1999년도 한국음향학회 학술발표대회 논문집 제18권 제2(s)호

문맥종속 반음소단위에 의한 자동 음운 레이블링 시스템의 구현 및 성능평가

박순철*, 김태환*, 김봉완*, 이용주*
*원광대학교 컴퓨터 공학과

Implementation of Automatic Phoneme Labelling System Using Context-dependent Demi-phone Unit and Performance Evaluation

(Soon-Cheol Park*, Tae-Hwan Kim*, Bong-Wan Kim*, Yong-Ju Lee*)
*Dept., of Computer Eng., Wonkwang Univ.

요 약

음소 단위로 레이블링된 데이터베이스는 음성연구에 있어 매우 중요하다. 그러나 수작업에 의한 음소분할 및 레이블링 작업은 많은 시간과 노력이 필요하기 때문에 자동 음소분할 및 레이블링 시스템에 대한 많은 연구가 진행되고 있다.

저자들은 자동레이블링 시스템에서 레이블링 분할의 단위로monophone과 triphone의 장점을 포함하는 문맥 종속 반음소 단위 모델을 이용한 자동 음소분할 및 레이블링 시스템을 제안한바 있다[1]. 본 논문에서는 문맥종속 반음소 단위 자동음소분할 및 레이블링 시스템의 성능을 개선하기 위하여, 반음소의 단위를 개선하였다. 기존에 제안된 반음소 단위는 음소의 중점을 기준으로 left/right의 반음소 단위로 양분하였다. 본 논문에서는 음소의 길이가 120ms 이상일 경우 음소의 천이구간의 특성을 잘 나타낼 수 있도록, 음소의 앞뒤구간 각각60ms를 전반음소와 후반음소로 나누고, 나머지 안정구간을 별도의 모델로 구성하였다.

본 논문에서 제안한 반음소 단위의 성능을 평가하기 위하여 PBW 452단어를 발성한 남자 30명분의 데이터를 이용하여 레이블링 시스템을 훈련하고, 훈련에 사용하지 않은 남자 4명분의 데이터를 이용하여 테스트 하였다. 실험결과, 기존의 반음소 단위에 비하여 10ms에서 69.09%로 1.65%, 20ms에서 85.32%로 1.02%의 성능향상을 가져왔다.

1. 서론

음성연구를 수행하기 위해서는 음소와 같은 기본 단위로 분할되고 레이블링된 대량의 음성데이터베이스의 구축이 필수적으로 요구된다. 음소와 같은 기본단위로 분할 및 레이블링하는 작업은 사람이 직접 수행할 수 있지만, 수작업에 의한 음성 데이터베이스 구축은 시간이 많이 소요되는 작업이며, 소수의 음성 전문가에 의존할 수밖에 없고, 구체적인 판단기준을 미리 정해놓더라도 상당부분 개인의 주관적인 판단에 의존해야 하기 때문에 일관성이 결여되는 문제가 있다[2,3].

이와 같은 문제를 해결하기 위하여, 음성을 자동분할 및 레이블링하는 기술이 다양하게 연구되어 왔다[4,5,6,7].

대용량 음성데이터베이스를 구축하기 위해 고려되어할 분야중의 하나가 레이블링 시스템의 인식단위에 대한 문제이다. 본 논문에서는 triphone에 비해 모델의 수를 훨씬 줄이면서, triphone과 같이 전후음소에 대한 조음효과를 잘 반영할 수 있는

인식의 단위로 제안되어 성능이 입증된바 있는[1] demiphone의 모델을 개선하여 성능을 향상시켰다.

2절에서는 본 논문에 사용된 demiphone를 정의와 모델 개선내용에 대하여, 3절에서는 자동음소분할 시스템의 구성에 대해서 기술한다. 4절에서 성능을 분석한 후 마지막으로 결론을 맺는다.

2. Demiphone

2.1 Demipone의 정의

일반적으로 음소는 정상시점을 기준으로 선행음소의 영향을 받는 전반부와, 후속음소의 영향을 받는 후반부로 분류할 수 있다.

이와같이, 음소를 두 부분으로 나누어 생각해볼 수 있는 이유는, 많은 경우 선행음소와 후속음소가 미치는 영향이 음소의 전반부 및 후반부에 국한된다는 점에서 [그림 1] 의 (d)와 같이 음소를 성질이 서로 다른 두 부분으로 구분 지울 수 있다[8,9].

