# PC를 이용한 합성음성평가용 워크스테이션 구성요소에 관한 고찰

창원대학교 제어계측공학과, 음향및 음성연구그룹
윤재원, 조철우

# Some Considerations about Specifications of PC-based Workstation to Assess Synthetic Speech

Acoustic and Speech Group
Dept. of Control and Instrumentation Engineering
Jae-Won Yoon, Cheol-Woo Jo

요 약

제작된 합성기를 평가하는 데는 많은 노력과 시간이 필요하다. 본 논문에서는 이러한 평가 과정을 정규화 해줄 수 있는 평가용 워크스테이션을 PC를 이용하여 작성하는 데 필요한 요소들을 제안하고 진행중인 구성과정을 제시한다. 합성기 평가용 워크스테이션은 복잡하고 번거로운 합성기 평가과정을 균일하게 유지시켜 줄 수 있으며 단시간에 평가결과를 얻을 수 있게 해 줌으로써 합성기의 평가 및 개발 기간을 단축시키는 데 유용할 것으로 생각된다.

## 1. 서 론

현재 많은 연구실에서 음성 합성 기술을 연구하고 있고 그 기술은 상당한 수준에까지 올라있지만 개발된 합성기의 평가에 관한 연구는 미비한 실정이다.

우리 나라의 음성 평가 현황은 각 연구실마다 음성 합성기를 제작시에 그 합성기를 평가하기 위해 개별적으로 만든 환경 조건 속에서 실시하고 있고 평가 단어군 선정 자체도 합성기 테스트를 위한 공통 표준 단어군이 아닌 단어군을 임의로 정해서 사용하고 있는 실정이다. 또한 평가 방법에 있어서도 M.O.S. 등의 방법을 사용하고 있지만 각 테스트마다 나온 결과를 비교하기가 어렵게 되어 있다. 연구실마다 각기 다른 환경 조건에서 평가를 하다 보니 공통된 파라미터를 찾기 어렵고 횟수 또한 1회 내지 수 회 정도에 불과하고 인원도 일반적으로 피험자의 수가 10 ~ 20명 정도이고 보통 수명 정도인 경우가 많다. 각 연구실마다 그 때마다 필요한 방법을 택해서 평가를 하고 환경 조건과 시험 횟수, 등이 각기 달라 시험 결과치의 신뢰도에 객관성이 결여되어 질 수 밖에 없다. [1]

이에 따라 많은 시간과 열정으로 만든 음성합성기를 평가는데 소요되는 시간과 노력을 줄여줄 수 있고, 잘못된 점을 정확히 짚어 주는 그러한 합성기 테스트 시스템을 요구하게 되었고 외국의 경우 SAM과 같은 개발계획의 부산물로 만들어진 합성기의 test 소프트웨어들이 있는 반면 우리 나라의 경우는 시험을 위한 단어군 조차도 미비한 실정이다. 즉 평가의 기준이 될 수 있는 DB(데이타 베이스)와 전문적인 음성 평가 장치가 요구 되고 있으며, 우리말에 적합한 평가 방법의 개발이 요구된다.

본 논문에서는 PC를 이용한 음성 평가용 워크 스테이션을 구현하기 위한 구성요소들을 고찰하고 전체적인 구성과 설계 시 필요한 블럭 다이어그램을 통해 현재 구현중인 워크스테이션의 전체적인 윤곽을 제시한다.

