

음성데이터 수집을 위한 발성내용 제시시스템

Utterance display system for speech data acquisition

김 경 태*, 이 용 주**, 정 유 현**

(Kyung Tae KIM*, Yong-Ju LEE**, Yu Hyeon JUNG**)

요 약

본 논문은 발성자의 자연스러운 음성데이터를 수집하기 위한 발성내용 제시시스템의 구현에 대하여 기술한다. 대량의 음성정보의 수집 및 처리를 위해서는 이와같은 시스템이 필수적이다. 왜냐하면, 음성정보처리의 성능 평가는 음성데이터와 발성방법에 따라 좌우되므로 실제의 환경에서 사용되는 자연스러운 음성으로 평가되어야만 객관적인 결과를 얻을 수 있기 때문이다. 따라서 이러한 음성데이터를 효율적으로 수집하기 위한 방법으로써 발성내용 제시 시스템에 관하여 기술하고자 한다. 특히, 본 논문에서는 발성해야 할 데이터를 제시하기 위한 요구사항, 기능, PC에 의한 구현에 대하여 기술한다. 본 시스템은 음성수집 단계뿐만 아니라 수집 후의 편집 작업의 편리성을 고려하여 구현하였으며, 4연속 숫자음 등 96명이 발성한 63,840개의 단어를 수집하는데 적용하였고 수집 과정에서 종래의 리스트를 보고 발성하는 방법에 비해 훨씬 효율적이고 자연스러운 발성을 유도할 수 있었다.

Abstracts

This paper describes the implementation of utterance display system for natural speech data acquisition with individual utterance habits. This system is considered important for very large-scale speech signal processing since the performance of speech signal processing depends on the speech data as well as the way of utterance, and eventually must be evaluated by the natural speech data obtained from real worlds. Thus, as the method of collecting the speech data, this paper deals with the requirements, the functions, and the implementations based on personal computer (IBM-PC) for displaying the speech data to be uttered. Especially this system was implemented considering not only the acquisition of natural utterance but also the convenience of editing works after data acquisition phase. It was applied for 63,840 words uttered by 96 speakers and, from the view-point of utterance naturalness, showed the improved results better than the list method.

I. 서 론

음성정보처리 분야에서, 음성 데이터베이스 구축의 필요성 및 중요성이 더욱 높아 가고 있다.[1,2,3,

4]. 체계적인 음성연구를 위해서는, 가능하면 많은 종류, 많은 양(성별, 나이, 방언, 화자수 등)의 데이터가 필요하며, 음성연구가 고도화 되고 신뢰성이 있는 연구가 되기위하여는 더욱 절실히 요구되는 사항이다. 이러한 많은 양의 음성을 수집함에 있어서, 가장 중요한 점은 평상시처럼 자연스럽게 발생된 음성 데이터를 수집하는 일이다.

