

증권거래시스템 구축을 위한 음성 DB의 구축

어 범석*, 김 학진*, 김 순협*

*:광운대학교 컴퓨터공학과

A Study on the Construction of Speech DB to build a Stock sales system

*Bumsuk Eo, *hakjin kim, *Soonhyob Kim

scpol106@explore.gwu.ac.kr, kimsh@daisy.gwu.ac.kr

Abstract

음성 인식 시스템의 개발을 위해서는 음성 데이터베이스 구축이 중요한 과제의 하나로써, 많은 시간과 노력이 요구된다. 본 논문은 ARS 주식 거래 시스템에서 사용되는 주식의 매수, 매도, 증시 현황에 관련된 문장과 숫자음에 대하여 DB 구축한다. 이 DB 구축을 위하여 Dialogic 사의 D/41ESC보드를 장착하고, Window NT4.0 플랫폼에서 음성을 수집하였다.

본 논문에서는 음성 수집을 위해 전국의 20대에서 50대까지의 남녀에 대해 1명당 50개의 문장 또는 숫자음에 대하여, 유선 및 무선통화하여 데이터를 수집하였다. 또한 화자 독립 음성 인식을 위하여 1200명의 화자로 구성되어 있다. 지역별로 보면, 서울 및 경기, 강원 지역과 영호남, 충청 지역으로 나누었으며, 일반폰, 휴대폰, 공중전화의 환경에서, 그리고 실내와 실외환경에서 각각 수집하였다.

1. 서론

음성분야의 연구는 크게 음성 인식, 언어 번역,

음성 합성 분야로 크게 나눌 수 있다. 이 세 분야는 그 연구 방법이나 성능 측정 방법에서 서로 상이한 점이 많다. 그러나 한가지의 공통점은 대량의 데이터베이스가 필요하다는 점이다.

하지만 이러한 음성 데이터베이스를 구축하려면 시간적, 물적, 인적 자원의 효과적인 필요가 요구되므로 개개인의 연구자로서는 엄두도 내지 못할 일이다. 따라서 국외의 경우를 살펴보면 이러한 음성 DB의 구축은 국가과제로써 미국의 경우 국방부의 DARPA 프로젝트의 연구에 의해 구축된 음성 DB를 상무부 산하의 국립 연구소인 NIST (National Institute of Standards and Technologies)가 중심이 되어 통합적으로 운영하고 있으며, 유럽의 경우 각국이 단독 혹은 EC 여러 나라간의 음성 및 언어연구공동프로젝트가 구축되어 있다. 또한 일본의 경우 전자공업진흥회, 일본음향학회 등이 중심이 되어 각 대학, 연구소 및 기업이 공동으로 구축한 음성DB를 학회를 중심으로 보급 활용되고 있다.[2][3][4][5]

국내의 경우에는 각 대학이나 기업의 연구소별로 필요시 관련 데이터베이스가 만들어서 사용하고 있는 실정이다. 따라서 국내에서 국가적 주도 차원에서 음성 데이터베이스의 구축을 지속적으로 추진해야 한다.

이제 지금까지 본 연구실에 수행된 증권거래시스템 구축을 위한 음성 데이터 베이스 구축 작업에 대해 살펴보겠다.

2. DB 수집 계획

우선 DB를 효율적으로 수집하기 위하여 수집 계획을 작성함으로써 DB 구축을 좀 더 효과적으로 할 수 있다.

2.1 수집 인원

데이터 수집을 위한 효과적인 인원은 2인 1조로 하여 다수의 조를 운영함으로써 수집속도를 증가시킬 수 있다.

또한 전체 수집 인원 1232명 대해 문장음과 숫자음을 1번씩 발음하였고, 일반종목과 코스닥 종목은 400명을 대상으로 1번씩 발음하였다.

2.2 수집 대상

연령별로 20대 22%, 30대 30%, 40대 30%, 50대 18%로 구성하였으며, 남녀의 성별 비율은 남 57%, 여 43%로 구성되어 있다.

