug-in 카드 형태로 개발되었으며 저장된 문장 혹은 키보드를 통하여 입력되는 문장을 실시간으로 음성신호로 바꾸어 발생시킬 수 있다. 본 음성 변환 시스템의 하드웨어는 제어 및 PC 접속부, 디지탈 신호처리부, 아나로그 회로부의 세부분으로 구성된다. 해당 블럭도가 그림 2에 있다.

제어 및 PC 접속부의 역할은 PC로 부터 명령을 받

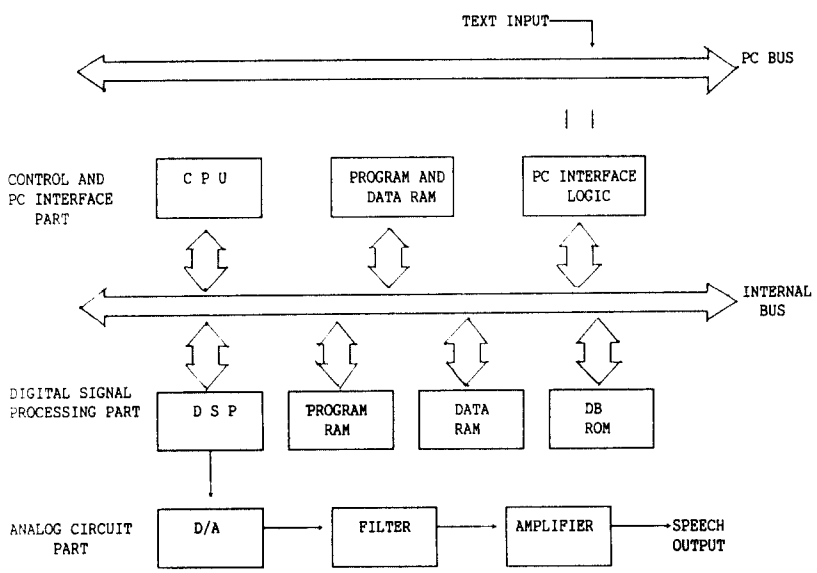


그림 2. 하드웨어 구성도

아 발화속도, 운율등과 같은 문자음성변환 시스템의 파라미터를 조정하고 문자음성 변환의 결과를 PC에 전달하는 일을 한다. 또한 문장 전처리나 발음표기 변환 등과 같은 언어처리의 일부도 수행한다. 제어 및 PC 접속부는 PC와 polling이나 interrupt 방식으로 통신할 수 있다. PC는 가라사대와 I/O 맵 방식으로 통신하며 4개의 번지중 하나를 점퍼로 선택할 수 있다. 이상의 제어 및 PC 접속부는 8 bit one-chip 마이크로프로세서에 의하여 수행된다.

디지탈 신호처리부는 강세와 억양을 포함하는 운율의 조정을 맡는다. 이를 위해서 한국어의 강세규칙이 단어단위로 적용되고 입력문장의 피치 패턴과 지속시간을 제어하기 위한 문장분석 알고리즘이 수행된다. 또한 디지탈 신호처리부는 합성단위의 접속과 음성신호의 발생을 맡고 있다. 합성단위의 자연스러운 연결을 위하여 음성신호의 spectral coefficient, 단구간 에너지, 피치 패턴등을 interpolation 하는 기법을 사용하였다. 신호발생 과정에는 LPC lattic 필터를 사용하였으며 합성 필터의 사양이 표 1에 있다. 이처럼 디지탈 신호처리부는 많은 계산량을 필요로 하므로 16 bit 부동 소수점 디지탈 신호처리기를 사용하였다.

표 1. 합성 필터의 사양

파라미터	반 사 계 수
차 수	10
형 태	Lattice
샘플링 주파수	10kHz
Cut-off 주파수	4.5kHz
여기신호	Stored waveform 또는 잡음

아날로그 회로부는 D/A converter, 저역통과 필터, 증폭기 등으로 구성되어 있다. D/A convert는 16 bit 이며 cut-off 주파수가 4.5kHz인 4차 butterworth filter가 저역통과 필터로 사용되었다. 헤드폰이나 스피커를 외부 증폭기 없이 연결하기 위하여 최대 0.5W의 증폭기를 내장시켰다. 신호출력 level은 가라사대 뒤쪽의 volume 다이얼을 돌려서 조정할 수 있다.