* 한남대학교 정보통신공학과,

** 한국전자통신연구소

* Depts. of Inform. & Comm., Han Nam Univ.,

** ETRI

접수일자: 1992년 10월 6일

일반적으로, 음성을 수집하기 위하여 녹음을 한다고 하면 정상시의 자연스러운 발성이 나오지 않고 긴장하게 되고 조심스럽게 발성이 되어 실제적인 효과를 얻기가 상당히 어렵다. 따라서, 어떻게 하면 보통 때의 발성과 똑같이 발성된 음성을 얻을 수 있을지가 중요한 연구의 대상이 된다. 즉, 음성데이터 수집을 위한 발성내용 제시 방법에 대한 연구이다. 보통 생각될 수 있는 방법으로는, 첫번째, 발성단어 혹은 문장의 리스트를 만들어 두고 발성자로 하여금 읽게 하는 방법이다. 이는 가장 간단한 방법이지만 녹음 중에 종이를 넘기는 소리와 같이 불필요한 잡음이라든가 여러번 발성을 할 때 몇번째 발성인지, 혹은 방금 발성한 단어가 어느 단어인지 잊어버리기 쉬운 단점이 있다. 두번째, 발성해야 할 단어를 헤드폰을 통하여 음성으로써 알려주는 방법이다. 이는 리스트를 보고 발성하는 방법보다는 좋지만 들고 발성하므로써 발성자 나름대로 발성하기보다는 억양, 발성방법, 소리의 크기 등이 헤드폰에서 나오는대로 발성하려는 경향이 있어 개인성이 상실될 염려가 있다. 세번째, 발성해야 할 대상단어, 또는 문장을 개인용 컴퓨터 등의 화면에 디스플레이 시켜 발성자가 읽게 하는 방법이다. 이 방법은 발성자에게 발성에 대한 불안과 부담을 덜어 주며, 발성 등의 잘못이 있을 때는 미리 구현되어 있는 컴퓨터의 제어기능을 이용하여 수정작업이 쉽게 이루어지는 장점이 있다. 또한 발성 시간, 나아가서는 발성의 크기 등을 적절히 조정할 수 있어 매우 좋은 방법이다. 그러나, 이는 발성 데이터들을 화면에 디스플레이 시키고 발성자와 운영자를 위한 각종 기능을 제어할 수 있는 시스템의 구현에 관한 연구가 필요하다. 네번째, 가장 좋은 방법은, 화면 등을 이용하여 원하는 발성이 나오도록 유도 질문하는 방법이다. 즉, 화면상에서 원하는 대담(발성)이 나올 수 있도록 질문하는 방법이다.[5] 지금까지는 주로 첫번째 방법인 발성리스트를 이용하는 것이 가장 보편적이고, 세번째의 방법도 시스템의 구현문제 때문에 널리 이용되고 있지 않은 실정이다.

본 논문은 많은 양의 자연스러운 발성 데이터를 수집하고 효율적인 편집을 할 수 있으며, 신뢰도 높은 음성 데이터베이스를 구축함을 목적으로 발성내용 제시시스템의 구현에 대한 연구로써, 먼저 고려해야 할 발성내용 제시시스템의 요구사항을 정의하고, 범용 시스템으로써의 기능과 구현에 대하여 기술한다. 마지막으로 본 시스템을 4연속 숫자음 35개, 숫자음 25개를 각각 96명이 9회씩 발성한 데이터와, 계산기

제어용 40개단어를 60명이 5회씩 발성한 데이터를 수집하고 레이블링하는 데 적용하였다. 그 결과, 발성자의 자연스러운 발성의 유도화, 운영자의 작업 효율성이 향상된 결과를 얻었다.

II. 시스템의 요구사항

발성자는 편하게 발성할 수 있고, 발성을 모니터링하고 시스템을 운영하는 운영자의 작업효율을 향상시키기 위하여 발성내용 제시시스템의 요구사항을 다음과 같이 정의한다.

첫째, 부담없이 평소의 발성대로 긴장하지 않고 자연스럽게 발성할 수 있도록 발성자를 도와주어야 한다. 둘째, 발성자의 발성을 모니터링하는 운영자의 작업량이 상당히 많기 때문에 작업의 부담을 줄여야 한다. 셋째, 음성데이터 수집이 끝나고 정확하고 편리하게 데이터의 편집작업을 할 수 있도록 보조정보를 자동으로 제공할 수 있어야 한다. 넷째, 시스템은 범용의 개인용 컴퓨터에 의해 구현되어야 하며 소프트웨어의 확장과 호환이 쉬워야 하고, 시스템이 동작하기 위하여 필요한 소프트웨어, 하드웨어의 지원을 최대한 줄여 기본 시스템으로 동작될 수 있어야 한다. 다섯째, 화면에 발성해야 할 내용이 디스플레이 될때 디스플레이 되기까지 걸리는 시간이 가능한 짧아야 한다. 즉, 발성 내용의 글씨가 스캐닝되어 나타나는 것처럼 보이는 형태가 아니라 시작과 끝부분이 동시에 디스플레이 되어야 한다. 여섯째, 화면에 디스플레이 되는 발성내용은 운영자가 원하면 언제든지 내용의 첨가, 수정, 삭제할 수 있어야 하고 변경된 사항은 쉽게 확인할 수 있어야 한다. 일곱째, 운영자는 컴퓨터를 전혀 모르는 사람으로 가정하여 화면에 나오는 지시만을 보고 모든 작동을 할 수 있어야 한다.

