

일본의 합성음 품질 평가 동향

이용주(yjlee@wonnms.wonkwang.ac.kr)

원광대학교 컴퓨터정보통신공학부

개요

- 합성음 평가의 목적
 - 사용자 측면 - 객관적 평가 근거를 이용한 최상의 제품 선택
 - 개발자 측면 - 시스템을 구성하는 각 모듈의 진단 결과로 활용
- 합성음의 품질 척도
 - 전달하고자 하는 메시지의 내용을 정확하게 전달할 수 있어야 함.
 - 부자연스러움에 의해 문맥이해 오류나 거부감이 없어야 함.
- 평가항목 - 언어학적 단위별
 - 명료도 평가
 - 음절단위 명료도
 - 단어 및 구 단위 명료도
 - 문장단위 명료도
 - 자연성 평가
 - 시스템 전체 평가

일본 JEIDA의 합성음 평가법 연구

- 1991 : 합성기 평가법에 대한 연구 시작 및 사례조사
- 1992 : 합성기 평가법에 대한 초기 버전 발표
- 1993 : 합성기 평가법에 대한 초안 발표
- 1994 : 텍스트 분석 모듈의 성능 평가법 추가

명료도 평가 - 음절단위(1)

- 일본어의 음절 특성
 - 기본 단음절 수 : 100 (200: foreign & historical 단음절 포함)
 - 상대적으로 적은 수의 단음절을 가짐
 - 단음절 평가
 - 조음효과에 의해 단어 초 평가만 유용.
 - 단어 중, 단어 말 음절은 조음효과가 내재됨
 - 2음절 단위 평가
 - 단음절 평가의 단점 극복
 - 그러나 조합의 수가 10000개 : 전체 조합 테스트가 어려움
- 음절단위 평가를 위한 3 레벨의 평가 단위 정의
 - level 1 : 단순평가 목적
 - 고유 100 단음절 + 200단음절(외국어 음절 포함) + CVCV 음절 로 구성된 평가셋
 - level 2 : 진단평가 목적
 - set1 (고유 100 단음절)
 - set2 (외국어 단음절 포함한200 단음절 + 12 고유 단음절)
 - set3 (/{alilu}+CV/ 로 구성된 2음절 평가셋)
 - level 3: 진단평가 목적
 - 일본어 PBW에 준한 고빈도 음소열 셋

명료도 평가 - 음절단위(2)

- 평가실험
 - 평가척도
 - 평균조음점수(Average articulation score)
 - 무의미 음절열 듣기 평가 및 정량적 어려움
 - 음소단위 오인식 Confusion matrix
 - 오인식 음절위치 : 2음절 평가의 경우
 - 평가환경 구성 및 리포트 항목
 - 평가자 인적사항
 - 성별, 연령, 실험 경험 여부, 청취능력
 - 청취방법
 - 스피커, 헤드폰(단귀/양귀)
 - 실험실 환경
 - 방음실, 회의실, 일반사무실
 - 스피커나 헤드폰의 출력레벨 및 잡음레벨, 주변잡음의 유형 명시
 - 합성음 생성 방법
 - 합성음 생성에 사용된 합성기의 종류 : 포먼트합성기, 규칙기반 합성기, 연결형합성기 (LPC기반, 파형접합기반)
 - 합성음의 성별
 - 합성음의 합성단위(diphone, demissyllable, ..)
 - 샘플링 주파수
 - 평가자의 응답방법
 - 음절열이나 단어 또는 문장을 청취후 2분법(correct/incorrect)으로 표시하도록 함.

명료도 - 악센트

- 악센트의 중요성
 - 동음이의어는 단어 내 강세위치에 따라 그 의미가 결정되므로, 악센트 오류를 내재한 합성음은 의미전달 오류를 유발함.
 - 일본어의 악센트 특성
 - 각 단어는 고유의 악센트를 가짐
 - 악센트 구에서는 오직 한 음절에만 악센트가 부여됨
- 합성시스템의 악센트 결정 방법
 - 악센트 구 예측 - 악센트 구 범위 예측 오류
 - 악센트 구내 악센트 음절 결정 - 악센트 음절 예측 오류
- 정량적 평가
 - 평가기준 : 부자연스러운 악센트
 - 평가척도
 - 악센트 구 결정 에러율
 - 악센트 구내 악센트 음절 결정 에러율

명료도 평가 - 포우즈

- 포우즈 평가의 문제점
 - 객관적 기준 부재
 - 포우즈는 화자의 발성습관에 종속된다.
 - 포우즈는 화자의 발성속도에 종속된다.
 - 포우즈에 대한 부자연성은 주관적 의견이다.
- 평가척도

(부자연스러운 포우즈 삼입+부자연스러운 포우즈 누락)

포우즈 에러율 =

텍스트 길이

명료도 평가 - 단어단위(1)

