



교육 및 연구용 한국어 TTS 플랫폼 개발

2004.5.8
울산대학교
이정철



목차

- ◆ 개발 목적
- ◆ 필요성
- ◆ 사례조사
- ◆ 개발 방향
- ◆ 요구기능
- ◆ 연구추진 일정
- ◆ 추진체계



개발 목적

◆ 목적:

교육 및 연구용 한국어 TTS 플랫폼을 개발하여 대학의 음성합성 연구 활성화와 산학연 연구결과 교류에 기여함.

◆ 목표

- 한국어 TTS 기본 플랫폼 설계 및 구현
- 플랫폼 모듈 및 API에 대한 문서 작성
- 음성합성 교육에 적합한 tool 제작
- TTS관련 새로운 알고리즘의 수용을 위한 확장성, 유연성 구현
- 산학연 연구결과 교류의 공동 플랫폼으로 활용

◆ 기간: 2004.5.1. - 2005.2.28. (10개월)

◆ 연구비: SITEC



개발 필요성

◆ 음성합성분야 교육 및 연구의 문제점

- 기존에 공개된 한국어 음성합성기 플랫폼이 거의 없음
- 독자적인 알고리즘의 이식 및 검증이 어려움
- 음성합성기 개발을 위한 작업량이 많아 음성합성 연구개발 분야의 신규 진입이 어려움
- 대학 및 연구소 별로 음성합성 연구결과를 발표하고 있으나 공통의 플랫폼을 사용하지 않아서 성능향상 결과의 검증이 어려움

◆ 공통 플랫폼의 필요성

- 합성기 기본 모듈과 문서를 제공함으로써 합성관련 기본지식의 이해와 연구에 기여
- 확장성 및 유연성을 고려한 모듈 설계로 다양한 알고리즘의 이식과 검증이 가능하여 새로운 알고리즘의 구현을 용이하게 함
- 공통 개발 도구는 각 분야의 연구역량을 핵심기술 개발에만 집중할 수 있도록 함
- 공동 플랫폼 사용으로 각 분야의 연구결과 공유가 원활
- 국내 공동연구 그룹간의 공동 엔진으로 사용 가능함



사례 조사

◆ The Festival Speech Synthesis System

- General multi-lingual speech synthesis system developed at CSTR (The Centre for Speech Technology Research University of Edinburgh)
- Full TTS system with various APIs, as well an environment for development and research of speech synthesis techniques
- Version 1.4.3 (January 2003) is the most recent version available free for any purpose, commercial or non-commercial alike (Ver.1.1.1, Jan.1997)
- Written in C++ with a Scheme-based command interpreter for general control
- Sponsorship The Engineering and Physical Science Research Council (EPSRC grant GR/K54229), Sun Microsystems, AT&T Labs -- Research, and BT Labs



사례 조사

◆ Festival Features include:

- English (British and American), Spanish and Welsh text to speech
- Externally configurable language independent modules:
 - ◆ Phone sets
 - ◆ lexicons
 - ◆ letter-to-sound rules
 - ◆ tokenizing
 - ◆ part of speech tagging
 - ◆ intonation and duration
- Waveform synthesizers
 - ◆ diphone based: residual excited LPC (and PSOLA not for distribution)
 - ◆ MBFCLA database support
 - ◆ distributed under a free X11-type licence
 - ◆ generalization of stats modules, ngram, CART, wfst with viterbi so they can be shared more easily
 - ◆ Initial JSAPI support
 - ◆ XML load for Relations
- Portable (Unix) distribution
- On-line documentation
- SABLE markup, Emacs, client/server (including Java), scripting interfaces



사례 조사

◆ Festival distribution includes:

- Full English (British and American) TTS
- Full C++ source for modules, SIOD interpreter, and Scheme library
- Lexicons based on CMULEX and OALD (distributed with permission)
- Edinburgh Speech Tools, low level C++ library
- British English diphone database (for residual LPC resynthesis) (8k (2.1M) and 16k (7.5M))
- 2 American English diphone databases (for residual LPC resynthesis) (8k (3.0M) and 16k (5.6M))
- 4 voices (1 British male, 2 American male and 1 American female) using the MBROLA system. (The mbrola program and voice databases are not distributed from this site).
- Castilian Spanish diphone database (for residual LPC resynthesis)
- Full documentation (html, postscript and GNU info format)



사례 조사

◆ MBROLA projects :

