

맞춤형 동화구연 시스템구연에 관한 연구

박원, 배명진
승실대학교 정보통신공학과

A Study on the Fairy tale Narration System with Key-word Exchange

Won PARK, MyungJin BAE
Dept. of Info. &Telecomm. Engr., Soongsil Univ. Seoul 156-743, Korea

mjbae@saint.soongsil.ac.kr

Abstract

최근 유아기 아이들을 위한 교육매체의 발달로 각종 CD-ROM이나 테이프 등에서 성우의 목소리로 동화를 읽어주는 시스템이 많이 나와 있고, 또한 Web Book이 점차적으로 보편화가 되 가고 있다. 하지만 이런 획일적이고 균일화된 매체들은 아이들에게 금세 실증을 내게 하기 때문에 흥미 유발을 위해 동화의 주인공을 자기자신이나 친근한 사람의 이름 등으로 바꾸어 발성해 준다면 더욱 친근감 있게 받아들일 것이다. 본 논문에서는 기존의 성우가 발성하는 동화의 주인공 이름을 Test화자가 새로운 이름으로 발성을 해주면 기존 성우의 목소리패턴으로 바꾸어 동화를 읽어주는 시스템에 대해서 제안하고자 한다.

우선 Test화자가 발성한 목소리를 성우의 목소리로 바꾸어 주기 위해서 기존의 성우가 발성한 동화 주인공 이름과 Test화자가 발성한 이름과의 운율패턴을 비교하여 성우의 운율패턴에 일치시키고 성우의 목소리 패턴으로 변경된 새로운 주인공의 이름만을 기존의 동화 DB에 삽입하였다. 또한 에너지 패턴 조절은 기존의 성우가 발성한 기존패턴에 근사화 시켰고 끝점을 스프딩 시킴으로써 자연스런 발성이 되게 만들어주었다. 결과적으로 Mos Score가 3.873로 비교적 좋은 결과를 얻을 수 있었다.

I. 서론

사회가 고도로 산업화가 되어감에 따라 개인화 및 고립화가 심화되고 있다. 이러한 예로 부모와 자식간

의 대화가 줄어들고 유아기의 교육도 가족들에게 의존하기보다는 방송매체나 비디오, 오디오 또는 CD-ROM 타이틀에 의해 대체되고 있는 실정이다. 하지만 이러한 매체들은 아이들에게 획일화 된 교육을 제공하고 실증을 내게 만든다. 이런 이유로 유아기 아이들에게 흥미를 유발시키고 상상력을 키워주기 위한 대안이 필요한데 요즘은 멀티미디어의 보급으로 인간과 인간사이의 정보전달 뿐 아니라 인간과 기계사이의 정보전달도 매우 중요해 지게 되었다. 이에 인간의 가장 기본적인 정보전달 수단인 음성을 인간과 기계사이의 정보전달 수단으로 이용하려는 연구가 음성합성, 음성인식 등 여러 분야에 빠른 속도로 진행되고 있다[1]-[3]. 특히, 음성합성기술은 음성인식기술에 비해 구현이 용이하여 여러 분야에 응용되고 있다. 이러한 분야로서 기존의 매체에 저장된 동화를 부모의 음색이나 정감이 가미된 음색으로 구연해 줌으로써 자녀들의 정서함양에 도움을 주는 동화구연 시스템을 제안한 바 있다. 본 논문은 기존에 제안되었던 동화구연 시스템에서 성우의 목소리를 부모의 목소리로 음색을 변경하는 것과는 달리 동화 주인공의 이름을 아이들 자신의 이름이나 친근감 있는 사람의 이름으로 바꾸어 발성하게 한 다음 발성한 사람의 목소리를 성우의 목소리 패턴으로 바꾸어 이름부분이 빠져있는 동화 DB에 이름만 삽입하여 주는 시스템으로 동화 전체를 바꾸어 줄 필요 없이 주인공 이름 부분만을 성우의 운율패턴으로 바꾸어 끼워 넣기만 하면 되기 때문에 기존의 동화구연보다 구연이 쉽다는 장점을 가지고 있다.

본 논문에서는 동화로 “아기돼지 삼형제”를 이용하였다. 또한 음색변환을 위해 운율특성과 음운특성을 이용하였으며 기존의 이름부분이 빠진 동화DB와

음색이 변경된 이름 사이의 자연스런 연결을 위하여 전후 운율 변화특성이 충분히 고려된 몇 가지 문장을 화자가 발성하게 함으로써 이름부분의 DB만을 이용하였다.