III. 발성내용 제시시스템의 기능

앞 절에서 정의한 요구사항을 만족시키기 위한 발성내용 제시시스템의 기능은 크게 나누어, 발성자로 하여금 부담을 느끼지 않고 자연스러운 발성을 유도할 수 있도록 하는 발성자를 위한 기능과 수집과정 중의 운영기능, 음성수집이 끝난 후 음성데이터 편집작업을 정확하고 편리하게 하는 예비작업을 위한 기능이 있다.

3.1 발성자를 위한 기능

발성준비시간과 발성시간을 서로 다른 색깔을 이용하여 시각적인 막대가 채워지는 형태로 시간의 흐름을 시각적으로 나타내어 줌으로써 다음의 일을 예측할 수 있게 하고 화면에 「발성한 단어수/전체 단어수」를 타나내 현재의 진행정도를 알 수 있게하여 심리적인 부담을 줄인다. 이때, 발성시간은 운영자에 의해서도 조정되지만 발성자가 키보드 혹은 마우스를 이용하여 조정할 수 있게 한다. 발성 도중에 언제든 지 쉴 수 있으며, 발성을 시작하기 전에 몇 개의 단어를 발성하고 쉬기를 원하는지 그 단어의 수를 입력함으로써, 원하는 일정 간격으로도 쉴 수 있게 한다.

3.2 운영자를 위한 기능

운영자를 위한 기능에는 크게 두 가지의 기능이 있는데, 즉 발성자가 발성하는 음성데이터를 수집할 때 필요한 기능과 음성수집이 끝난 후 음성 편집작업을 할 때 필요한 기능이 있다.

3.2.1 음성데이터 수집시 필요한 기능

발성 도중에 운영자가 할 수 있는 일들을 화면의 밑부분에 윈도우형 메뉴 형식으로 표시해주고, F1 키를 누르면 HELP 기능이 있어 그때 그때 필요한 기능을 설명해 줌으로써 운영자가 모든 절차를 쉽게 알고 제어할 수 있다.

발성을 시작하기 전에 운영자가 녹음환경, 발성자 인적사항, 운영자 관련 사항 등을 키로써 입력하며, 이 일을 가급적 줄이기 위하여 입력을 한 번 끝내고 발성자가 바뀌어 다시 시작할 때에(처음화면), 그전에 입력된 사항을 그대로 남게 하여 발성자의 이름 등 변동이 있는 부분은 BACKSPACE로써 수정하여 다시 입력하고, 녹음환경 등 변동이 없는 부분은 ENTER 키로써 다시 입력하는 노력을 줄인다. 그리고, 내용을 전체적으로 수정, 변경해야 할 부분이 있으면 상하 화살표키를 이용하여 원하는 장소로 움직여 수정할 수 있게 한다.

발성도중에 발성자가 발성을 잘못했을 경우와 발성을 하지 않았을 때 다시 발성해야 할 경우가 생긴다. 이때, 보통의 방법으로써는 운영자가 발성내용의 번호를 메모지에 메모해 두었다가 일차 발성이 끝나면 잘못된 발성들을 함께 모아 다시 발성하게 한다. 이렇게 하면 발성자와 운영자 모두가 작업의 복잡성 등에 의한 오류가 발생하기 쉽다. 이를 위하여 잘못된 발성내용을 운영자가 시스템에 기억시켜 자동으

로 발성내용을 다시 제시하게 한다. 이 작업은 물론 수동으로도 임의의 발성내용을 제시할 수 있게 한다.