3.2 소프트웨어의 구현

가라사대의 소프트웨어는 다음의 세 부류중의 하나에 속한다. 첫번째는 가라사대 하드웨어 상에서 동작하는 언어처리 프로그램, 음성처리 프로그램, PC 접속프로그램을 포함하는 음성합성 소프트웨어이다. 두번째는 PC 상에서 수행되는 응용 프로그램이다. 예를 들면 문장을 읽어주는 기능을 가진 문서편집 소프트웨어, 전자메일이나 정보서비스의 내용을 음성으로 바꾸어주는 소프트웨어, 사용자와 음성으로 대화하는 게임 프로그램등이 개발되었다. 또한 시각 장애자를 위하여 컴퓨터에 display된 내용이나 키보드 입력을 읽어주는 프로그램이 개발되어 널리 사용되고 있다. 이러한 소프트웨어는 문자화일이나 키보드 입력을 간단한 명령을 내려 음성으로 들어볼 수 있으며 발화속도나 운율같은 시스템 파라미터도 소프트웨어에 의하여 제어할 수 있다. 최근에는 전화접속기능을 가진 보드와 가라사대를 연결하여 전자메일의 내용을 전화기를 통하여 들을 수 있는 소프트웨어가 개발되어 사용되고 있다.

세번째는 응용 프로그램과 음성합성 프로그램을 접속시켜주는 I/O 드라이버 프로그램이다. 드라이버 프로그램은 응용 프로그램으로 부터 제어 파라미터나 문장 데이타를 받아 이를 음성합성 프로그램에 전달하는 일을 한다. I/O 드라이버 프로그램을 이용하면 응용 프로그램 작성자가 직접 가라사대 보드를 제어할 필요가 없다. 즉 프로그램 작성자는 문자 데이타를 GARASADE라는 장치에 직접 쓰거나 제공되는 각종 응용함수를 호출하면 된다. 이러한 방식을 선택하므로써 가라사대 version 사이의 호환성과 단순한 프로그램 환경을 제공할 수 있었다.

3.3 기타기능 및 응용 예

가라사대는 범용의 CPU와 DSP를 갖춘 보드이므로 여기에서 동작하는 소프트웨어를 교체하면 다른 기능을 추가할 수 있다. 이러한 특성을 이용하여 문자음성변환 알고리즘에 PCM 및 ADPCM으로 저장된 화일을 재생하는 기능과 보통의 A/D 보드에 의하여 얻어진 음성데이타를 PCM이나 ADPCM으로 압축하는 기능을 별도로 갖추었다. 여기서 사용한 PCM이나 ADPCM은 국제표준인 μ-law PCM 및 CCITT G.721 ADPCM으로 다른 음성화일과의 호환성을 고려하였다. 또한 음성데이타를 LPC 방식에 의하여 2.4 Kbps로 압축하고 재생하는 기능도 개발하였으나 현재 시판중인 장치에는 이를 포함하지 않았다. 이상의 재생기능은 음성합성기능과 혼용하여 사용이 가능하므로 사용자의 필요에 따라서 음성재생 기능이나 합성기능을 선택하여 응용 프로그램을 구성할 수 있다.

현재 가라사대를 이용하고 있는 응용제품중 몇가지 예는 다음과 같다. 첫째 맹인용 컴퓨터에 활용하고 있는 예이다. 이 컴퓨터는 가라사대를 기본으로 장착

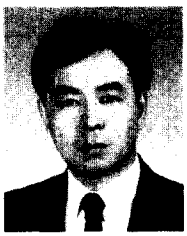
하고 맹인들의 키보드 입력을 즉시 들려주는 기능이나 화면에 표시된 내용을 줄 단위나 전체로 읽어주는 등의 기능을 갖춘 소프트웨어를 올려서 구성한다. 현재는 맹인용으로 활용되고 있지만 앞으로는 농아자를 위한 발성장치로서도 활용할 수 있을 것이다. 두번째는 문서편집기의 내용을 확인하여 주는 제품이다. 이는 선택된 부분의 내용을 들려주는 기능뿐만 아니라 문서편집기 동작상의 오류를 사용자에게 알려주는 기능도 갖고 있다. 세번째는 전자메일의 내용을 음성으로 바꾸어 전화를 통하여 이용자에게 전달하는 내용이다. 이 서비스는 전자메일 가입자가 데이타 단말을 이용하기 어려운 상황에 있을때 매우 유용한 방법이다. 이 서비스를 위해서 가라사대는 전화선 접속보드와 연결되어 하나의 시스템을 구성하여야 한다. 현재 이러한 서비스를 위한 장치가 개발되어 운용되고 있다.

