Text-to-Speech 합성음 품질 평가

정 유 현, 최 준 혁, 한 민 수
한국전자통신연구소 음성응용연구실

Assessment of Synthesized Speech by Text-to-Speech Conversion

Y.H. Jeong, J.H.Choi, M.S.Han
Electronics and Telecommunications Research Institute

요 약

본 논문은 한국전자통신연구소 음성응용연구실에서 개발한 문자-음성변환 시스템(Text-to-Speech Conversion System)의 음질개선 연구의 일환으로 Phoneme-Balanced Words 110개에 대해서 개선전 시스템(V.1)과 개선 후 시스템(V.2)을 대상으로 각각 실시한 명료도 실험결과에 대하여 기술하고 있다. 본 실험의 목적은 연구개발자 입장에서 합성음 개선에 대한 정량적 성과 및 문제점 파악을 위한 진단형 평가이며 남자 5명, 여자5명을 대상으로 1회 실시한 청취 시험결과 V.1에 대해서는 최저 37.3%(41개) ~ 최고 55.5%(61개)이고, V.2에 대해서는 최저 39.1%%(43개) ~ 최고 60.9%(67개) 결과를 얻었다.

I. 서 론

한국전자통신연구소 음성응용연구실에서는 통신처리장치의 미디어 변환(Text-to-Speech) 서비스를 제공하기 위하여 반음절 단위의 LSP방식에 의한 한국어 문장-음성변환(Text-to-Soeech Conversion) 시스템(글소리 I)을 개발하였다[1]. 초기의 글소리 I은 합성음에 익숙하지 않은 사람이 듣기에는 거부감이 많았다. 따라서 당 연구실에서는 글소리 I의 합성음 음질개선을 위한 여러가지 시도를 하고 있으며[2][3], 동시에 연구개발자 입장에서 이러한 개선결과를 객관적으로 평가하기 방법을 연구중에 있다.

문자-음성변환 시스템의 합성음 품질을 평가하기 위한 객관적인 방법은 아직 확립되어 있지 않으나 일반적으로 피험자가 합성음을 듣고 언어의 내용을 얼마나 이해할 수 있는가를 평가하는 명료도(Intelligibility)와 합성음으로서 얼마나 자연스럽고 듣기 쉬운지를 평가하는 자연성(Naturalness)을 평가하는 주관적인 방법을 많이 이용하고 있다[4].

본 논문에서는 현재 당 연구실에서 수행중인 문자-음성변환시스템의 음질개선 결과에 대한 정량적 평가 및 문제점 파악을 위한 진단형 평가로써 1회 실시한 명료도 실험에 관하여 기술하고 있으며, 시험 단어들은 당 연구소에서 국민학교 고빈도 단어 4,084개를 모집단으로 하여 추출한 Phoneme-balanced word 445개[5] 중 110개를 선택한 것이다.

II. 글소리 I 시스템 개요

1. 반음절 DB

본 시스템에서 사용한 반음절 DB는 기본형인 CV, VC, V의 반음절을 대상으로 총 640개로 음성 자료를 12 bit, 10Khz로 샘플링하여 10ms의 프레임 단위로 12차의 LSP파라메터로 분석한 것이다.

2. 문자-음성변환 시스템의 전체 흐름도

한국어 문장(완성형 코드)을키보드로 입력하면 숫자음 및 약어처리, 예외규칙 처리, 운율 및 경계분석을 한 후 여러가지 음운규칙을 적용하여 소리나는 형태의 발음기호열로 변환시킨다. 생성된 발음 기호열에 대해서 우선 경계정보를 이용하여 각 음소의 지속시간을 산출한 뒤 각 반음절 DB 검색 코드를 생성한다. 이 코드를 이용하여 계산된 지속 시간만큼 반음절 DB로 부터 합성 파라메터를 읽어와서 인접 반음절사이에서 평활화처리를 한다. 다음에 에너지 조절 요소를 이용하여 반음절 단위로 에너지값

을 계산해 윤곽선을 생성하며 경계점 정보를 이용하여 기본 주파수의 제어를 함으로써 최종적인 합성 파라메터 열을 생성한다. 마지막으로 이 파라메터들을 사용하여 합성 필터를 구동함으로써 합성음을 만들어낸다. 이와 같은 합성과정의 전체흐름도는 그림 1과 같다.