많은 양의 발성내용을 제시하는 순서에 있어서, 한 세트의 단어를 한 번만 제시하는 경우와 여러번 제시해야 할 경우(발성횟수)가 있다. 이때, 여러번 제시하는 방법에 있어서 발성내용의 순서를 각 횟수 내에서 같게 할 수도 있으며, 랜덤으로 각각 다르게 할 수도 있고, 각 횟수 내가 아닌 전체의 발성내용이 랜덤한 순서로 할 수도 있는데 이를 운영자가 선택할 수 있다. 물론, 발성내용을 입력할 때도 운영자가 쉽게 입력하고 수정할 수 있다.

3.2.2 음성데이터 편집시 필요한 기능

음성데이터 수집 작업이 끝난 후 편집작업을 정확하고 편리하게 할 수 있도록 편집 작업에 필요한 화일을 자동 저장하게 하는 기능이 있다. 즉, 음성데이터의 수집작업을 끝내면, 자동으로 두 종류의 화일이 생성된다.

운영자는 발성자가 발성할 내용이 들어 있는 화일명을 입력시키고, 현재의 발성자가 누구인지를 알기 위하여 발성자의 이름을 입력한다. 그리고 발성자의 발성이 끝나면, 음성데이터를 편집할 때 필요한 화일이 생성된다. 이는, 어느 발성자가 어떠한 순서로 된 단어를 발성했는지? 혹은 잘못 발성한 단어가 어느 단어인지를 알아야 편집을 정확하고 쉽게 할 수 있다. 이때, 화일을 두 종류로 나누어 누구의 발성이 어떤 화일명으로 저장되었는지를 알 수 있는 화일과, 발성자의 인적사항과 발성환경, 조건, 그리고 실제로 발성한 내용이 저장되어 있는 화일이 생성된다.

IV. 시스템의 구현과 적용

본 시스템은 크게 두 부분으로 나누어 구현된다. 첫째는, 발성자가 발성할 발성 리스트의 작성 부분과, 둘째는 단어를 화면에 디스플레이 시키고 제어해주는 발성내용 제시부분이다. 본 시스템의 구현과 실행에 필요한 기본환경과 앞에서 기술한 시스템의 요구사항을 만족시키기 위한 기능을 구현하는 방법과 시스템의 적용에 대하여 기술한다.

4.1 기본 환경

시스템은 누구라도 사용할 수 있도록 하는 범용성을 고려하여 특별한 환경은 사용하지 않음을 원칙으로 하였기 때문에 거의 대부분이 범용의 개인용 컴퓨터

환경이다.

4.2 발성 리스트의 작성

발성자가 발성할 발성내용의 리스트를 작성한다. 운영자는, 발성단어의 반복횟수, 반복된 단어들의 나열 방법 선택, 저장될 화일명을 입력한다. 현재의 입력 데이터를 디스플레이 시켜주고, 데이터의 수정, 확인, 편집을 선택할 수 있다. 이러한 절차는, 그때 그때 필요한 사항이 화면에 디스플레이 되기 때문에 컴퓨터에 대한 기본 지식이 없는 운영자라도 쉽게 운영할 수 있다.

표 1은 10개의 4연숫자 2회를 반복횟수내에서 랜덤하게 구성되는 방법으로 작성된 단어리스트를 예로써 나타내고 있다. 이 표에서 알 수 있듯이, 4연속 숫자 "7502"의 경우에 1회째는 두번째 제시되고 (2번)2회째는 여덟번째 제시(18번)됨을 알 수 있다.

표 1. 작성된 데이터 리스트 : 4digit.dat

1. 9568	2. 7502	3. 6737	4. 7254	5. 5625
6. 1451	7. 4656	8. 8245	9. 5625	10. 8653
11. 8653	12. 9568	13. 5626	14. 1451	15. 8245
16. 4564	17. 7254	18. 7052	19. 5625	20. 6737