#	나	비	#

a. 음절단위

#	n	a	b	i	#

b. 음소단위(monophone)

#	#-n+a	n-a+b	a-b+i	b-i+#	#

c. triphone단위

#	#-n	n+a	n-a	a+b	a-b	b+i	b-i	i+#	#

d. demiphone 단위

그림 1. demiphone의 경계

demiphone은 음소 또는 변이음을 그것의 전후 음소의 영향을 받지 않는 정상상태시점을 중점을 경계로 하여 음소를 양분함으로써 얻어지는 음성단위이다. 이렇게 나누어진 demiphone는 선행음소에 의한 조음효과를 포함하는 left-demiphone과 후속음소에 의한 조음효과를 포함하는 right-demiphone의 두 부분으로 나누어진다. 즉, 음소의 경계와 diphone의 경계를 동시에 가지는 단위라고 할 수 있다.

또한, [그림1]에서 보는 것과 같이 demiphone은 선행음소의 영향을 받는 left-part와 후속음소의 영향을 받는 right-part로 분류되어 전·후 음소와의 문맥조음 관계를 포함하게 된다.

2.2 Demiphone의 모델개선

Demiphone은 음소의 천이구간을 모델링하여 전·후음소의 문맥조음 관계를 포함하게 된다. 그러나, 음소의 길이가 길어진다면 전체음소에서 안정화 구간이 길어진다. 따라서, demiphone의 천이구간을 모델링하는 특성이 약화된다.

본 논문에서는 위와 같은 이유로 120ms이상의 지속시간을 가지는 음소에 대해 [그림 2]와 같이 음소를 3부분으로 나누어 모델링 하였다. 음소의 전·후 60ms를 각각 left·right-demiphone으로 모델링 하고 음소의 안정구간을 별도로 모델링 하였다.

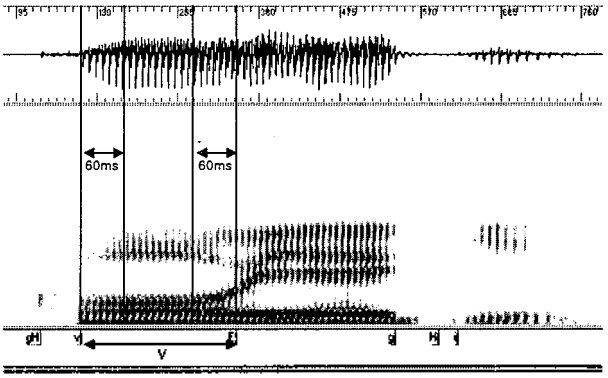


그림 2. Demiphone의 모델개선

[그림 2]는 '거액이'라는 단어의 예이다. 'v'의 경우 전반 60ms 'gh-v', 'v', 'v+E'의 3부분으로 모델링 하였다.

3. 시스템의 구현

본절에서는 개선된 demiphone의 성능을 평가하기 위하여 구축한 자동 음소분할 레이블링 시스템에 대해서 기술한다.

[그림 3]은 자동 음소분할 및 레이블링 시스템에서 HMM모델의 일반적인 구성 절차를 보이고 있다.

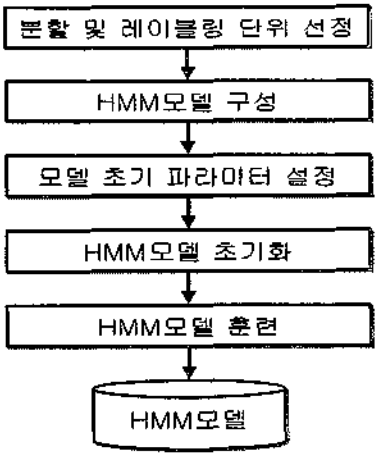


그림 3. HMM모델 생성 절차

3.1 레이블링 단위 선정

구현된 레이블링 시스템에서는 레이블링의 기본 단위로 당 연구실에서 "국어 정보베이스"사업의 일환으로 제작한 음소적으로 균형이 잡힌 PBW(Phonetically Balanced Word) 452 단어에 사용된 유사음소를 선정하여, triphone과 demiphone으로 확장하여 사용하였다.