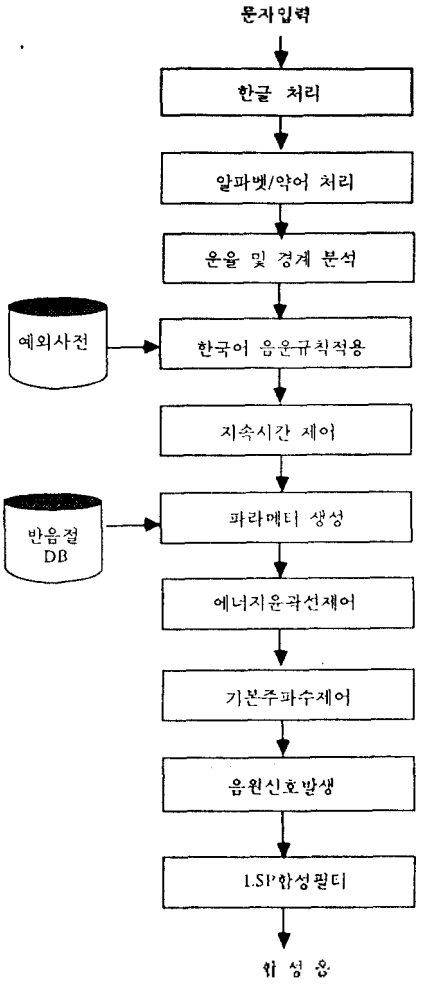

그림 1 문자-음성변환 시스템 전체 흐름도

III. 음질개선 연구

당 연구실에서는 합성 시스템의 기본적인 합성방식을 변경하지 않는 상태에서 음질을 개선하기 위하여 에너지 윤곽선 조절 및 Excitation Source Mode 에 의한 음질개선에 관한 기초 연구를 수행하였다.

1. 에너지 윤곽선 조절

에너지는 합성음 음질에 영향을 미치는 요소로써 문장중에서 에너지 흐름을 에너지 윤곽선이라고 한다. 본 연구에서는 최소한의 기본적인 에너지 윤곽선에 대한 규칙을 얻고자

- 모든 모음은 같은 에너지를 가진다.
- 자음은 그룹별 반음절 형태에 의해서만 다른 에너지를 가진다.

와 같은 가정 아래 모음의 시작구간, 모음의 끝구간, 반음절이 음절의 앞 반음절인 경우, 반음절이 음절의 뒷 반음절인 경우의 모음부분에 대해서 에너지를 구하고, 만약 자음이 있는 경우에는 초성 또는 종성의 에너지를 구한 다음 서로 연결되는 부분에서 이차 함수를 이용하여 매끄럽게 이어주었다. 반음절 단위로 에너지가 구해지면 인접 반음절사이의 음소간의 영향을 고려하여 모음 또는 유성음, 모음과 유성자음, 모음과 무성자음별로 평활화 모델을 선택하여 반음절과 반음절사이의 접속구간에서 에너지를 평활화처리를 하므로써 실험 단계이지만 청취를 통해 개선된 합성음을 얻었음을 확인하였다.

2. Excitation Source Model

본 연구에서는 differentiated 단일 pulse음원의 단점인 제1 포먼트의 약화를 개선하고 피치간의 결합 부분을 보상하기 위해 LF 모델을 수정하여 사용하였으며, 이를 통해 명료도를 향상시킬 수 있었다.

상기 방법들은 아직 초기 단계로 보다 많은 실험 데이타의 분석을 통해서 보다 정확한 알고리즘에 관한 연구가 계속되어야 하며, 현재 이에 대한 연구가 진행중에 있다.