## 2. 음성 평가용 워크스테이션

합성음의 평가작업은 우선 공통의 평가절차와 데이타 베이스의 확립이 선행되어야 하며 이와 관련된 도구들이 개발되어 규칙합성시스템을 연구하는 팀들에게 제공되어야 한다. [1]

음성 평가 시스템은 음성을 평가한 후 잘못된 음절을 정확히 찾아 내고 신뢰성 있는 인식율을 보여 주어야 한다. 이러한 조건을 만족하기 위해 많은 시험 횟수와 비슷한 환경 조건을 만들어야 한다. 본 연구실에서는 전문적인 음성 평가용 시스템을 제작하고 특정 합성기가 아닌 일반적인 음성 합성기를 평가할 수 있는 시스템을 제작 진행중이다. 그러기 위해서 각 합성기에 대해 객관성 있는 시험을 하고 그에 부과되는 각 합성기의 다양한 파라미터에 대해서 충족시키기 위해 여러 개의 헤더화일 을 만들었다. 현재 세계적으로 사용하고 있는 합성음을 평가하는 방법은 D.R.T. (Diagnostic rhyme test), M.R.T. (Modified rhyme test), M.O.S. (Mean opinion score), D.A.M. (Diagnostic acceptability measure) 등 여러 방법이 있으나 환경 조건이나 parameter의 값들이 시험 방법에 따라 일정한 값을 나타내어 주지는 못하고 있다. 합성음성의 질을 각 음절간의 명료도(intelligibility), 자연성(naturalness), 쾌적성(ease of communication)으로 평가를 하여 보다 정확히 어느 부분이 문제가 있는가를 보여 줄 수 있어야 하고 수정 보완시에 지침이 되어야 한다. [1] [2] [3]

일반적으로 음성합성기를 평가할 경우 녹음테이프의 스피커 또는 방음 시설이 된 시청각실을 빌려야 하고 정답 체점시 거의 손으로 정답과 오답을 찾아내고 사람이 도표에 기입해야 한다. 또 피험자를 제시간에 필요한 수의 사람을 시험 장소에 모두 모으는 것도 쉽지 않고 제차 시험 방법에 대한 교육과 또 다른 테스트를 하는 경우에 다시 피험자를 모으기도 어렵다. 이런 과정은 만약 평가용 워크스테이션이 있다면 많은 문제점을 줄일 수 있게 된다.

이렇게 해서 많은 데이타 베이스를 확보하기도 힘들 뿐더러 한번에 많은 파라미터를 조정해야 하기 때문에 과연 그 평가가 올바른 평가였는가, 객관적으로 행하여 졌는가를 고려해야 할 필요도 있다. 객관적 평가는 엄청난 양의 데이타 베이스가 필요하고 계산 또한 복잡하다. 주관적 평가는 객관적 평가보다는 다소 객관성이 떨어지지만 데이타 베이스의 양이 많아졌을 때 이 평가도 객관성을 가질 것이라고 본다. 본 문에서는 PC를 이용해서 대용량의 데이타 베이스를 관리할 수 있고 녹음도 보다 효율적으로 행할 수 있으며 평가시에도 가능한한 조건들을 종래의 경우에 비해

일관성있게 유지 함으로써 객관성을 유지할 수 있게 할 수 있을 것이다.

본 연구에서 이미 기존에 작성되어 있는 단어군세트 또는 단어발생기에 의한 임의의 단어군세트를 선택가능할 수 있고, 단어의 입력을 KEY board 또는 화일로도 입력이 가능하게 만들었다.

시험용 단어군은 무의미 단어군도 사용할 수 있게 한다. 무의미 단어군은 음소 단위로 무작위 생성해서 의미를 가지는 단어는 제외하고 무의미 단어만 택해서 만든 단어군이다. 그러나 너무 많은 음절을 택할 경우 운율의 영향을 받게 되어 적당한 음절을 선택해야 한다.

무의미 단어군을 사용하는 이유는 우리가 피시험자가 되었을 때 단어를 정확히 듣지 못해도 단어의 뜻을 추측하여 기입하여 생기는 시험의 오류를 줄이기 위해서이다. 통계적으로 구해진 음소나 음소환경의 출현 빈도수를 바탕으로 시험에 사용한 무의미단어를 발생시켜 평가에 사용하며, 평가에 사용되는 음소환경의 수가 합성기 음성에서 발생할 수 있는 여러 음소환경의 경우를 다루기에 충분하다.