레이블링을 위한 유사음소를 선정하기 위해 먼저 수작업으로 레이블링된 PBW 452 단어의 음성 데이터베이스를 분석하였다. 데이터의 분석은 [표 2]에서 나타낸 유사음소만을 대상으로 하였으며, 그 결과는 각각의 유사음소에 대해 음소, 출현 빈도, 평균 지속시간과 표준편차로 구하였다. 거의 출현하지 않거나 전혀 출현하지 않는 음소들은 레이블링 단위에서 제거하였다.

한국어 기본 음소단위를 기준으로 하여, 파열음, 파찰음, 그리고 유음의 폐쇄구간과 파열/마찰구간으로 나누고, 나뉘어진 각 구간의 유성음화를 고려하였다. 공명음화, 파열음·파찰음에서의 불파음화, 그리고 유음의 탄설음화의 경우 단일 구간으로 취급하였다.

분석을 통해서 선택된 레이블링 단위는 PBW 452어절 음성 데이터베이스에 사용한 유사음소 94개 중 [표 1~4]에 나타낸 바와 같이 마찰음의 경우 유성음화와 공명음화를 고려하였으며, 이중모음인 'ㅖ' 와 'ㅒ', 'ㅚ' 와 'ㅞ'를 하나의 음소로 취급하였다. 이렇게 89개의 유사음소와 1개의 묵음을 포함하여 총 90개의 레이블링 단위를 선정하였다.

최종적으로 선택된 레이블링의 기본 단위는 [표 3]에 나타낸 것처럼 89개의 유사음소와 1개의 묵음을 포함하여 총 90개이다.

	음소	폐쇄구간	폐쇄구간의 유성음화	파열/마찰구간	파열/마찰구간의 유성음화	불파음화	공명음화
파열음	ㄱ	g	gV	gH	gHV	gC	gR
	ㄲ	G	GV	GH	GHV		
	ㅋ	k	kV	kH	kHV		
	ㄷ	d	dV	dH	dHV	dC	dR
	ㄸ	D	DV	DH	DHV		
	ㅌ	t	tV	tH	tHV		
	ㅂ	b	bV	bH	bHV	bC	bR
	ㅃ	B	BV	BH	BHV		
	ㅍ	p	pV	pH	pHV		
파찰음	ㅈ	z	zV	zH	zHV		zR
	ㅉ	Z	ZV	ZH	ZHV		
	ㅊ	c	cV	cH	cHV		

표 1. 파열음과 파찰음의 기호 목록

	음소	마찰성분	유성음화	공명음화		음소	기호
마찰음	ㅅ	s	sV		비음	ㅁ	m
	ㅆ	S				ㄴ	n
	ㅎ	h	hV	hR		ㅇ	N

표 2. 마찰음과 비음의 기호 목록

	음소	폐쇄구간	폐쇄구간의 유성음화	기식음	기식음의 유성음화	공명음화	탄설음화
유음	ㄹ	r	rV	rH	rHV	rR	l

표 3. 유음의 기호 목록

	음소	기호	음소	기호	음소	기호	음소	기호
모음	ㅏ	a	ㅣ	i	ㅑ	ja	ㅝ	wv
	ㅓ	v	ㅔ	e	ㅕ	jv	ㅞ	wE
	ㅗ	o	ㅐ	E	ㅛ	jo	ㅟ	wi
	ㅜ	u	ㅢ	Ui	ㅠ	ju	ㅙ, ㅞ	we
	ㅡ	U			ㅖ, ㅒ	je	ㅘ	wa
묵음	C							

표 4. 모음과 묵음의 기호 목록

3.2 음성 분할을 위한 단위 선정

레이블링 단위로 선정된 유사음소의 목록 중 유음의 폐쇄음화, 마찰음의 유성음화, 파열음의 불파음화와 'ㄱ, ㄷ, ㅂ'을 제외한 파열음 폐쇄구간, 기식구간의 유성음화를 제외한 68개의 유사음소 그리고, 1개의 묵음을 포함하여 총 69개의 음소단위를 분할을 위한 단위로 선정하였다.

선정된 68개의 유사음소를 중점을 기준으로 나누어서 demiphone으로 확장하고, 음소의 길이가 120ms이상인 앞에서 설명한 방법으로 경우에는 3등분으로 분할하였다.

출현빈도수가 낮은 음소의 경우 훈련량이 충분하지 않아 훈련에 어려움이 있으므로, [표 5.]에서처럼 음소를 17개의 전·후 문맥정보로 클러스트링하여 각각의 단위로 사용하였다.