4.3 발성내용 제시 시스템

발성내용을 제시하고 제어하는 기능을 가진 시스템을 구현한다. 발성자는 부담없이 자연스럽게 발성하고, 운영자는 쉽게 사용할 수 있도록 구성되어 있으며, 운영자가 해야 할 일, 발성자가 발성해야 하는 일, 그리고 발성자가 속으로 읽기만 해야 하는 일의 부분으로 나누어 글씨의 색깔로써 구분하고 있다. 특히, 발성자는 화면에 노란색으로 된 큰 글씨만 발성하면 되고 진행 도중에 운영자가 할 수 있는 일 등을 화면의 아래쪽에 메뉴박스를 만들어 사용 가능한 커맨드 종류를 나열하고, 색깔을 이용하여 현재 상태에서 사용가능한 커맨드가 어떤 커맨드인지도 쉽게 구분할 수 있다. 아래에 발성에 들어가기 전에 수집과 편집에 필요한 정보를 얻기 위한 운영자의 입력사항과 발성내용의 제시에 관하여 기술한다.

4.3.1 입력사항

운영자는 발성자가 발성을 시작하기 전에, 발성자에 대한 정보와 시스템 운영과 편집시에 필요한 사항을 입력한다. 그림 1은 운영자가 발성자의 발성에 앞

서 미리 입력해야 할 사항을 나타내고 있다. 그 입력 사항으로써, 운영자의 이름, 녹음장소, 발성자의 인격사항, 발성순서와 단어가 저장될 화일명, 읽어들일 데이터 화일명이 있다. 그리고, 발성자가 발성내용을 발성할 때 발성을 위한 준비시간과 발성시간의 길이를 각각 1초에서 3초의 시간 범위 내에서 선택할 수 있다. 그리고, 발성자가 발성 도중에 휴식을 취하게 하기 위하여 몇 개의 단어를 발성하고 쉬기를 원하는 지 그 갯수를 입력한다.

<< 운영자께서는 다음 사항을 입력하십시오 >>

1. 운영자의 이름을 입력하십시오(5자 이내) : *****
2. 녹음장소를 입력하십시오(영/한 30자 이내) : *****
3. 발성자의 인격사항
이름 : ***** 성명 : ** 생년월일 : year, mo, dd
출생지 : ***** 12세 이전의 거주지 : *****
신장(Cm) : cm 학력 : ***** 직업 : *****
- 4-1. 발성순서와 단어가 저장될 화일명을 입력하십시오.(영문8자 이내) : *****
- 4-2. 읽어들일 데이터 화일명을 입력하십시오.(영문 8자 이내) : *****
5. 발성준비 시간과 발성시간을 다음중에서 선택하십시오 : r
(예 - 2, default:5)

	준비시간	발성시간
1)	1	2
2)	1	3
3)	2	1
4)	2	2
5)	2	3
6. 쉬었다갈 단어수를 입력하십시오 : r
(예 - 70, default: 쉬지않고 계속함, 단 F에 의해서는 될 수 없음)

그림 1. 운영자의 입력사항

4.3.2 발성내용의 제시

운영자가 그림 1의 지시에 따라서 필요사항을 입력하면 여러 종류의 안내 화면이 나오고, 발성자가 발성하기 전 혹은 쉬었다가 다시 발성할 경우 시작 단어의 오류를 막기 위하여 세계의 연습용 단어를 제시하여 발성 연습을 하게하여 발성의 오류를 최대한 방지한다. 또한, 발성 데이터를 수집한 후 녹음 데이터가 누구의 것인지 쉽게 알 수 있게 하기 위하여 발성자의 이름을 디스플레이 시켜 발성자가 자신의 이름을 발성한 후 발성해야 할 발성내용 제시 화면이 나온다. 그림 2는 발성내용 제시 화면의 예를 나타내고 있으며, 여기서 "1451"은 발성해야 할 내용, 그 밑의 막대기들은 빨강색으로 채워지는 발성 준비 시간과 녹색으로 채워지는 발성개시 시간을 나타내어 올바른 발성을 유도하고, 발성단어 간의 시간도 어느정도 일정하게 유지한다. 메뉴박스에 대한 설명은 다음 절의 키의 기능에서 자세히 기술하며, 왼쪽의 6/20은 현재의 진행상황 즉, [발성된 단어수/전체의 단어수]를 표시하여 발성자의 심리적 부담을 줄인다. 오른쪽의 'saved'는 발성자가 잘못 발성하였을 경우 다시 발성하게 하기 위하여 기억시켜 두어야 하는데 발성을 모니터링하고 있는 운영자의 키 입력에 의해

기억되었음을 확인시켜 주는 표시이다. 이럴 경우에 물론 전체단어수도 증가한다.

