전화음성데이타 수집에관한 연구

최 승 호*, 김 형 근**, 방 만 원***
동신대학교 정보통신공학과*
한국방송통신대학교 전자계산학과**
목포대학교 전자공학과***

A STUDY ON THE COLLECTION OF TELEPHONE SPEECH DATA

Seung-Ho choi*, Hyung-Kuen Kim**, Man-Won Bang***
Dept. of Infomation and Comunications Eng., DongShin Univ.*
Dept. of Computer Science, Korea Air Coresponddence Univ.**
Dept. of Electronics Eng., MokPo Univ.***

요 약

본 논문에서는 전화망을 이용하여 다양한 환경을 내포한 음성시료를 수집하여 스펙트럼 기울기의 변동을 검토함으로써 지역간 전화회선 특성을 파악하고, 시료 수집은 발화자수 선정모델을 근거로 광주·전남지역에 거주하는 84명을 대상으로 영역별, 연령별, 성별로 구분하여 수집하였다.

또한, 데이타 수집방법은 수집장치와 발화자가 시나리오의 정해진 순서에 따라 작성되었다.

Ⅰ. 서 론

음성 정보처리는 음성언어의 특징을 분석하는 음성분석, 이러한 특징을 이용하여 기계가 음성을 알아듣는 시스템인 음성인식, 그 요구사항을 처리하여 다시 음성으로 알려주는 음성합성으로 나누어질 수 있다. 효율성과 경제성이 높은 음성신호를 전송하기위해 음성의 분석기술, 부호화기술, 합성및 인식기술 그리고 통화품질 기술등의 발전이 요구되고 있으며 인간중심의 통신서비스를 만들어내기 위해서는 음성 정보처리 연구가 필연적으로 선행되어야 한다.

정보화 사회에서 가장 널리 구축된 통신망인 전화망을 이용하여 다양한 형태의 서비스를 제공하기 위해 전화 음성인식에 대한 연구가 요구되고 있다. 이를 위해서는 음성 정보처리의 연구가 조직적이고 장기적으로 실행되어야 한다. 최근, 전화시스템을 통한 다양한 서비스 제공을 위해 전화음성인식에 대한 연구가 활발히 진행되고 있다. 따라서, 본 논문에서는 실용성 있는 전화음성 인식장치의 개발을 위해 지역적 특성과 실제 회선잡음이 존재하는 상황에서 음성데이타를 수집하고, 분석하고자 한다.

Ⅱ. 전화음성 수집을 위한 모델

음성 정보처리의 연구는 H/W와 S/W등 관련기술의 진보에 따라 음성인식및 합성시스템에 대한 연구와 개발이 활발히 진행되고 있다. 인식시스템의 개발을 위해서는 여러사람이 발성한 대량의 각종 데이타를 수집분석하여 개인차와 조음결합의 현상을 파악하여야 한다. 이를위해 92년 통계청 인구자료를 근거로 지역별에 따른 인구비를 기준으로 데이타 수집인원을 정하는 모델을 제시하였다.

전국의 음성수집 표본대상을 1000명으로 정해놓고 전국 시·도별 인구분포를 근거로하여 연령별 인구비를 0~65세 이상까지 5세간격으로 나누어 각 지역별로 표 1에 나타내었다.

표 1. 전국 연령별 인구비에 따른 수집인원

연령	전국	서울	인천·경기	강원	충북	대전·충남	대구·경북	부산·경남	전북	광주·전남	제주
0 - 4	7.7	7.4	8.5	6.6	7.0	7.0	6.9	7.6	6.4	6.6	7.0
5 - 9	8.6	8.4	9.5	8.4	8.9	8.9	8.6	9.2	8.9	9.2	8.9
10 - 14	9.2	9.0	8.9	9.9	9.7	9.7	9.3	9.4	10.5	10.5	9.7
15 - 19	10.2	10.3	8.5	10.5	10.5	13.5	10.6	10.1	11.9	11.5	11.6
20 - 24	9.9	10.9	9.7	9.4	9.7	10.1	10.2	9.8	9.8	9.4	11.4
25 - 29	9.8	10.6	12.4	8.3	8.2	8.7	9.2	9.9	7.3	7.8	9.7
30 - 34	9.7	10.1	12.1	8.4	8.4	8.[illegible]	8.4	7.0	7.3	7.6	8.5
35 - 39	7.5	8.1	7.6	7.0	6.8	6.9	7.2	7.9	6.6	6.2	6.2
40 - 44	6.6	6.6	5.3	6.1	4.9	5.7	5.6	6.1	5.7	5.7	5.6
45 - 49	5.2	5.6	4.2	5.6	5.1	5.0	5.1	4.9	5.1	5.3	5.0
50 - 54	4.1	4.4	3.7	5.7	5.5	4.6	5.0	4.4	5.4	5.2	4.5
55 - 59	3.6	3.0	2.8	4.4	4.7	4.2	4.2	3.6	4.9	4.6	3.6
60 - 64	2.7	1.9	2.2	3.3	3.5	2.8	3.0	2.6	3.8	3.5	2.4
65 -	6.1	3.3	4.3	7.4	7.1	6.0	6.5	4.5	7.0	6.6	5.8
인원수 [명]	1000	264	183	37	32	71	117	172	48	84	12