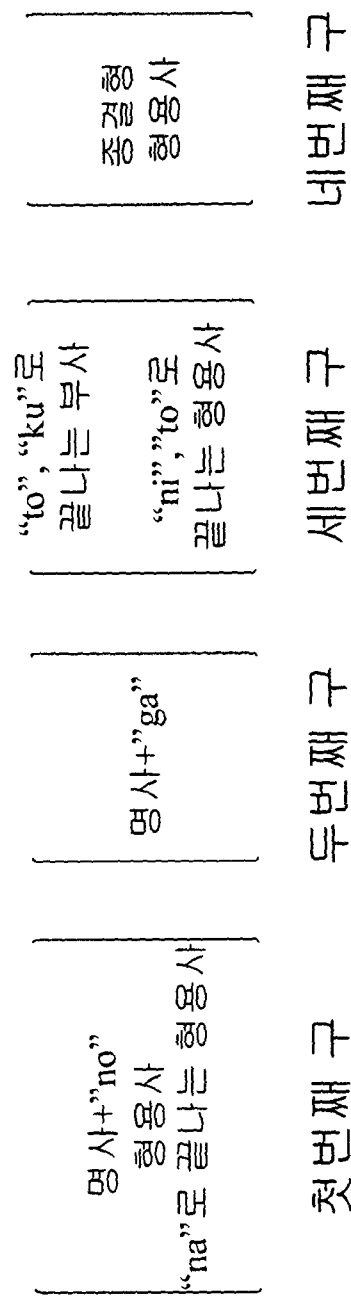
- 단어단위 합성음의 명료도 평가의 중요성
 - 단어는 문장의 의미를 전달하는 기본 단위이므로
 - 접합기반 합성시스템의 합성음질 평가에 유용
- 단어단위 합성품질 평가법
 - 그림단어 단위 평가법
 - 무의미 Carrier sentence를 이용하는 방법

명료도 평가 - 단어단위(2)

- 고립단어 단위 합성음 평가법
 - 평가용 단어 선정 시 고려사항
 - 평가용 단어의 길이
 - Heida(1985)의 실험(인간의 short-term 기억장소에 기억될 수 있는 단어의 최대 길이는 8 mora)에 준해 단어의 길이는 다음 3 부류로 선정.
 - 1) 2~3 moras, 2) 4 moras, 3) 5~8 moras
 - Familiarity(친숙도?)
 - 단어의 Familiarity를 5단계(1:very familiar, 5: very unfamiliar)로 구분하고 이를 다시 3부류로 분할
 - set1(1~2 familiarity), set2(2~3 familiarity), set(3~5 familiarity)
 - 유사도
 - Familiarity에 의해 구별된 각 셋내에서 다시 단어간 문맥 유사도를 고려하여 셋을 구별함('1': 1개의 음소만 다른 단어, '1이상', '0':동음 이의어)
 - 강세
 - 4 mora 이하인 단어에서 강세의 위치에 따라 다음 3가지 유형
 - 0type (강세없음), 1~N-1type, N type(단어의 마지막에 강세가 있음)
 - 단어의 음소 구성
 - 1st mora(7 categories), 2th mora(7 categories + syllabic nasal, double consonants)
 - 선정된 단어의 대표성

명료도 평가 - 단어단위(3)

- 무의미 Carrier sentence를 이용한 단어단위 합성품질 평가법
 - 문장 내 사용된 단어의 명료도가 떨어져도, 문장의 의미로 단어를 추측할 수 있으므로, 문장 내 단어의 정확한 명료도를 평가하기 위해서는, 무의미 Carrier sentence내에 평가용 단어를 삽입하여 듣기평가를 수행함.
- Carrier sentence의 구성
 - 문장의 구성은 매우 복잡할 수 있으나, Heida의 실험에 의하면 인간의 단기 기억 장소에는 저장될 수 있는 가장 적합한 문장의 길이는 4개 이하의 구로 구성된 문장이다. 이를 근거로 무의미 Carrier sentence는 다음과 같은 문법 구조를 가지며 구(Phrase)간에 지배구조가 다른 5가지 유형을 사용함.



명료도 평가 - 단어단위(3)

- 평가실험
 - 모집단 단어 선정
 - “Shin Meikai Kokugo Jiten(2th)” 사전에서 고빈도 42,385 단어 추출
 - 실험에 사용될 단어목록 선정
 - Watanabe(1988)의 testing word selection algo'에 의해 1200단어 추출
 - 200단어를 1set으로 한 6셋 이용(각 피험자마다 실험시간 20분 이내가 되도록)
 - 청취용 데이터 녹음: 문장 전체 또는 일부
 - 피험자에게 문장의 오해가 없도록 무의미 문장임을 미리 알림
 - carrier sentence에 단어를 “대치”, “삭제”, “삽입”하여 평가용 문장 생성

명료도 평가 - 문장단위(1)

- 문장의 명료도를 저해하는 요인들
 - 문장의 중요부분에 사용된 음절의 명료도가 떨어지는 경우
 - 문장의 중요부분에 사용된 단어의 명료도가 떨어지는 경우
 - 형태소 분석오류로 인한 발음변환 오류 및 운율오류
- 문장의 명료도 평가시 고려해야 할 사항
 - 문장의 난이도(평가용 문장 자체가 읽기 어렵거나 전문지식을 요구해서는 안된다.)
 - 평가에 적합한 문장의 유형이란 ?
 - 평가에 적합한 의문문 및 이에 대한 응답 유형