1451



그림 2. 발성내용 제시

(“1451”의 발성내용이 제시되어 발성자가 잘못 발성했을 때)

4.3.3 자동으로 작성되는 화일

3.2.2절에서 기술했듯이, 정확한 편집과 작업의 편리성을 위하여 두 종류의 화일을 자동 작성한다.

첫째, [발성자가 발성할 내용이 저장되어 있는 데이터 화일명,{1st}]로써, 운영자는 발성자가 발성할 내용이 저장되어 있는 데이터 화일명을 입력한 후, 시스템을 빠져 나오면, 위의 화일명에 발성자의 이름과 발성자의 인적사항, 발성내용이 저장되어 있는 발성자 화일명이 쓰여진다. 물론, 발성자 화일명은 운영자가 임의로 줄 수 있으며, 이는 발성자가 발성한 내용이 어느 화일에 들어있는지를 알 수 있다. 그리고, 데이터 화일명이 같으면 발성자 이름과 발성자 화일명이 계속 첨가되고, 경우에 따라서는 발성자 화일명을 저장할 필요가 없는 경우에는 저장하지 않고 작업에서 빠져 나올 수도 있다(Q에 의해서). 표 2는 4digit.dat를 읽어 들일 때 생기는 4digit.lst를 나타내며 여기에는 4digit.dat를 발성한 발성자 리스트인 csk.lst, jyy.lst, pjs.lst가 등록되어 있음을 알 수 있다.

표 2. 4digit.dat를 읽어들일 때 생기는 4digit.lst의 내용

최선규 csk.lst
조양연 jyy.lst
박종성 pjs.lst

둘째, [발성자 화일명.lst]로써, 운영자가 임의의 발성자 화일명을 주면 [발성자 화일명.lst]라는 화일명이 생성된다. 여기에는, 화일명, 녹음일시, 녹음장소, 발성자의 인적사항(이름, 성별, 생년월일, 출생시, 12세 이전 거주지, 신장, 학력, 직업), 발성내용(발성일련번호, 단어번호, 단어명), 운영자의 이름이

저장되어 있다. 또한, 발성자가 발성 내용을 잘못 발성했을 경우, 운영자는 발성자로 하여금 다시 발성하도록 이 단어를 기억시켜 두어야 하고 또한 이러한 정보가 표시되어 있어야 한다. 그렇지 않으면, 음성데이터 수집이 끝나고 데이터 편집작업을 할 때 현재의 데이터가 올바른 발성인지, 잘못된 발성인지를 모르기 때문에 오류를 범하게 된다. 따라서, 잘못 발성하여 다시 발성한 단어에 대하여는 *를 표시한다. (이 단어는 편집을 하지 않도록 한다). 표 3은 위의 내용의 한 예를 나타내고 있다. 즉, 표 2에 등록된 리스트 중의 하나인 jyy.lst를 보면 표3과 같이 화일명, 녹음일시, 장소, 인적사항, 발성내용이 들어 있다. 특히, 발성내용에서 발성일련 번호 6번의 “9567”은 처음에 잘못 발성하여 *표가 붙어서 13번째 다시 제시되어 발성하였으나 또 잘못 발성하여 15번째 다시 제시되어 발성했음을 보여주고 있다.