그러나 이러한 방법은 너무 포괄적이고 광범위하기 때문에 제주도 지역을 제외하고 각 지역을 생활권 영역으로 나누었다. 생활권 영역으로 나눌때 서울은 7개, 인천·경기 9개, 부산·경남 11개, 대구·경북 13개, 광주·전남 6개, 대전·충남 6개, 전북 7개, 강원 5개, 충북 4개이다.

이에대한 광주·전남의 예를보면 다음과 같다.(표 2-3) 광주·전남지역은 광주시가 그 주변의 인접군과 서로 사방으로 연결되어있어 광주시를 독립해놓고 수집대상자수를 정할 수 없기 때문에 인접군과 통합하여 영역을 나누었다.

표 2. 광주전남의 영역별 수집대상자 수

영 역	지 역 명	비율 (%)	인원(명)	[남/여]
1	광주시, 곡성군, 담양군, 영광군, 장성군, 화순군	42.8	36	[18/18]
2	목포시, 신안군, 무안군, 영암군	13.1	11	[5/6]
3	순천시, 고흥군, 구례군, 보성군, 승주군	13.1	11	[7/4]
4	여수시, 여천시, 여천군, 동광양시, 광양시	13.1	11	[5/6]
5	해남군, 장진군, 완도군, 진도군, 장흥군	10.8	9	[4/5]
6	나주시, 나주군, 함평군	7.1	6	[3/3]
계		100.0	84	[42/42]

표 3. 영역별 · 연령별 수집대상자 수

나이 \ 영역	1	2	3	4	5	6
10 ~ 14	4	1	1	3	3	0
15 ~ 19	5	0	1	1	1	2
20 ~ 24	5	2	2	0	0	1
25 ~ 29	6	1	1	1	1	0
30 ~ 34	3	3	2	0	0	1
35 ~ 39	3	1	0	0	2	0
40 ~ 44	2	1	0	2	0	0
45 ~ 49	3	0	2	2	0	0
50 ~ 54	2	1	1	0	1	0
55 ~ 59	3	1	1	0	1	1
60 ~	1	0	0	2	0	1
계	36	11	11	11	9	6

이와같이 각 지역을 영역별로 나누어 음성을 수집함으로써 지역에 따른 표준어와 방언의 편차를 최소화 시킬 수 있다고 사료된다.

Ⅲ. 전화음성 수집장치

영역별로 정해진 수집대상자를 무작위로 추출한 발성자의 음성을 녹음하고 녹음된 음성을 재생할 수 있는 전화음성 수집장치를 그림 1과 같이 구성한다. 구성된 수집장치는 음성데이타 관리 S/W와 시나리오관리 S/W를 포함하게 되며 이들 S/W는 음성수집에 관련된 모든 환경을 효과적으로 제공하게 된다.

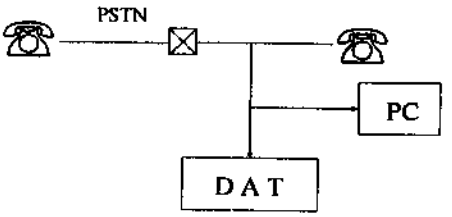

그림 1. 음성수집장치 구성도

Ⅳ. 수집시나리오

음성데이타를 수집하기위해 발화자와 음성데이타 수집장치가 시나리오의 정해진 순서에 따라 데이타를 수집하도록 작성하였으며 그 과정은 그림 2와 같다.

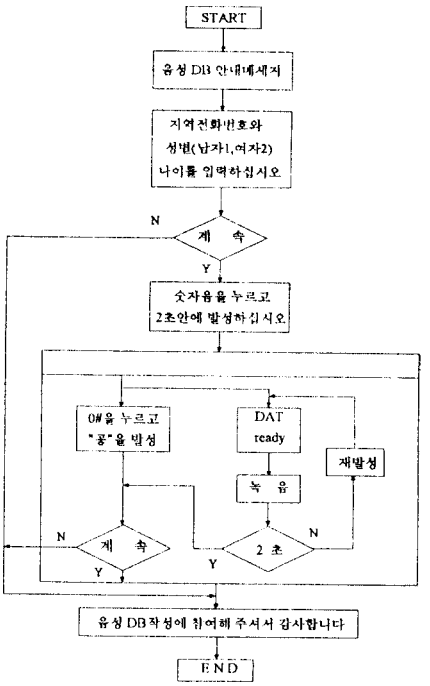

그림 2. 수집시나리오 흐름도

V. 전화 음성데이타의 편집

전화음성 데이터구축에 대한 전 과정도는 그림 3과 같으며, 각부분에 대한 세부적인 방법은 다음과 같다.