IV. Phoneme-balanced words에 의한 합성음 평가

합성음의 품질을 객관적으로 평가하기 위해서는 합성음을 듣고 언어의 내용을 얼마나 이해할 수 있는가를 평가하는 명료도(Intelligibility)와 합성음으로써 얼마나 자연스럽고 듣기 쉬운지를 평가하는 자연성(Naturalness)을 평가척도로 생각할 수 있다. 그러나 아직 우리말에 대한 객관적인 평가법이 확립되어 있지 않은 상태이므로 본 논문에서는 개발자의 입장에서 합성음의 음질개선 결과를 정량적으로 평가 및 문제점을 도출하기 위한 방법으로 음소환경을 모두 갖춘 Phoneme-balanced words를 대상으로

한 명료도 실험방법을 평가 척도로 이용하였다.

1. Phoneme-balanced words

당 연구소에서 국민학교 교과서에 나오는 고빈도 단어 4,084개를 모집단으로 선정한 phoneme-balancedwords 455개[7]를 모두 이용하기에는 너무 양이 많아 1/4(110개)를 선택하였으며, 선택된 110개 단어들은 다음과 같다.
훈민정음, 끊임없이, 좌표평면, 위생, 처들어오다, 정육면체, 즐거움, 층층대, 요일, 휴지, 여보세요, 쓰임, 최고, 무역, 섬유, 아홉, 팽이, 입체도형, 교양, 귀뚤귀뚤, 부채꼴, 서양, 특별히, 기준량, 꽃밭, 뻐대, 우체국, 가운데, 토요일, 의논하다, 여섯째, 떼럼, 빗방울, 어떤, 의무, 뛰놀다, 더위, 왼쪽, 회수, 둘러앉다, 경례, 태권도, 풍습, 외삼촌, 붙잡다, 드디어, 얕다, 여학생, 돌쇠, 괜히, 혜택, 마음씨, 법원, 받침대, 어촌, 하여금, 금빛, 몸집, 씨앗, 사회과, 달걀, 미역, 에벌레, 의체, 벗집, 갯수, 열심히, 으뜸, 케네다, 쇠똥, 이웃, 처음, 햇볕, 백분율, 내용알기, 나눗셈, 쉽다, 오랫동안, 윗몸,왕릉, 유통, 공예품, 자취, 아주머니, 야단, 이제까지, 짜임새, 용액, 해수욕장, 습관, 화음, 어업, 되찾다, 원양, 예금액, 뻐꾸기, 온도계, 바뀌다, 소위, 와치다, 싸움터, 예절, 뜻밖에, 느낌, 차려, 왜구, 높이뛰기, 규칙, 일제, 휴식

2. 명료도 시험 결과 및 고찰

Phoenem-balanced words 110개를 단어들을 구 문자-음성변환 시스템(V.1)과 개선된 문자-음성변환 시스템(V.2)에 입력순서를 다르게 하여 합성시킨 합성음을 DAT에 녹음한 후 조용한 사무실에서 합성음을 청취한 경험이 없는 남성 5명 여성 5명을 대상으로 1회에 실시한 청취 실험을 행하였으며, 그 결과는 표 1과 같다.

V.1 시스템은 최저 37.3%(41개) ~ 최고 55.5%(61개)이고, V.2 시스템은 최저 38.2%(42개) ~ 최고 60.9%(67개)의 단어 인식율을 보였다.

V.1 시스템에서 10명이 모두 인식한 단어는
"훈민정음, 마음씨, 이등변삼각형, 예벌레, 여보세요, 이웃, 빗방울, 나눗셈, 오랫동안, 아주머니, 이제까지, 어업, 높이뛰기"로 모두 13개 이고, 10명이 모두 인식하지 못한 단어는
"위생, 처들어오다, 층층대, 회수, 외삼촌, 혜택, 받침대, 어촌, 케네다, 내용알기, 공예품, 자취, 예절, 야단, 원양, 예금액, 외치다, 예절, 일제"로 모두 19개이다.