표 3. 4digit.lst의 내용중 jyy.lst의 내용

<< 발 성 리 스투 >>		
1. 화일명 : jyy.lst		
2. 녹음일시 : 1992년 7월 1일 15시 8분 43초		
3. 녹음장소 : 삼광전자통신		
4. 발성자 인적사항		
-이름 : 조양연		
-생년월일 : 1970.1.2		
-출생지 : 서울		
-12세 이전 거주지 : 서울		
-신장 : 177		
-학력 : 대재		
-직업 : 학생		
5. 발성내용		
발성일련번호	단어번호	단어명
1	1	2566
2	2	9758
3	3	3567
4	4	2567*
5	5	6886
6	6	9567*
7	7	6738
8	8	7633*
9	9	8546
10	10	3462
11	2	9758
12	4	2567
13	6	9567*
14	8	7633
15	6	9567
6. 운영자 이름 : 김경태		

4.3.5 각 키의 기능

키보드와 마우스를 동시에 사용할 수 있으며, 키보드의 경우 특정 키를 누를 때 특정 키 주변의 다른 키가 눌러져서 오류를 범하는 실수가 없도록 구현되어 있다. 마우스의 경우는, 버튼 3개의 마우스를 사용할 수 있게 하고 그렇지 않을 때에도 부분적으로는 사용할 수 있다. 이러한 사용방법은 화면에 모두 디스플레이 되기 때문에 쉽게 알 수 있다. 그림 2의 아래쪽에 있는 메뉴박스는 각 커맨드 키의 기능을 나타낸다. 마우스를 사용할 때는 왼쪽버튼으로 메뉴박스의 키를 선택하여 한번 클릭함으로써 그 기능이 실현된다. 그리고, 다음 화면으로 진행시키는 커맨드는 많이 사용되므로 마우스의 오른쪽 버튼을 이용하여 편리성을 추구한다. 각 키의 기능을 설명하면 다음과 같다.

ENTER : 다음 화면으로 진행한다.

A(gain) : 발성자 한사람이 발성을 완전히 끝내고 다음 발성자가 발성할 때 혹은 다른 종류의 데이터를 발성하고자 할 때 사용한다.

J(jump) : 이미 발성된 단어 혹은 다시 발성해야 할 필요가 있을 때 그 단어의 번호를 입력함으로써 화면에 다시 나타내 준다.

P(ause) : 쉬고 싶을 때, 혹은 문제가 들발했을 때 사용한다.

SPACE : 발성 도중에 발성자가 틀리게 발성했을 때 다시 발성하게 하기 위하여 기억해 두었다가 나중에 화면에 나타나게 한다. 이 기능의 확인을 위하여 화면 오른쪽 밑에 'saved'라는 표시가 나타난다.

E(nd) : 발성의 끝을 나타낸다.

ESC : 본 시스템에서 탈출하여 도스 상태로 간다.

Q(uit no saving) : ESC와 같은 기능이지만, 다른 점은 운영자가 입력한 발성자의 인적사항 등의 데이터가 화일에 저장되지 않는다(ESC는 정장됨)

4.4 시스템의 적용

본 시스템의 유효성을 확인 위하여 음성수집 대상 자료로써 4연속 숫자음 35단어와 숫자음 25단어를 각각 9회씩 96명, 계산기 제어용 40개단어를 5회씩 60명이 발성한 총 63,840개의 단어를 약 2개월간에 걸쳐 수집하는 단계에서 시스템을 적용한다. 시스템을 적용할 때 평가 고려 항목으로써 발성자의 심적인 부

담감, 발성한 결과에 대한 자연성, 운영자의 부담감, 사용상의 편리성 등을 조사한 결과 종래의 리스트를 작성하여 읽는 방법에 비하여 훨씬 좋은 결과를 얻었다. 또한, 음성수집을 한 후 음성 데이터의 편집작업을 할 때에도 발성자에 관련된 많은 정보와 발성순서 등이 자동으로 생성되어 작업시간을 대폭 줄일 수 있었다.

V. 결 론

본 논문에서는, 음성 데이터베이스를 구축하기 위하여 음성데이터가 필요하고 이 음성 데이터를 수집함에 있어서 발성자의 부담을 줄이면서 정상시의 자연스럽게 발성된 음성 데이터를 수집하고, 편집 과정에서 정확하고 편리한 편집작업이 가능한 발성내용 제시시스템의 구현에 관하여 연구하였다. 이를 위하여 발성내용 제시시스템의 요구사항 및 기능을 정리하여 범용의 개인용 컴퓨터로서 구현하였으며 사람이 하는 일을 최소화하여 발성자와 운영자의 부담을 줄였다. 이를 4연속 숫자음 등의 63,840개의 단어를 실제 수집하는 단계에서 적용한 결과, 종래의 방식보다 발성의 자연성과 운영자의 부담감, 시간의 절약면에서 만족할 결과를 얻었다.