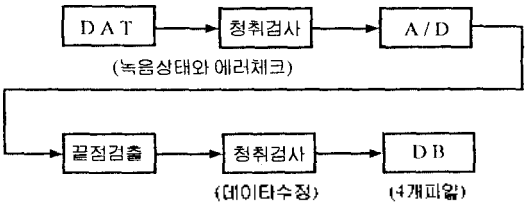

그림 3. 데이터 편집의 과정도

공중전화회선망을 통하여 들어온 전화음성은 DAT를 거쳐 음성특성 분석장치에 저장된다. 저장된 음성을 대상으로 잘못 발성되었거나 level over 등의 경우를 조사하여 양호한 음성을 선정한다. 음성데이타의 검사와 부분적인 수정을 위해 선정된 녹음데이타를 대상으로 숫자음별로 청취검사를 수행한다.

수집된 음성데이타는 차단주파수가 3.4[KHz]인 저역통과필터에 통과시켜 고주파성분을 제거하며, 8[KHz]의 샘플링주파수와 12비트 분해도의 A/D변환을 거쳐 저장된다.

A/D변환된 음성데이타는 성대의 스펙트럼형태를 정확히 추출하기 위하여 음성신호 스펙트럼의 동적구간을 줄이는 Preemphasis과정을 통해 성도특성을 부각시킨다. 음성신호의 분석조건은 표 4에 나타내었다.

표 4. 음성신호의 분석조건

sampling freq.	:	8 [Hz]
cut off freq. of LPF	:	3.4 [Hz]
A/D Resolution	:	12 [bit]
Analysis Window	:	20 [ms]
Overlapping interval	:	10 [ms]
window function	:	Hamming win.
LPC Order	:	14

또한, 수집된 데이타의 환경조건은 표 5와 같다.

표 5. 수집된 데이타의 환경조건

수집 시간	수집 장소	전화기 종류	소음 종류
AM 8 - AM 12 : 27	가정집 : 54	일반 전화 : 75	자동차 소리 : 3
PM 1 - PM 6 : 45	사무실 : 17	무선 전화 : 6	떠드는 소리 : 9
	가 게 : 10		음악 소리 : 5
PM 7 - AM 7 : 12	공중전화부스 : 3	공중 전화 : 3	혼선 : 2
			기타 : 7

수집음성 데이타는 단모음 7개와 숫자음 22개를 3회씩 발성한 총 7308개 데이타이다.

VI. 전화음성의 특성분석

시내의 전화선로에서는 대역폭이 0.3~3.4㎑ 정도이기 때문에 고대역의 정보를 상실 할 수 있고, 시외에서는 고대역의 정보가 전송될 수 있기 때문에 전화음성의 특성을 파악하기가 매우 어려운 문제이다.

국내의 경우 전화음성에 대한 선로의 특성조사가 이루어지지 않고 있는 실정이기 때문에, 현 인식 시스템에 전화음성을 입력으로 사용하면 인식률 저하가 현저하게 나타나고 있다. 이는 선로의 회선잡음과 왜곡및 누화등에 의한 잡음의 영향이라 사료된다. 이와같이 음성에 섞여진 잡음을 제거하는 방법은 short-time spectrum amplitude estimation 사용하는데, 이는 사람이 음성을 들을때 위상정보 보다는 진폭정보가 훨씬 유효하기 때문이다.

6.1 단시간자기상관함수

단시간 구간의 음성신호 Sn(m)을 이용한 자기상관함수를 단시간 자기상관함수 R(k)라 하며 식 1과 같다.

$$R(k) = \sum_{m=-\infty}^{\infty} S(m)S(m-k)h_k(n-m) \qquad (1)$$

여기에서, R(k)는 m이 n-N+1로부터 n까지의 단시간 구간에 해당하는 음성신호 s(m)에 대해 K만큼 지연된 단시간 자기상관함수이다. 식 (1)의 시스템에 대한 구성도는 그림 4와 같다.