II. 맞춤형 동화구연을 위한 음성합성 시스템

2-1. 시스템의 구성

기존의 동화구연을 위한 음성합성 방식으로는 주로 파형 녹음방식이 사용되었다. 파형 녹음방식은 보통 비디오, 오디오테이프, 자기디스크 등의 매체에 음성 및 음향 신호의 파형을 그대로 기록하고 필요시에 VTR이나 오디오플레이어, 컴퓨터 등을 이용하여 재현해 내는 방법이다. 이러한 방법은 녹음 당시의 상황을 그대로 재현해 주기 때문에 음질이 우수하다[1].

파형 녹음방식은 음성신호를 부호화 할 때 파형이 갖는 불필요한 잉여성분만을 제거하고 부호화 하여 저장하고 이를 이용하여 다시 재생하는 방법으로 음질이 우수하다. 하지만 인간의 개성과 감정을 나타내는 여기(excitation)정보와 의사정보를 나타내는 포먼트(formant)정보를 분리하지 않고 부호화 하기 때문에 발성자의 특징을 기존동화 성우의 특징으로 변경시켜야 하는 맞춤형 동화구연시스템의 합성법으로는 여러 가지 어려움이 따르게 된다[3]. 이러한 문제를 해결하기 위해서는 파형 녹음방식을 사용하면서도 피치를 변경시킬 수 있는 피치변경법과 더불어 운율과 음운을 제어할 수 있는 기술을 필요로 한다.

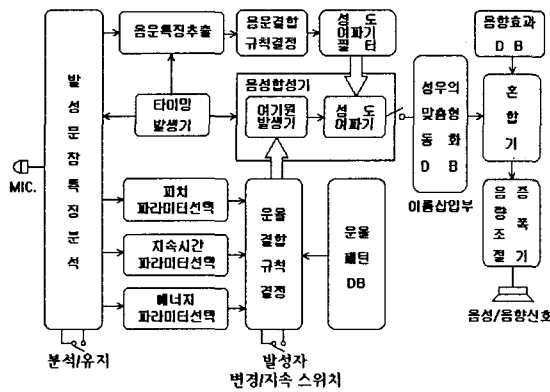


그림 2-1. 맞춤형 동화구연시스템 구성도

맞춤형 동화구연시스템의 전반적인 구성은 그림 2-1에 나타나 있는 바와 같이 화자특징 분석부, 운율 결합규칙 결정부, 음성 합성부, 음색 변경된 이름DB삽입부 등의 블록으로 구성할 수 있다. Test화자에 의해 발성된 문장을

화자의 특징 분석부에서 운율 정보를 추출, 기존의 성우 패턴 DB와 비교하여 결합규칙을 결정하고 합성부에 제공한다. 음성합성부는 이를 이용해서 합성음성을 성우의 패턴에 맞춘다.

2-2. 맞춤형 동화구연용 데이터베이스

음성의 규칙합성을 위해서는 합성이 기본이 되는 단위를 가능한 적은 단위로 구현하는 것이 효율적이다. 규칙합성에 사용되는 합성단위는 임의 어휘합성이 가능해야 하며, 음운의 명료성과 자연성에는 손상을 주지 않으면서 고음질을 얻을 수 있어야 한다. 또한, 합성에 필요한 모든 음성단위를 간단히 작성할 수 있는 단위를 사용해야 한다. 이러한 합성단위로는 음소, 반응절, 음절, 다이폰 등이 사용되고 있으며 음절, 다이폰에서 실현되지 않는 조음결합에 의한 음운변동을 흡수하기 위해 CVC나 VCV와 같은 더 큰 단위도 이용된다. 최근에는 고정된 길이의 합성단위보다 합성하고자 하는 음소열에 대한 다량의 음성데이터베이스를 이용하여 가장 적합한 환경 중에서 임의길이를 선택적으로 사용하는 방법도 시도되고 있다.

본 시스템에서는 동화의 전체를 합성하는 것이 아니라 문장에서 중요 주인공이름만 빠진 전체 동화 DB를 사용한다. 발성화자가 발성한 주인공 이름DB는 전처리로 기존의 동화에서 성우가 발성한(비어있는) duration길이 만큼 자른다. 또한, 기존의 성우가 발성한 동화에서 빠진 부분 전후의 운율특성을 충분히 반영하기 위하여 몇 가지의 특정 대표 문장을 발성하여 그 중에서 선택하여 사용한다.

III. 음색 변경법

3-1. 음운 패턴 추출

맞춤형 동화구연시스템을 구성하기 위해 동화로 야기 돼지 삼형제를 선정하였다. 발성화자의 음색을 이 동화의 구연자(성우)의 음색으로 변환하기 위해서는 먼저 발성자의 운율 및 음운 패턴을 추출해야 한다. 발성자의 운율패턴은 아래 그림 3-1의 블록도에 나타난 과정으로 추출하여 데이터베이스를 구성하였다.