음소분류	기호	음소분류	기호
파열음의 폐쇄구간	sS	파찰음의 폐쇄구간	sA
파열음 폐쇄구간의 유성음화	sVS	파찰음의 폐쇄구간의 유성음화	sVA
파열음의 파열구간	bS	파찰음의 마찰구간	bA
파열음의 파열구간의 유성음화	bVS	파찰음의 마찰구간의 유성음화	bVA
마찰음	F	비음	N
성문마찰음	hF	모음	V
유음	L	이중모음	yV
유음의 공명음화	RL	묵음	sil
유음의 탄설음화	tL		

표 5 문맥정보를 위한 음소 분류

3.3 시스템의 훈련

시스템의 훈련에 사용한 음성 데이터베이스는 PBW 452 단어 데이터베이스에서 레이블링된 남성화자 30명분을 사용하였다. PBW 음성 데이터베이스는 방음 부스에서 Senheizer HMD224X를 사용하여 녹음되었으며, DAT(Digital Audio Tape)에 저장되었다. AD/DA 변환은 KAY CSL 4300B를 이용하여 16kHz로 Sampling하고 16Bits로 양자화되었다.

음성 분석은 10ms단위의 Hamming 윈도우를 5ms 간격으로 이동시키면서 분석하고, 특징 파라미터로는 12차의 MFCC (Mel-frequency cepstrum Coefficient)과 MFCC의 시간축 미분값, 그리고 정규화된 에너지와 그 미분치를 사용하였으며, 각 특징 파라미터들에 가중치를 주고, 각기 독립적인 벡터로 사용하였다. 특징파라미터에 부여한 가중비율은 MFCC와 MFCC의 시간축 미분값, 정규화된 에너지를 각각, 5, 3, 2의 비율로 적용하였다.

음소 모델의 확률분포모델로는 연속확률 분포를 사용하였고, 모델의 형태는 [그림 4.]와 같은 도약경로가 존재하지 않는 5상태7천이를 가지는 left-right 모델로 각 상태당 3개의 mixture를 사용하였다.

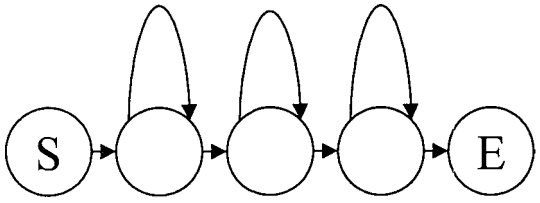


그림 4. 5상태 7천이를 가지는 left-right모델

HMM 모델을 초기화하기 위하여 Viterbi 알고리즘을 이용하여 HMM 모델을 초기화한 후 Baum-Welch 알고리즘을 이용하여 초기화된 HMM 모델을 훈련하였다

4. 실험 및 결과

본 논문에서 개선된 demiphone단위의 성능을 평가하기 위한 실험은 자동 레이블링 시스템을 구축하여 수작업으로 레이블링한 결과와 경계오차를 비교하는 과정을 통하여 수행하였다.

레이블링 시스템의 훈련은 PBW(Phonetically Ballanced Word) 452단어의 남성화자 30명분의 데이터를 demiphone과 개선된 demiphone으로 확장하여 사용하였다.

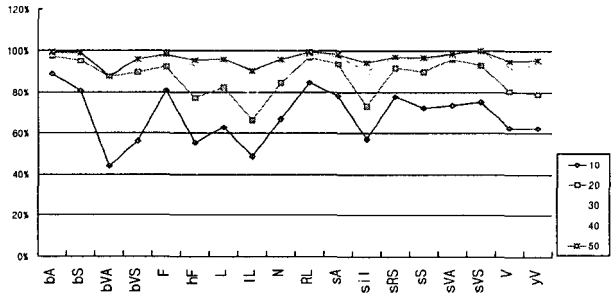
시스템의 평가를 위해 훈련에 사용하지 않은 PBW 452단어의 남성화자 4명분의 데이터에 대해서 수행하였다. 평가 결과는 수작업으로 레이블링된 결과와 자동 레이블링 시스템에 의해 레이블링된 결과를 경계의 위치가 벗어난 정도를 10ms에서 50ms까지 10ms단위로 경계 인식률을 비교하였다.

[표 6]은 음소의 지속시간이 120ms이상되는 음소의 경계인식률을 비교한 것이다. 지속시간이 120ms 이상인 총 10730개의 음소에 대해 개선된 demiphone이 전체적으로 성능이 우수함을 볼 수 있다.