Ⅳ. 결 론

본 논문에서는 한국어 문자음성 변환시스템인 가라사대에 관하여 살펴보았다. 우선 한국어 문자음성 변환시스템의 일반적인 고려사항과 가라사대의 특징을 살펴보았다. 그리고 가라사대의 문자음성 변환 알고리즘을 소개하였다. 문장 전처리, 발음표기 변환과 운율조절에 관하여 살펴보고 합성단위 생성 및 연결, 음성 신호발생과 같은 음성신호처리 과정도 서술하였다. 다음으로 가라사대 구현에 있어서 고려된 사항들을 하드웨어 구성과 소프트웨어 구현으로 나누어 설명하였다. 이 밖에 하드웨어의 DSP를 이용하여 다른 기능을 부가하였으며, 현재는 가라사대의 음질을 개선하고 여성음성을 추가하기 위한 연구가 수행중이다.

참 고 문 헌

1. Jung-chul Lee at. al., "Speech synthesising using demisyllables for Korean:a preliminary system," in Proc. int. Conf. Spoken Language Processing, pp. 19. 3. 1-19. 3. 4
2. 성광보 외, "반음절 데이타베이스를 이용한 MPLPC 한국어 무제한 단어음성 합성기의 제작," Korean-Japan Joint Workshop Advanced Technology of Speech Recognition and Synthesis, pp. 298-303, 1991. 6.
3. 김상용 외, "한국어 문자음성변환 시스템 개발," Korea-Japan Symposium on Acoustics, pp. 261-270, 1991. 6.
4. 허웅, 국어음운론, 샘 문화사, 1985.
5. J. Allen, M. Hunnicutt and D. Klatt, From text to speech, Cambridge University Press, 1987.
6. H. Fujisaki and H. Kawai, "Realization of linguistic information in the voice fundamental frequency contour of spoken Japanese," in Proc. ICASSP, pp. 663-666, 1988.
7. CH. H. Shadle and B.S. Atal, "Speech synthesis by linear interpolation of spectral parameters between dyad boundaries," in Proc. ICASSP, pp. 577-580, 1978.
8. 이현복, "한국어 리듬의 음성학적 연구"
9. H. Kaeslin, "A systematic approach to the extraction of diphone elements form natural speech," IEEE Trans. ASSP, pp. 264 271, Apr. 1986.
10. K. Hirose and H. Fujisaki, "Analysis and synthesis of voice fundamantal frequency contrours of spoken sentences," IEEE Trans. ASSP, pp. 950-953, 1982.
11. 권칠홍, "Diphone을 이용한 한국어 음성합성 시스템에 관한 연구," 한국 과학기술원 석사학위 논문, 1989.
12. 안승권 외, "한국어 문장-음성 변환기의 운율 제어용 구문분석기," 제9회 음성통신 및 신호처리 워크샵 논문집, pp. 218-223, 1992.
13. 다지소, 가라사대 사용설명서, 1991. 12.



권 철 홍

- 1963년 7월 30일생
- 1987년 2월 : 서울대학교 전자공학과 학사
- 1989년 2월 : 한국과학기술원 전기 및 전자공학과 석사
- 1994년 8월 : 한국과학기술원 전기 및 전자공학과 박사
- 1989년 1월 ~ 1994년 8월 : 디지콤 정보통신연구소 연구원
- 1994년 9월 ~ 현재 : 디지콤 정보통신연구소 선임 연구원
- 주관심분야 : 음성신호처리, 디지탈 신호처리, 음성정보 시스템

구 준 모

- 1963년 1월 13일생
- 1985년 2월 : 서울대학교 전자공학과 학사
- 1987년 2월 : 한국과학기술원 전기 및 전자공학과 석사
- 1991년 8월 : 한국과학기술원 전기 및 전자공학과 박사
- 1987년 7월 ~ 1992년 2월 : 디지콤 정보통신연구소 연구원
- 1992년 3월 ~ 현재 : 디지콤 정보통신연구소 선임 연구원
- 주관심분야 : 음성신호처리, 디지탈 신호처리, 음성 및 팩스 메세지 시스템

정 원 국

- 1964년 5월 5일생
- 1986년 2월 : 서울대학교 전자공학과 학사
- 1988년 2월 : 한국과학기술원 전기 및 전자공학과 석사
- 1993년 8월 : 한국과학기술원 전기 및 전자공학과 박사
- 1988년 3월 ~ 1993년 8월 : 디지콤 정보통신연구소 연구원
- 1993년 9월 ~ 현재 : 디지콤 정보통신연구소 선임 연구원
- 주관심분야 : 음성신호처리, 디지탈 신호처리

김 형 순

- 1960년 8월 21일생
- 1983년 2월 : 서울대학교 전자공학과 학사
- 1989년 2월 : 한국과학기술원 전기 및 전자공학과 박사
- 1986년 9월 ~ 1992년 8월 : 디지콤 정보통신연구소 선임연구원
- 1992년 9월 ~ 현재 : 부산대학교 전자공학과
- 주관심분야 : 음성신호처리, 디지탈 신호처리