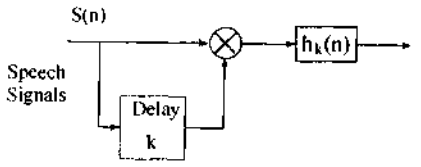

그림 4. 단시간 자기상관함수의 구성도

6.2 스펙트럼 기울기의 검토

본 연구에서는 전화회선 특성을 확인하기 위하여 지역간의 스펙트럼 기울기를 검토하였다. 왜냐하면, 음성 스펙트럼의 기울기 분포가 다르면 인식율에 악영향을 미치기 때문에 스펙트럼의 기울기를 나타내는 파라메터

인 단시간 자기상관함수로부터 첫번째 자기상관 계수 A1을 사용하여 지역간의 기울기 정도를 측정하였다. 그림 5는 제 6영역에서 수집하여 1~5영역으로 부터 보내온 음성 데이타를 자기상관함수 A1의 평균값에 대한 상대빈도 분포도를 나타낸것으로써 발성영역에 따라 그 분포의 차이가 있음을 볼 수 있다. 이 차이는 화자의 개인차, 방언, 전화상의 회선특성등에 대한 영향이라 사료된다.

표 6은 자기상관함수 A1의 평균값에 대한 상대분포를 계산한 예이다.

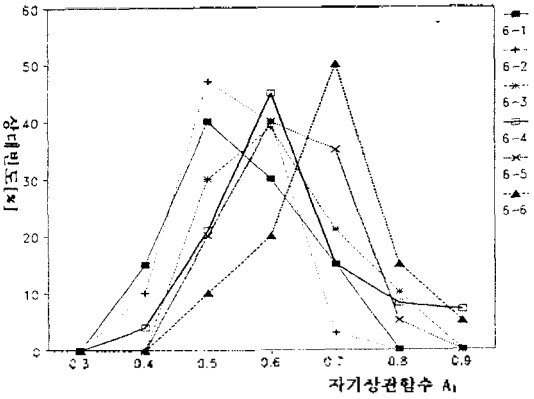

그림 5. 첫번째 자기상관함수 A1의 평균값의 분포

표 6. 평균 자기상관함수 분포의 예

대표값	범위	갯수	상대빈도
0.0	0.000~0.025	0	0%
0.05	0.026~0.075	0	0%
0.10	0.076~0.125	0	0%
.			
0.30	0.276~0.325	0	0%
.			
0.50	0.476~0.525	30	15%
.			
0.70	0.676~0.725	80	40%
.			
1.00	0.976~1.000	0	0%

Ⅷ. 결론

본 연구에서는 실용성있는 전화음성 인식장치의 개발을 위해 실제 회선잡음이 존재하는 상황에서 음성데이타를 수집분석하였다.

음성시료의 수집은 지역간, 남녀간, 연령간의 개인차를 극복하기 위해 발화자 수를 통계청의 인구자료를 근거로 광주·전남지역을 6개 영역으로 광역화하여 10~64세 까지의 연령층을 5세 간격마다 연령별, 성별로 나누어 배정하였으며, 실제 현장상황하의 음성을 수집하기 위해 공중전화 회선망을 통해 직접 전화통화를 가정한 시나리오를 작성하여 이 시나리오에 따라 자연스럽게 발성된 발화자의 전화음성을 DAT에 녹음하였다.

수집된 음성시료에 대해서는 평균적인 자기상관함수를 사용하여 지역간 스펙트럼 기울기의 변동을 검토하였으며, 향후 변동에 대한 대책과 기울기의 정규화 방법에 대해 연구할 예정이다.

감사의 글

본 논문은 1993년도 통신학술지원과제인 "전화망에 의한 연구용 음성 DB및 구축에 관한 연구"의 지원으로 이루어 졌읍니다.

참고 문헌

[1] 이용주, 김경태, "우리말 음성DB의 구축을 위하여", 한국음향학회 음성처리 Workshop pp.92-97, 1988

[2] 정유현, "음성 데이타베이스의 연구동향", 전자통신, 10.1992

[3] 한국통신, "각 시도 전화번호부", 1993

[4] 통계청, "인구주택 총 조사 보고서", 1992

[5] 이용주외 7인, "단어음성 데이타의 수집및 DB구성 시스템", 대한전자공학회 추계종합학술대회, Vol.9, No.2, pp.332-335, 1986

[6] 방만원외 3인 "전화회선을 이용한 음성DB수집에 관한연구", 한국통신학회학술논문집, Vol.11, No.1,PP 331-333, 7, 1992

[7] P. C. Millar, I. R. Cameron, " A very large telephone speech database collected using an automated voice-interaction dialogue", British Telecom, Reserach L., IEEE, 1988

[8] Dialogic corporation, *"DIALOG/4x Multi-Line voice communication system user's guide"*, ver.2.31, 1989

[9] L.Villarrubia, et al,"influence & the telephon line on automatic speech recognition", proc, of European conference on speech comm.and tech., vol.2, pp.1363-1366, 1991

[10] S. Itahashi, "A Japanese Language Speech Database", proc. ICASSP 86, TOKYO, 1986

[11] 長島廣海 他, "電話音聲認識における變動要因の檢討", 2-3-9, 音講論集, 1983