다음절 무의미 단어는 같은 수의 무의미 단음절에 비해 많은 종류의 음소와 음소환경을 시험할 수 있다. 그리고 실제 사용하는 단어군보다 청취자의 학습효과가 적어 보다 정확한 평가가 가능하다. [4]

합성 음성을 평가시에 고려해야 할 사항은 첫째 시험 방법의 세부 사항들을 어떤 것인가, 둘째 시험 환경에서 문제되는 점들을 고려해야 한다.

일반적으로 시험 방법에 관하여는 다음과 같은 몇 가지의 세부 사항들을 고려하여야 한다.

◆ 시험 단어 선정 과정의 문제
- 공인된 평가용 단어군은 없다.
- 수동 입력, 자동 발생
(무의미 단어 발생시 유리하다)
- 단음절, 다음절
- 유의미 단어, 무의미 단어

◆ 응답 양식

- OPEN 방식 - 예를 들면 음성을 듣고 받아 적는 것
- CLOSE 방식 - 주어진 틀 속에서 번호를 선택 하는 것

◆ 시험 장소의 소음

- 방음 시설의 유무
- 헤드폰, 스피커
- 시험 장소의 고유 소음

◆ 시험시에 단어 반복 횟수

- 시험관이 정의
- 시험자의 요청시

◆ 문제간의 공백 시간

이상과 같은 문제점들을 평가용 워크스테이션의 구성시 고려하여 자동으로 시험이 진행될 수 있게 해 주어야 한다. [5] [6]

3. 하드웨어적 구성

본 연구에서는 사용한 IBM PC-486 DX-67Mhz에 사운드 카드로는 Sound Blaster 16과 연결하여 음성 자료의 녹음과 평가시에 사용하고 있다. 녹음 파일은 INKEL YM-1500 마이크를 사용하여 볼랜드 C++로 만든 프로그램에서 생성된 단어군을 모니터를 보고 녹음한다. 녹음시 샘플링 주파수는 5000~44100Hz, 16비트로 녹음을 한다. 녹음시에 생기는 여러 파라미터들을 헤더 화일에 기록하여두고 통계시에 적절히 이용할 수 있도록 한다. 청취실험에서는 헤드폰을 사용하여 청취할 수 있게 하고 입력은 키보드나 마우스를 통해 행할 수 있다.

4. 소프트웨어적 구성

본 연구실에서는 PC를 이용한 전문적인 음성 평가 프로그램이기에 피험자가 시험관이 없어도 시험이 가능한 방향할 수 있도록 구성한다. 볼랜드 C++로 작성되고 있는 프로그램은 PULL DOWN 방식으로 개발 중이다. 평가의 진행 방식은 그림 1 과 같은 형식으로 진행 될 것이다.

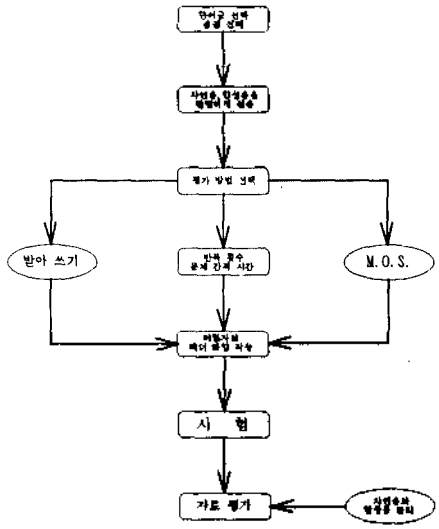


그림 1. 평가과정의 흐름도

메뉴가 화면에 뜨면 평가 방법과 반복 횟수, 문제 제시 간격 등을 KEY로 사용하여 선택하면 된다.

시험 방법은 시험자가 헤더 화일을 작성하면서부터 시작되고 앞에서 열거한 여러 변수들이 기록되어진다.

대부분의 합성기 평가에서 M.O.S.을 사용하여 자연성 판별을 하고 있는 반면 본 연구에서의 장점은 한 합성기를 가지고 OPEN방식과 CLOSE방식을 선택하여 평가할 수 있다는 것이다.