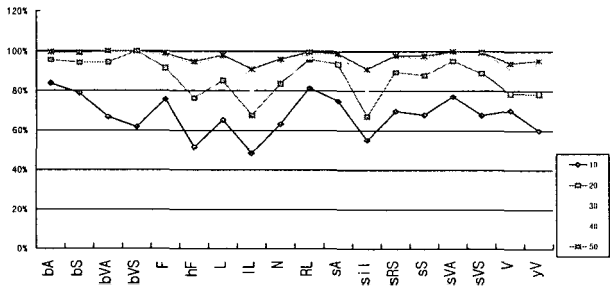
	demiphone	개선된 demiphone
10ms이하	61.26 %	62.09 %
20ms이하	74.63 %	79.35 %
30ms이하	82.39 %	87.49 %
40ms이하	86.83 %	91.81 %
50ms이하	89.60 %	94.76 %

표 6. 지속시간 120ms이상 음소의 인식률 비교

[그림 5]는 [표 5]의 문맥 정보별로 분류한 음소 그룹별 인식률 결과이다.



a) 기존 demiphone의 음소그룹별 인식률 비교



b) 개선된 diphone의 음소 그룹별 인식률 비교

[그림 6] 음소 그룹별 인식률 비교

[표 7]은 기존의 demiphone단위와 개선된 demiphone단위의 전체 인식률을 비교한 것이다. 개선된 demiphone단위가 기존의 단위에 비하여 10ms이하에서 1.65%, 20ms이하에서 1.02% 의 성능향상을 보이고 있다.

	demiphone	개선된 demiphone
10ms이하	67.44 %	69.09 %
20ms이하	84.30 %	85.32 %
30ms이하	89.98 %	90.66 %
40ms이하	93.10 %	93.88 %
50ms이하	95.29 %	96.01 %

표 7. 전체 인식률 비교

5. 결론

.본 논문에서는 기존의 demiphone단위를 개선하기 위하여 음소의 지속시간이 120ms이상일 경우에 음소의 천이구간의 특성을 잘 나타낼 수 있도록 하기 위하여, 음소의 앞뒤구간 각각60ms를 전반음소와 후반음소로 나누고, 나머지 안정구간을 별도의 모델로 구성 하였다.

실험결과, 기존의 반음소 단위에 비하여 10ms에서 69.09%로 1.65%, 20ms에서 85.32%로 1.02%의 성능향상을 가져왔다.

< 참 고 문 헌 >

[1] 김태환, 김봉완, 박순철, 이용주, "문맥종속 반음소단위 모델을 이용한 자동 음소분할 및 레이블링 시스템의 구현", 한국음향학회, 17권 2호, 1998.11.

[2] B. Eisen, H. G. Tillman, and C. Draxler, Consistency of judgments in manual labeling of phonetic segments: The distinction between clear and unclear cases, Proc of the ICSLP (Banff), 1992, pp. 871-874.

[3] 김종진, 김봉완, 이용주, 한국어 음성데이터베

이스 구축을 위한 한국어 레이블링 기준에 관한 연구, 제 13 회 음성통신 및 신호처리 워크샵 논문집, KSCSP '96 13권 1호, PP. 250-255., 1996.8.

[4] O.Mella, D.Fohr, "Semi-Automatic Phonetic Labelling of Large Corpora," Eurospeech97, pp.1732, 1997

[5] Andreas Kipp, Maria-Barbara Wesenick, Florian Schiel,"Automatic Detection and Segmentation of Pronunciation Variants in German Speech Corpora,"

[6] Ryszard Gubrynowics, Adan Wrzoskowics, "Labeller -A System of Automatic Labelling of Speech Continuous Signal," Eurospeech '93 pp.297, 1993.

[7] 성종모, 김형순 외 2, 한국 음성 데이터베이스 구축을 위한 반자동 음성분할 및 레이블링 시스템 구현, 제 13 회 음성통신 및 신호처리 워크샵 논문집, KSCSP '96 13권 1호, PP. 161-166., 1996.8.

[8] José B. Mariño, Albino Nogueiras, Antonio Bonafonte, "The Demi- phone: An efficient subword unit for continuous speech recognition",

[9] 이종락, "반음소 : 새로운 음성합성 및 인식단위", 제 10회 음성통신 및 신호처리 워크샵 논문집, pp. 208-212, 1993.