앞으로는, 본 시스템보다 더욱 개선된, 즉, 사람이 하는 일을 대폭 자동화 할 수 있는 시스템을 개발하여야 한다. 현재의 과정은 발성 -> 음향기기(녹음기, VTR, PCM 레코더, DAT)에 녹음 -> A/D 변환 -> 컴퓨터에 저장 -> 편집작업 -> 단어의 화일화로 되어 있다. 이렇게 여러 단계를 거치므로, 시간과 노력이 들고, 신뢰성에도 문제가 있다. 개선된 발성 제시시스템은, 위의 6단계를 동시에 자동으로 수행되는 시스템이 되어야 하고, 이를 위하여는, 음성을 직접 컴퓨터에 저장하기 위한 A/D변환 장치, A/D된 데이터를 실시간으로 하드디스크에 저장하는 기술, 수집된 데이터를 분석하여 자동으로 음성구간을 검출하는 기술, 편집된 내용을 화일화하는 기술 등(화일명의 명명법, 디렉토리의 구조 등)에 관한 연구가 필요하다.

-감사의 글-

본 연구는, 한국전자통신연구소 「보급형 음성데이터베이스 구축에 관한 연구」의 일부로 수행되었으며, 시스템 설계를 위하여 조언을 해주신 강철화 통신시스템 연구단장님, 윤병남 통신처리 연구부장님과 자

동통역 연구실의 최준혁, 한남용 연구원께 감사드립니다. 그리고, 본 프로그램의 코딩을 도와준 정보통신공학과와 조양연, 박관식, 최선규, 박종성 군에게 감사의 뜻을 전합니다.

참 고 문 헌

1. 김경태, 최준혁, 이용주, "음성 데이터베이스 및 관리 시스템의 구축," Proceeding of Korea-Japan joint symposium on acoustics, July 24-25, 1991.
 2. 板橋秀一, "單語音聲デ-タベ-ス," 音響學會誌 41, 10, pp.723-726, 1985-10.
 3. S. Itahashi, "A Japanese Language Speech database," Proc. ICASSP 86, Paper 7.4, pp. 321-324(1986).
 4. 板橋秀一, "音聲デ-タベ-ス," 電子情報通信學會誌, 70, 4, pp.433-438, 1987.
 5. Victor Zue et al, "The MIT ATIS system: Preliminary development, spontaneous speech data collection, and performance evaluation," EUROSPEECH '91.
- 김 경 태**
- 1949년 5월 9일생
 - 1972년 2월 : 경북대학교 전자공학과
 - 1980년 8월 : 연세대학교 대학원 전자공학과(석사)
 - 1985년 3월 : Tohoku Univ., Tapan
전기 및 통신 전공(박사)
 - 1986년 2월 ~ 1991년 2월 : 한국전자통신연구소 신호처리 연구실
 - 1991년 2월 ~ 현재 : 한남대학교 정보통신공학과
- 이 용 주**
- 1954년 1월 17일생
 - 1976년 2월 : 고려대학교 전자공학과 (학사)
 - 1983년 8월 : 고려대학교 대학원 전자공학과 (석사)
 - 1992년 8월 : 고려대학교 대학원 전자공학과 (박사)
 - 1980년 8월 ~ 현재 : 한국전자통신연구소 자동통역 연구실 실장
- 정 유 현**
- 1956년 8월 10일생
 - 1980년 2월 : 광운대학교 전자계산학과 (학사)
 - 1989년 2월 : 광운대학교 전자계산기공학과 (석사)
 - 1980년 8월 ~ 현재 : 한국전자통신연구소