

LG TTS의 특징 및
합성음 평가를 위한 제안

이윤근

LG종합기술원 정보기술연구소

LG TTS의 특징

1. 전처리

- 형태소 분석
- Text-to-phoneme conversion : 규칙 기반 + 사전처리
- 특수 문자 처리 (기호, 영어, 일본어, 한자어 등) : 사전 처리 + 오토메타

2. 운율 구현 (Break, Intonation, Phoneme duration, Amplitude)

- 소규모 합성기 : 규칙 기반 처리
- 대규모 합성기 : 규칙 기반 + 통계적 처리(CART, HMM)

3. 합성 단위

- 소규모 합성기 : 고정 합성 단위
- 대규모 합성기 : Corpus based units

4. 합성 방식

- PSOLA (또는 Modified-PSOLA)

5. 합성기 size (database 포함)

- 다양한 size의 TTS 보유 : 2Mbytes~300Mbytes

6. 합성 속도

- 500bytes 합성시 500MHz Pentium급 PC에서 약 0.1 ~ 0.3sec 소요 (합성기 version에 따라 차이가 있음)

상용화 실적

- **E-mail** 읽기, 인터넷 음성 읽기, 공장 자동화, **bundle** 등에 응용
 - **PC 통신 (E-mail Reader)** : 데이콤(천리안), Hitel
 - 언론 기관 (기사 낭독) : 문화방송, 한국일보
 - **통신 업체 (E-mail Reader)** : LG텔레콤, SK텔레콤, 하나로통신, 서울이동통신, 나레이동통신
 - **공장 자동화 (경고방송, 메시지)** : LG하니웰, LG 마이크론, 현대 자동차
 - **공공기관** : 대구 달서구청(One-stop 민원서비스), 대구소방본부(경고방송),
국립특수교육원(시각 장애자용 홈페이지 읽기)
 - 기타 : Dialogic Korea, MJL, JCH시스템, 그로리 미디어, LG홈쇼핑(ARS), 아시아인터넷(UMS)

평가를 위한 제안 (1)

1. 평가 방법

1-1. 음질 평가

* 평가 항목 별 MOS 또는 Preference Test

- 평가 항목 : 음색, 발음의 정확도, 끊어 읽기의 정확도, 억양의 자연스러움 등
- Preference Test가 MOS Test에 비하여 정확한 결과를 도출할 수 있다고 생각되나, 평가대상 시스템이 많을 경우 MOS test도 무방함
- 평가자는 합성음을 들어 본 경험이 있는 사람이 적당함

1-2. Application에 필요한 주요 기능 평가

- * 필요 시 평가 항목 추가
- * 각 기능, 사양에 대한 정량적 평가

평가를 위한 제한 (2)

2. 고려할 점

- TTS 시스템이 사용하는 resource를 비슷한 조건으로 맞추어 비교하여야 함.
 - * 엔진의 크기, 합성 알고리즘의 복잡도 (합성 속도) 등
- TTS 시스템의 용도에 따라 resource가 제한되는 경우가 있으므로, 여러 가지의 경쟁 부문으로 나누어 평가, 음질 평가 외에 상용화에 필요한 평가 항목 추가.

<Examples>

* CTI 서버용 TTS

- PC상에서 동작하므로 메모리 또는 하드디스크 용량에 대한 제한은 완화.
(수백 MBytes의 메모리, 수 GBytes의 하드 디스크 정도)
- 멀티 채널을 실시간으로 지원해야 하므로 합성 속도에 대한 평가가 중요함.
(예를 들어, 합성 속도가 2배이면 사용자는 같은 수의 채널을 지원하기 위해 2배의 시설(서버) 투자가 필요함)

* Embedded용 TTS :

- 소 용량의 메모리만을 사용하여야 함. (1Mbyte, 2Mbyte, 4Mbyte 등)
- 합성 속도는 상용 DSP(20 ~ 30 MIPS)에서 단일 채널 실시간 합성이 가능한 정도.