- Towards a Freely Available Multilingual Speech Synthesizer
- Initiated by the TCTS Lab of the Faculté Polytechnique de Mons (Belgium)
- The ultimate goal is to boost academic research on speech synthesis, and particularly on prosody generation in TTS
- a speech synthesizer based on the concatenation of diphones.
 - ◆ Input: phonemes list, prosodic information (duration of phonemes and a piecewise linear description of pitch)
- NOT a TTS (not accept raw text as input)
- Diphone databases tailored to the Mbrola format
- Incite other research labs or companies to share their diphone databases



사례 조사

◆ Non-commercial TTS systems and components compatible with MBROLA :

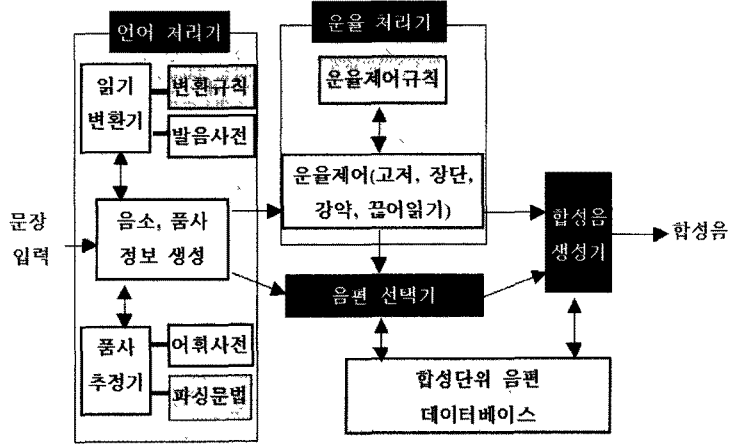
Language	Definition	Provider	
French	eLute, Online TTS	Multitel ASBL	http://www.multitel.be/TTS/eLute_french_demo.html
French, Arabic, English, Dutch, Spanish, Turkish	EULER, Freely available (GNU C++) TTS for Win95/98/NT	Multitel-TCTS LAB	http://cts.fims.ac.be/cyr-buss/euler
English, Welsh, Spanish	FESTIVAL, Multilingual TTS (can be connected to MBROLA)	University of Edinburgh	http://www.cstr.ed.ac.uk/projects/festival/ http://festival.org/
French, Arabic, English, Dutch, Spanish, Turkish	MERDICO, Multilingual talking dictionaries	Multitel-TCTS LAB	http://cts.fims.ac.be/cyr-buss/merdico/
German	Txt2Pho, TTS binary, (Linux, SUN SOLARIS, Win)	Institut für Kommunikationsforschung und Phonetik	http://www.ky.uni-korn.de/de/forsch/chonetik/haefl/HAEFLI/forMBROLA.html
English	TexTalk, TTS	Ben Wooler	http://cts.fims.ac.be/pub/mbrola/US/Engzsr/txtalk.zip
English	Freesepeech, TTS	Stephen Isard, Alistair Cooke	http://ics.fymc.ac.be/pub/mbrola/US/Engzsr/Free_s10r_talk1.zip http://www.sntl.com/col/cndr/wrgrts/wrgrts2.zip