명료도 평가 - 문장레벨

- 평가용 문장
 - 일반적 뉴스문, 에세이에서 추출한 1분 이내의 짧고 간단한 문장(전문분야 지식이 요구되는 문장은 배제)
 - 지도나 그림 등 텍스트외 정보를 이용하는 경우 평가에 비 음성적 요인이 영향을 미칠 수 있으므로 주의
- 질문과 응답방법

자연성 평가(1)

- 평가항목 및 평가척도
 - 평가항목 설계시 고려사항
 - 명료도, 합성음 음질, 리듬, 억양을 평가할 수 있어야 한다.
 - 실험규모 및 범위를 제한하기 위해 평가항목은 최소화 되어야 한다.
 - 평가항목은 피험자들이 쉽게 이해할 수 있어야 한다.
 - Watanabe(1991)의 Global 평가법에 근거하여 설계된 평가 항목들
 - (A) 다른 음소와 혼동되는 부분이 있는가?
 - (B) 알아듣기에 불분명한 부분이 있는가?
 - (C) 음소간 연결은 충분히 부드럽고 자연스러운가?
 - (D) 잠음처럼 들리는 음성이 있는가?
 - (E) 합성음이 사람의 목소리처럼 느껴지는가?
 - (F) 문장의 리듬은 충분히 자연스러운가?
 - (G) 무음구간의 위치 및 길이는 자연스러운가?
 - (H) 단어내 또는 구(Phrase)내 강세의 위치는 자연스러운가?
 - (I) 문장의 억양은 자연스러운가?
 - (J) 합성음의 전반적인 품질은 어떻게 느껴지는가?

자연성 평가(2)

- 평가항목에 대한 평가자의 응답
 - (A)~(I) 항목 : 자연스러움이나 이해의 정도를 5 단계로 응답
 - 5(매우 자연스러우며, 이해하는데 전혀 어려움 없음)
 - 4(약간 부자연스러우나 이해하는데 어려움 없음)
 - 3(약간 부자연스러우며, 이해를 위해 약간의 노력이 요구됨)
 - 2(매우 부자연스러우며, 이해하기가 약간 어려움)
 - 1(매우 부자연스러우며, 이해하기가 매우 어려움)
 - (J)항목 : 7단계로 응답
 - 7(훌륭하다)
 - 6(매우 좋다)
 - 5(좋다)
 - 4(틀을 만 하다)
 - 3(부족하다)
 - 2(매우 부족하다)
 - 1(도저히 이해하기 어렵다)

자연성 평가(3)

- 평가 실험 방법
 - 실험1 : 동일한 문장으로 전체 항목 평가. 합성문장에 익숙해지는 단점.
 - 실험2 : 각 평가항목에 사용되는 문장은 임의로 선택되며, 선택된 문장은 해당 항목의 평가 완료시까지 반복 청취함.
 - 실험3
- 피험자
 - 이전에 합성음 평가에 참여해 본 경험이 있는 자를 중심으로
- 청취환경
 - 일반 스피커
 - 청취실험실 환경(방음실, 사무실, ..), 스피커와의 거리등을 변환시켜 가면서 청취
 - 헤드폰
 - 양귀 헤드폰/단귀 헤드폰
 - 전화기
 - 다양한 제품의 전화기를 사용함.

자연성 평가(4)

- 평가용 문장
 - 문장의 유형, 지배구조, 리듬, 억양등이 다양한 의미있는 문장을 이용함
 - 상대적으로 긴 문장(50 mora 이상)을 사용

Overall quality Evaluation

- 맨머신 인터페이스 측면에서 평가
 - 사용자의 목적, 운영환경, 활용방법, 인터페이스 요소 등, ...
- 평가실험 방법
 - 평가항목
 - 명료도, 음질, 리듬 및 발생속도, 운율, 선호도, 사용자 목적 부합도, ...
 - 평가용 텍스트를 반복 청취 후 다음 7레벨로 응답
 - 3(very much), 2(much), 1(a little), 0(hard to say which), -1(a little), -2(much), -3(very much)
- 평가자 구성
 - 가능한 한 20명 이상이 평가에 참여토록 하였으며, 부득이 평가자의 수가 20명 이하인 경우, 통계적 신뢰도를 높이기 위해 각 평가자에 대해 여러번 실험함.
- 평가용 텍스트
 - JEIDA에서 권고한 평가용 텍스트 사용. 길이는 보통 30초 분량

- 평가환경 고려사항
 - 시스템 운영 환경
 - 사용자의 시스템 사용목적
 - 청취도구(스피커/헤드폰) 및 출력레벨
 - 스피커와 평가자 사이의 거리
 - 주변 잡음 환경
 - 평가실의 반향정도