2.3 수집 환경

전화 거는 위치에 따라 즉 주위 잡음환경의 영향을 고려하여 사무실과 일반 거리 다시 말해 외부환경에서의 비율은 사무실 58%, 외부 42%로 되어 있다.

2.4 수집 장소

우선 인구 통계표에 따라 그 비율을 나누었으며, 전체 수집 인원 중 632명에서 서울 56%, 경기 27%, 강원 16%로 하여 서울지역에서 357명, 경기에서 173명, 강원에서 102의 데이터를 수집하였다. 그리고 충청도 29%, 경상도 35%, 전라도 36%의 비율로 600명에 대한 데이터를 수집하였다.

2.5 전화기의 종류

전화기 종류의 특성을 고려하여 일반 유선전화와 핸드폰, 공중전화 3가지로 나누었으며, 유선전화 42%, 핸드폰 48%, 공중전화 10%로 나누었다.

2.6 대상어휘

증권 거래 시스템에서 쓰이는 어휘를 조사하여 문장음 1100문장, 숫자음 8850개, 일반종목 1900개, 코스닥 800개를 가지고 데이터 수집을 하였다.

문장음에는 증권종목과 숫자음이 섞여 있으며, 숫자음에는 주민등록 번호와 단독 숫자음, 가격 단위 세 종류가 있다.

표 1. 대상어휘의 예

구분	예제
문장음	한양증권 매도호가 얼마입니까? 하한가에 백칠십주를 매수하겠습니까 통일중공업 거래량이 얼마나 됩니까
숫자음	920209-7125891 148,200 7470718
일반종목	강원산업 금호석유유
코스닥종목	골드뱅크 가산전자

3. 시스템 구조

시스템의 기본 구조는 Windows NT 플랫폼에서 서비스 팩 4이상에서 데이터를 수집하였으며, 사운드 카드는 사운드 블라스터 라이브를 사용하였다.

또한 전화기라는 특수한 목적을 위해 Dialogic사의 D-41ESC 카드로써 4개의 아날로그 Loop-Start 전화 라인과 직접 연결하여 음성을 수집하였다.[6]

표 2. D/41 ESC 제품 사양

포트	4 Ports
최대 확장	16 / System
Analog Network interface	Loop-Start Interface
자원 공유	AEB : SCBUS or PEB
제어 프로세서	Intel 80C186
디지털 신호 프로세서	Motorola DSP56001

4. 수집시나리오

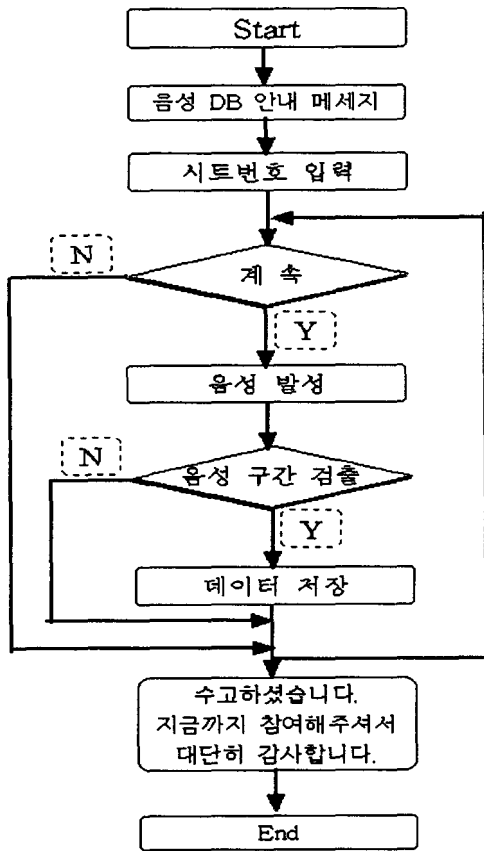


그림 1. 수집시나리오 흐름도

5. 음성 데이터의 편집

모든 음성 데이터는 A/D변환을 통하여 컴퓨터의 하드 디스크에 저장되며, 문장음, 숫자음, 일반 종목, 코스닥 종목별로 구분하여 별도의 디렉토리에 저장하였다. 편집은 끝점 검출 알고리즘에 의하여 자동으로 음성 구간만 검출을 하여 저장되며, 음성 분석 틀을 이용하여 음성 파형을 디스플레이하여 수정한다.