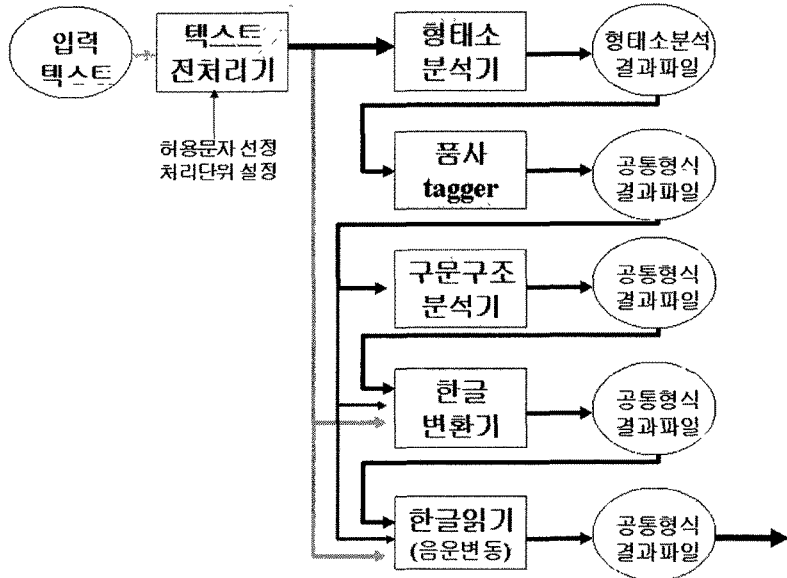
개발 방향

- ◆ 교육훈련 측면
 - 기능 모듈별 실행
 - 알고리즘, 프로그램 설명 문서화
 - 음성합성 교육에 적합한 tool 제작
 - 기존 제품과 비슷한 성능의 TTS 플랫폼을 사용함으로써 교육의 효율성 향상 및 저변 확대
- ◆ 소프트웨어 공학 측면
 - 재사용이 쉬운 프로그램 구조
 - ◆ 객체 단위 모듈 설계
 - ◆ 기능 모듈의 단독 실행을 위한 함수 지원
 - 모듈 구조(쉽게 다른 모듈로 치환 가능)
 - 일관성 있는 프로그램 인터페이스 구조
 - C/C++ 표준 언어 사용
- ◆ 산업적 측면
 - 공통 플랫폼에서 검증된 우수기술을 제품에 적용함으로써 경쟁력 강화
 - 적용분야: Multi-media, Server, Embedded

구조



요구 기능 (언어처리기)

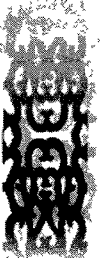




요구 기능 (언어처리기)

◆ Input text

- Input
 - ◆ keyboard *
 - ◆ file plane text -, e-mail text, HTML(XML) text, * hwp, *.doc
- 한글 코드
 - ◆ 조합형 코드 *
 - ◆ KS5601 기본 코드 + 확장형 코드 *
 - ◆ Unicode *
- 한자 *
- 숫자 *
- 외국어
- 심볼 *



요구 기능 (언어처리기)

◆ Text preprocessing *

- text cleaning 허용되지 않는 code 삭제 및 무의미 문자 제거
- Processing unit selection
 - ◆ syllable
 - ◆ word
 - ◆ sentence : 문장경계 검출 (구두점, new line)

◆ Morpheme analysis *

- tag set <운율 경계 추정 규칙과 연관>
- tagged corpus <형태소 사전 및 tagger와 연관>
- 형태소 사전 <tagged corpus → 형태소 사전 작성 tool>
- 형태소 분석 결과 출력 형식

◆ 품사 tagger

- 규칙 기반
- 통계 기반의 HMM (N-gram) *
- 개념망 기반 (의미구분, 장단 구분, 복합명사 의미 경계)
- 품사 tagging 결과 출력 형식



요구 기능 (언어처리기)

- ◆ 구문구조 분석 (parser)
 - Chart parsing
 - LR parsing
 - Link parsing
 - Example based parsing
- ◆ Text normalization (한글 변환기)
 - 숫자 *, 한자 *
 - 특수 기호 (-, ~) *
 - Abbreviations *
 - Acronyms *
 - 외국어
 - ◆ 영어 알파벳, 단어 *, 문장 (품사추정, 발음사전, 엑센트, 운율구)
 - ◆ 일어 알파벳 *, 단어, 문장 (어절경계, 품사추정, 발음사전)
 - ◆ 중국어 알파벳, 단어, 문장 (어절경계, 품사추정, 발음사전)
 - ◆ 기타

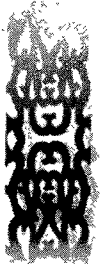
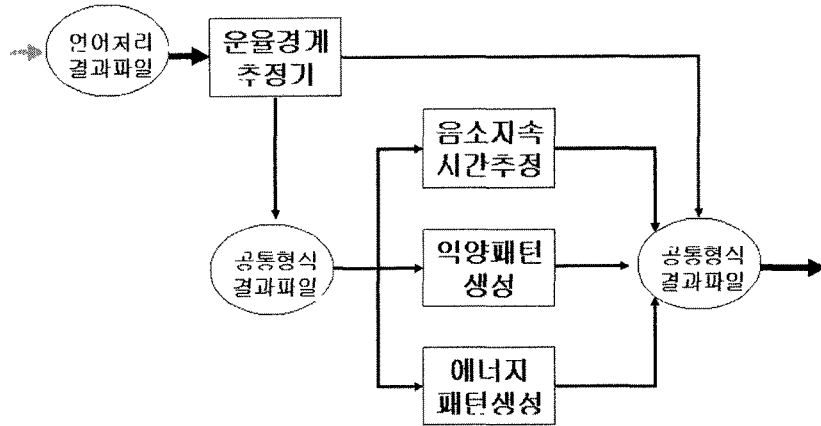