시험 방법에 있어서 두가지의 방법을 사용하는데, 스

피커로 음성을 듣고 판아 적는 방법과 스피커로 듣고 음성의 질을 5개 또는 입의의 개수의 보기가 제시하고 있는 것 중 하나를 선택하는 방법을 택했다. 첫번째 방법은 명료도를 판정하는데 용이하고 두번째 방법은 음성의 자연성을 척도하는데 용이하다. 시험자의 집중도를 파악하기 위해서 자연음과 합성음을 랜덤하게 섞어서 들려주고 그 비율은 사용자가 정할 수 있어야 한다. 사험자의 집중도를 파악하기 위해서 자연음과 합성음을 랜덤하게 섞어서 들려주고 그 비율을 사용자가 정할 수 있어야 한다. 이 단계는 평가방법을 선택하기 전 단계에서 만들어진 녹음 화일에서 시험관이 지정하여 주는 비율로 랜덤하게 섞어서 피험자에게 들려주고 평가 시험이 끝난 후에 자연음과 합성음을 분리시켜 통계를 내기 때문에 편리하게 사용할 수 있을 것이다.

시험 자료의 통계치는 필요한 데이타들을 불러서 통계를 낼 수 있을 것이다.

음성 자료 녹음의 순서는 그림 2 와 같다.

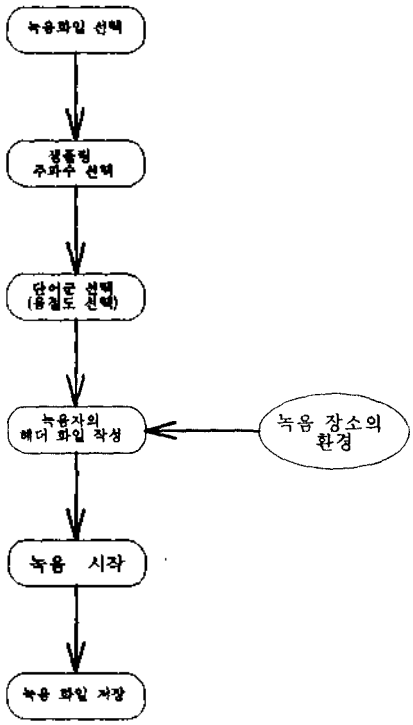


그림 2. 녹음과정의 흐름도

자연음을 녹음 할 것인가, 합성음을 녹음할 것인가를 선택한 후 샘플링 주파수와 단어군을 선택하게 된다. 이 때 음절수도 선택하게 되고 녹음자의 헤더 화일도 작성 한다. 이때 직접 합성기를 사용하지 않고 녹음 테이프를 사용할 경우도 있을 것이고 그럴 경우에는 테이프를 PC에 사운드 카드를 이용하여 화일로 저장한 후 사용할 수도 있다.

PC에서 녹음을 하고 평가를 함으로 인해 기억 용량이 큰 녹음화일을 처리하기 좋고 화일로 보관하기 때문에 이동과 보관이 용이할 뿐만 아니라 대용량의 데이타를 손쉽고도 오차 없이 계산이 가능하다.

본 연구에서 효과적인 데이타의 관리를 위해서 4가지 종류의 헤더화일을 사용하고 있다. 헤더화일은 단음절 생성시와 다음절 생성시, 시험 평가시 ,음성 자료 편집시로 세분화하여 데이타 처리가 편리하게 구성하였고 다음과 같다.