V.2 시스템에서 10명이 모두 인식한 단어는
"달걀, 열심히, 오랫동안, 훈민정음, 둘러앉다, 얕다, 싸움터, 높이뛰기, 여학생, 예벌레, 백분율, 이등변삼각형, 이웃, 나눗셈, 소위, 느낌, 여보세요, 해수욕장"로 모두 18개이고, 10명이 모두 인식하지 못한 단어는 외치다, 왜구, 최고, 금빛, 층층대, 받침대, 일제, 내용알기, 사회과, 어촌, 뻐대"로 모두 12종이다.
사람들이 인식하지 못한 단어들을 살펴보면
외치다 --> 오징어(2), 마지막(1), X치다(1), 멋지다(1)
왜구 --> 외국(5), 매우(1)
최고 --> 생옥(1), X업(1), 생후(1), 생업(1)
금빛 --> 풀잎(1), X립(2), 범위(1), 품위(1)
어촌 --> 왼손(1), X손(1), X손(1)
뻐대 --> 병해(2), 열해(1)
층층대 --> 심심해(5), 쓸쓸해(3), 지중해(1), 습습해(1)
받침대 --> 아침에(5), 확실해(2), 자치회, 바침에
일재 --> 일의(3), 일예, 인재, 말회, 일래, 밀레, 인내
내용알기 --> 전혀 인식 못함
사회과 -->사회과학(3), 상이와, 사회X, 사화와, 사회화
와 같으며, 현재 상기의 결과들을 분석중에 있다.
위와 같은 결과로 아직 음질 개선의 초보단계이지만 합성음 개선 연구가 피험자들의 주관적인 평가을 통하여 최저치가 37.3%(41개)에서 39.1%(43개)로 최고치는 55.5%(61개)에서 60.9%(67개)로 증가했슴을 확인할 수 있었다.

표1 Phoneme-balanced words에 대한 명료도 시험 결과

성별		시스템명료도(단어수110개)	
		V.1	V.2
남	1	57개(51.8%)	59개(53.6%)
	2	56개(50.9%)	63개(57.3%)
	3	41개(37.3%)	43개(39.1%)
	4	57개(51.8%)	67개(60.9%)
	5	56개(50.9%)	60개(54.5%)
여	6	60개(54.5%)	58개(52.7%)
	7	44개(40%)	60개(54.5%)
	8	61개(55.5%)	62개(56.4%)
	9	51개(46.4%)	54개(49.1%)
	10	48개(43.6%)	48개(43.6%)

V. 결론

본 고에서는 당 연구실에서 개발한 문자-음성변환 시스템의 합성음 품질 개선 연구의 일환으로 Phoneme-balanced words 110개를 대상으로 실시한 명료도 실험 결과에 관하여 기술하였다. 본 실험의 목적은 연구개발자 입장에서 합성음 품질개선에 대한 정량적 성과 및 문제점을 파악하기 위한 시스템 진단형 평가로 실시한 것이다.

향후 추진 계획으로는 실험 결과를 분석하여 이를 음질개선 연구에 반영하고자 하며, 동시에 현재 640개인 반음절 DB를 1228개로 시스템을 개선하고자 한다.

감사의 글

본 실험에 많은 관심을 보여 주신 윤병남 부장님과 실험에 협조하여 주신 통신처리연구부 부원들에게 진심으로 감사드립니다.

참고 문헌

[1] ETRI, "통신처리장치를 위한 음성정보변환 기술 개발", 최종보고서, 1990.6.

[2] 이승훈, 한민수, "Audiotex 음질개선연구" 제1회 ETRI 음성, 언어 및 음향정보처리 워크샵, 1993.4.

[3] 강동규, 한민수, "Exitation Source Model에 의한 ATX 음질 개선에 관한 연구" 제1회 ETRI 음성, 언어 및 음향정보처리 워크샵, 1993.4.

[4] Speech Input/output Assessment and Speech Database, Processing of ESCA WORKSHOP, 1989.

[5] ETRI, "대어휘 연속음성인식을 위한 음소인식기술개발", 최종보고서, 1991.7.