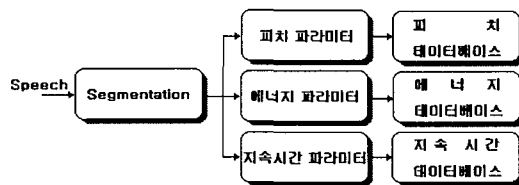


그림 3-1. 운율 패턴의 추출 과정

화자의 발성문장에서 우리가 원하는 운율 패턴을 추출하기 위해서는 특정한 단어나 문장을 발성해 주어야 한다. 이러한 단어나 문장은 성우의 운율 패턴을 충분히 반영해 주어야 한다. 본 시스템에서는 선택한 동화인 아기돼지 삼형제의 문장 중에서 주인공 이름 앞뒤 품사 등의 문법적 사항 등을 조사하여 성우 동화 DB에 끼워 넣을 때 좀 더 자연스런 발성이 되는 방향으로 대표문장을 만들었다.

다음 표 3-1은 맞춤형 동화구연 시스템에 사용될 주인공의 이름이 들어간 몇 가지 대표문장의 예를 나타낸 것이다.

주인공 단어목록 (연결어 포함)	대표문장의 예
와, # 돼지	문장 1 : # 돼지와, # 돼지는 # 돼지
# 돼지가	네로 갔습니다.
# 돼지도	문장 2 : 늑대는 #에게 문을 열라고 말했습니다.
# 돼지는	문장 3 : 아기돼지들은 #돼지가
# 돼지를	# 돼지를 도울 거라 생각
# 돼지에게	했습니다.
# 돼지네	문장 4 : #돼지도 말했습니다.

표 3-1. 아기돼지 삼형제의 대표 문장의 예

3-2. 음운 추출

음운 추출은 그림 3-2의 과정으로 처리하였다. 운율패턴과 마찬가지로 부모의 음운을 추출하기 위해서는 시스템의 특정 단어나 문장을 Test 화자가 입력해주어야 하는데 이러한 단어나 문장은 모두 변이음을 포함하는 문장이어야 한다. 이러한 문장의 예로 우리는 아래의 문장을 사용하였다.

/ 짧은 촛대형 빨간 모자를 눌러쓴 아이가 꿈을
꾸면서 카드표시 틱 비우기 놀이시합을 하고 있다. /



그림 3-2. 음운 패턴의 추출 과정

3-3. 음색 변경

성우의 목소리로 동화를 구연하기 위해 음성을 합성하는 규칙합성 시스템의 경우에는 합성어휘의 제한이 없어야 하고 화자의 음성으로부터 입력된 대표문장으로부터 음운정보와 운율정보를 추출하여 그에

따라 규칙적으로 음색을 조절하는 운율제어가 필요하다. 본 시스템에서는 고음질을 유지하면서도 음색을 변경하기 위해서 먼저 성우가 구연한 동화구연으로부터 추출한 운율패턴에 맞추어 피치, 에너지, 지속시간을 변경하였다. 이러한 변경을 위해서 본 연구실에서 제안한 피치변경률에 다른 최적의 피치변경법을 사용하여 운율변경을 수행하였다[4]. 한편 지속시간은 유성 및 무성 변이음 구간으로 나누어 처리하였다. 유성변이음 구간에서는 성우의 지속시간을 기준패턴으로 하여 피치구간을 반복하여 인터플레이션을 수행하거나 또는 피치구간을 몇 개증에서 슈아내는 과정을 수행하였다. 무성변이음 구간은 마찰성 음소와 파열성 음소구간을 별도로 처리하였다[5]. 에너지 패턴 조절은 성우가 발성한 기준 패턴에 일치시켰다. 유성 변이음 구간에 대해서는 피치주기 단위로 평균에너지를 구하여 일치시켰으며 무성음 구간에서는 2.5ms 단위로 평균에너지를 구하여 Test패턴을 일치시켰다.

다음에는 음운변경을 위해 운율패턴과 마찬가지로 발성화자가 변이음 단위로 구성된 문장을 발성한 다음 발성을 통해서 음운 파라미터를 채집한다.

이렇게 얻어진 음운과 음운파라미터를 가지고 음색변경과정에 적용하였다.

IV. 실험 및 결과

제안한 맞춤형 동화구연 시스템의 성능을 평가하기 위하여 IBM-PC 233MHz에 AD/DA를 인터페이스 하여 사용하였다. 선택동화인 아기돼지 삼형제에 임의의 주인공 이름이 들어간 대표문장을 2명의 남성과 2명의 여성에게 발성하게 하여 11kHz로 표본화하고 16bit로 양자화 하여 저장하였다.

문장 1 : 아기돼지 삼형제 중 ##돼지와, ##돼지는 ##돼지네 집으로 갔습니다.

문장 2 : 늑대는 ##돼지에게 문을 열라고 말했습니다.