요구 기능 (언어처리기)

- ◆ 음운변동규칙 *
 - 음운 첨가 및 탈락
 - 구개음화
 - 귀착 (중성법칙)
 - 연음법칙
 - 경음화
 - 자음점변
 - 임의 변동
- ◆ 형태소 경계, 단어경계에서 Blocking *
- ◆ POS 정보 활용*



요구 기능 (운율처리기)



요구 기능 (운율처리기)

◆ 운율경계 추정

- 운율 경계 유형 설정
 - ◆ syllable, morpheme, word, sentence, paragraph
 - ◆ accentual phrase
 - ◆ intonational phrase
- 운율 경계 유형 추정
 - ◆ 문장의 구둑점 이용법
 - ◆ POS sequence based N-gram통계 이용법
 - ◆ Syntactic structure based top-down approach
 - ◆ rule based approach*
 - ◆ POS sequence based NN learning
 - ◆ 품사열과 HMM기반 추정법
 - ◆ CART (Classification and Regression Tree)*
- Pause insertion
 - ◆ boundary *
 - ◆ breath group



요구 기능 (운율처리기)

◆ F0 contour 생성

• F0 패턴

- ◆ 기준역양: 점감
- ◆ 역양패턴의 변화: 변화방향, 변화율, 변화폭, 변화위치: early/middle/late
- ◆ Sequential constraints
- ◆ Intonational pattern

• F0 contour 생성 모델

- ◆ Rule based approach
 - O'Shaughnessy model: Two reference lines (topline, baseline)
 - Hat model
 - Pulley model
 - Fujisaki model
 - Pierrehumbert model
- ◆ Data driven approach
 - ToBI model
 - CART model



요구 기능 (운율처리기)

◆ 음소 지속시간 추정

• Factors that influence phoneme durations:

- ◆ intrinsic phoneme duration
- ◆ phonetic context
- ◆ stress level
- ◆ position within the word
- ◆ syntactic structure
- ◆ content word vs. function word
- ◆ etc.

• 음소 지속시간 추정 모델

- ◆ Klatt's duration model
- ◆ Lehiste: Degree of lengthening
- ◆ Fant : Duration index
- ◆ Campbell, Wightman : Normalized Dur.
- ◆ CART*



요구 기능 (음편 선택기)

◆ 합성단위 설계

- 이음소 (diphone)*
- 반음절 (demisyllable)
- 음소 (phoneme)*
- 3음소 (triphone)
- 4음소 (tetraphone)
- 비정형 합성단위 (non-uniform unit)

◆ 최적 후보 선택

- 단일 후보
- Large corpus 기반 복수 후보: *
 - ◆ node cost parameter set, attribute set, cost function
 - ◆ transition cost: parameter set, cost function



요구 기능 (합성음 생성기)

◆ 합성기

- 포먼트 합성기
- 조음기관 합성기
- 음편접합 합성기
 - ◆ Parametric concatenation
 - ◆ PSOLA
 - ◆ Resequencing*
- 속도 가변
- 피치 가변
- 음색 변환



요구 기능 (합성DB 제작 tool)

- ◆ Auto labeling tool
 - HTK 이용
- ◆ hand labeling tool
 - waveform, spectrogram, label marking
 - speech corpus
- ◆ DB construction tool
 - Symbolic transcription
 - ◆ Phonetic and prosodic matching
 - ◆ Initially based on language study (bottom up)
 - ◆ Refined by using prototype system (top down)
 - ◆ Includes segmental (e.g. phonetic transcription) and supra-segmental (e.g. stress, sentence type) information.
 - Acoustic transcription
 - ◆ Prosodic matching
 - ◆ Prosodic continuity
 - ◆ Segmental continuity
 - ◆ Includes pitch, pitch slope, duration, spectral measure



개발추진 일정 계획

- ◆ 1차년도
 - 기존 제품, 소프트웨어 기능 및 구조 분석
 - 기능 설계
 - 시스템 구조 선계
 - 모듈 정의
 - 모듈 입출력 규격 (API) 작성
 - 핵심 모듈(*표시) 구현
 - 검증
- ◆ 2차년도
 - 부가 기능 구현



추진 체계