단음절 헤더화일
1. 단어 개수
2. 생성 시간

다음절 헤더화일
1. 단어 개수
2. 생성 시간
3. 음절 수

평가 헤더화일
1. 시험일시
2. 장소
(소음 정도에 따라 1 ~ 5단계)
3. 이름
4. 나이
5. 15세 이전의 고향
6. 직업
7. 학력
8. 경험 회수
9. 시험관
10. 시험 방법
11. 데이타 번호
12. 문항 수
13. 문제 제시 간격 시간

음성 자료 헤더화일
1. 합성음, 자연음 구분
2. 녹음 장소
3. 녹음 장소의 소음 정도
4. 녹음 시간
5. 녹음 방법
6. 녹음자의 성별
7. 녹음자의 나이
8. 녹음자의 이름
9. 단음절, 다음절
10. 녹음 단어 갯수

단음절 헤더화일은 발생된 단어의 갯수와 생성시간을 기재하게 되어 있다.

다음절 헤더화일은 발생된 단어의 갯수, 생성시간과 음절의 수를 기재하게 되어 있다.

평가 헤더화일은 피험자가 인적 사항을 작성한 후 평가의 방법과 시험 번호, 문항 수와 시험관까지 기입하게 되어 있다.

헤더화일에서 3 - 8 까지가 피험자의 인적 사항인데 5번 항의 15세 이전의 고향을 기입하는 부분이 있다. 이것은 대부분의 사람이 고향의 사투리를 그대로 쓰고 그것이 평가에 영향을 줄 수도 있기 때문이다. 7

번 항도 통계 시에 고려해 주어야 할 사항이다. 학력이 음성 인식에 영향을 줄 수 있고 8번 항의 경험 회수 또한 합성음을 자주 들어봄으로 인하여 생기는 인식성에 차이도 고려하여야 할 것이다.

평가부분에서는 개방형 시험의 경우는 각 음소별 정오율, 혼돈율, 자연음과 합성음의 비교인식율등이 측정되며 폐쇄형 시험의 경우는 자연성이 측정되고 각 문항별 통계자료가 시험이 끝남과 동시에 출력될 수 있다.

## 5. 결 론

본 논문은 음성합성기를 보다 객관성 있게 평가하기 위해 필요한 PC를 이용한 워크스테이션을 구성하는데 있어서 시스템을 구현할 때 필요한 파라미터에 대해 고찰하고 구현중인 시스템의 전체적인 구조에 관하여 소개 했다. 본 시스템이 구현된다면 지금까지 시간이 걸리고 번거로운 과정이었던 합성기의 평가과정을 보다 쉽게 수행할 수 있게 될 것이다.

외국의 경우 연구소에서 합성기 테스트를 위한 시험단어군을 만들어 사용하고 있지만 우리 말의 단어군 DB는 아직 개발 되어 있지 않은 실정이다. 이 또한 시급히 개발되어야만 보다 정확한 음성 평가가 이루어질 수 있고 그럼으로 인해 양질의 음성합성기 또한 개발될 수가 있을 것이다.

## 6. 참고문헌


1. 조철우 , '규칙합성음의 평가 및 진단법에 관하여' , 昌原大學校 産業技術研究所 論文集 第 8 輯 (1994)

2. 김정환 외 , '명료도 평가용 단음절 목록의 개발' , 韓國音響學會誌 第 13 券 第 4 號 (1994)

3. Jekosch ,' Speech Intelligibility Testing : On the Interpretation of Results .'sub-mitted to Journal of the American Voice I/O society ,vol 15, pp.63-80, U. (1994)

4. 조철우, 김정태 외 , '합성음성평가를 위한 다음절 무의미단어 생성과 이용에 관한 연구', 韓國音響學會誌 第 13 卷 第 5 號 (1994)

5. Ute Jekosch & Louis C.W. Pols ,'A FEATURE-PROFILE FOR APPLICATION-SPECIFIC SPEECH SYNTHESIS ASSESSMENT AND EVALUATION', ICSLP'94, Yokohama, (1994)

6. Thomas Hegehofer , 'A DESCRIPTION MODEL FOR SPEECH ASSESSMENT TESTS WITH SUBJECTS', ICSLP '94 , Yokohama, (1994)

7. Jens Blauert, Thomas Hegehofer, Ute Jekosch, Arnd Mariniak,' An Adaptive For The Assessment Of Speech Intelligibility ', EUROSPPECH '91., (1991)