문장 3 : 아기돼지들은 ##돼지가 ##돼지를 도울 거라 생각했습니다.

문장 4 : ##돼지도 말했습니다.

위의 대표문장들은 동화 문장 전후의 운율변화 특성을 생각해서 만들었다. 문장 1, 2, 3, 4 중 ##표시가 된 동화 주인공 이름부분만을 기존의 성우가 발음한 지속시간 길이로 자른 후 운율패턴과 음운패턴을 성우패턴으로 변경한 후 맞춤형 동화 DB에 끼워 넣었

다. 음색변경은 본 논문에서 제안한 방법으로 변경을 하였다. 대표문장에서 Segment되고 운율 및 음운이 변경된 단어들을 끼워 넣을 때 발생할 수 있는 불연속 점을 최소로 줄이기 위하여 앞 뒤 끝점을 스므딩시켜주었다.

성우의 음색으로 변환과 동화에 삽입한 문장의 자연성에 대한 선호도를 10명의 청취자로 구성된 3개의 청취 그룹에게 기록하게 하여 그 결과를 표 4-1에 나타내었다.

표 4-1. 음색과 자연성에 관한 합성음 선호도

청취자 그룹	Test 화자 음색	동화 합성음(성우패턴)			
		남1	남2	여1	여2
1	3.5	3.9	3.8	3.75	3.8
2	3.62	4.0	4.1	3.9	4.0
3	3.55	3.91	4.0	3.6	3.7
평균	3.56	3.94	3.97	3.75	3.83

V. 결 론

최근 유아기 아이들을 위한 교육매체의 발달로 각종 CD-ROM이나 테이프 등에서 성우의 목소리로 동화를 읽어주는 시스템이 많이 나와 있다. 하지만 이런 획일적이고 균일화된 매체들은 아이들에게 금세 실증을 내게 하기 때문에 흥미 유발을 위해 동화의 주인공을 자기자신이나 친근한 사람의 이름 등으로 바꾸어 발성해 준다면 더욱 친근감 있게 받아들일 것이다. 본 논문은 기존의 성우가 발성하는 동화의 주인공 이름을 Test화자가 새로운 이름으로 발성을 해주면 기존 성우의 목소리패턴으로 바꾸어 동화를 읽어주는 시스템에 대해서 제안하였다.

성우 목소리의 동화구연 중 동화 주인공 이름에서 추출된 운율과 음운 패턴과 Test화자가 발성해준 대표문장에서의 주인공 이름의 운율과 음운 패턴을 비교해 성우의 운율과 음운 패턴으로 바꾸어 줌으로써 음색을 변경하였다.

결과적으로 동화 주인공의 이름을 성우의 음색으로 변경하고 성우동화DB에 삽입한 합성음에 대한 MOS Score를 살펴보았을 때 평균 3.873로 비교적 좋은 결과를 얻을 수 있었다. 하지만 음색이 변경된 주인공 이름을 삽입할 때의 불연속점을 없애주기 위해 스므딩을 해주었기 때문에 연결부위의 명료도가 낮아지는 단점을 보였다.

VI. 참고문헌

1. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
2. G. Bristow, *Electronic Speech Synthesis*, McGraw Hill, 1984.
3. Panos E. Papamichalis, *Practical Approaches to Speech Coding*, Prentice-Hall, 1987
4. 박형빈, 조왕래, 김종득, 박원, 심도식, 배명진, "피치 변경률에 따른 최적의 피치변경법에 관한 연구," 1997년도 음성통신 및 신호처리 워크샵 논문집, Vol. 15, No1 pp. 460-464, 1998년.
5. 조왕래, 이대근, 배명진, 송문섭, 기석철, "피치 검색에 의한 지속시간 변경에 관한 연구," 대한전자공학회, 1995년도 하계종합학술대회 논문집, Vol, 18, No.1, pp.1050-1053, 1995년 10월
6. 김규홍, 김일환, 최광호, 배명진, "부모의 음색을 고려한 동화구연시스템에 관한 연구 - 고음질 합성방식의 선정 -," 한국음향학회, 1996년도 학술발표대회 논문집, Vol. 15, No.1(s), pp. 163-166, 1996년 11월
7. JongDeuk Kim, HyungBin Park, MyungJin Bae "On A Design of the Fairy Tale Narration System with Parent's Voice Color", Proceedings of ITC-CSCC'98, July 13-15, Sokcho, Korea, pp. 673-676
8. 박원, 박형빈, 배명진, "부모의 음색을 고려한 동화구연 시스템의 연구(실시간 처리 알고리즘에 관한 연구)", 한국통신학회 하계종합학술발표대회 논문집(하), Vol.18, No.1, pp.495-498, 1999년 7월 8-10일.