6. 음성 데이터의 저장

음성파일은 μ -low 파일형태로 저장이 되며, 각각 따로 저장이 된다. 또한 녹음된 파일의 정보를 담고 있는 text 파일이 디렉토리별로 하나씩 존재한다.

본 연구에서는 각각의 데이터를 특성에 맞게 계층적 디렉토리로 구분하여 검색에 용이하도록

하였다.

또한 text 파일에는 header 정보를 두어 파일의 정보를 효율적으로 관리하도록 하였다.[1]

file name | 시트번호 | 화자 이름 | 지역 | 연령 | 성별 | 대상어휘 구분 | 환경 | 전화기 종류 | 단어수 | 음절수 | 음소수 | 어휘 | 특수기호

표 3. 지역별 코드표

지 역	코 드
서 울	1
경 기	2
강 원	3
충 청	4
전 라	5
경 상	6

표 4. 연령별 코드표

연 령	코 드
20 대	1
30 대	2
40 대	3
50 대	4

표 5. 성별 코드표

성 별	코 드
남	1
여	2

표 6. 대상 어휘별 코드표

대상어휘 구분	코 드
문장음	1
숫자음	2
일반종목	3
코스닥종목	4

표 7. 환경별 코드표

환 경	코 드
사무실	1
거리	2

표 8. 전화기 종류별 코드표

전화기 종류	코 드
유선전화	1
핸드폰	2
공중전화	3

7. 결론 및 향후계획

본 연구에서는 증권거래시스템이라는 특수한 목적을 위한 음성 DB 구축에 대해 살펴보았다. 전화회선에 존재하는 회선잡음의 영향을 고려하기 위해 직접 그 지역에서 음성을 수집하였으며, 주변 잡음을 고려하여 조용한 사무실과 일반 거리에서 음성을 수집하였다.

음성 DB의 구축을 위한 음성 수집은 지역별, 성별, 나이별 개인차를 극복하기 위하여 증권거래를 주로 하는 30대에서 40대의 비율을 높게 잡았으며, 그리고 지역별 인구 분포에 따라 전국을 6개 지역으로 구분하여 데이터를 수집하였다.

향후에는 지금까지 구축한 데이터베이스를 기초로 하여 Segment와 Labeling 작업을 들어갈 예정이다. 이렇게 함으로써 우선은 증권거래 시스템에서 단어 단위의 인식 시스템을 구현할 것이고, 나아가서 연속음 인식 시스템을 구현할 것이다.

8. Reference

- [1] 최승호, '전화음성데이터 수집에 관한 연구'-제11회 음성통신 및 신호처리 워크샵 논문집
- [2] NIST: Speech Copora Produced on CD-ROM Media by the National Institute of Standards and Technology(NIST), April, 1991
- [3] 이용주, "해외 음성 DB 구축 동향"-제 12회 음성통신 및 신호처리 워크샵 논문집
- [4] A. Kurematsu and K. Takeda, "ATR Japanese speech database as a tool of speech recognition and synthesis," Proc. ESCA'89, pp.2.3.1-4, 1989
- [5] Y. Sagisaka, K. Takeda, M. Abe, S. Katagiri, T. Umeda, and H. Kuwabara, "A large-scale Japanese speech database," Proc. ICSLP'90, pp. 1089-1092, 1990

[6] Dialogic corporation, "Dialogic D/41ESC User's Guide"

[7] 이용주, "국어공학센터의 한국어 음성DB 구축 계획"-제 12회 음성통신 및 신호처리 워크샵 논문집 pp.276-279

[8] 최 인정, "KAIST 통신연구실의 음성 데이터베이스 구축 현황"-제 12회 음성 통신 및 신호처리 워크샵 논문집 pp.272-275