K-ToBI 기호에 준한 F0 곡선 생성 알고리듬*

이용주**, 이숙향***, 김종진**, 고현주***, 김영일**, 김상훈****, 이정철****

<차 례>

<Abstract>

A computational algorithm for F0 contour generation in Korean developed with prosodically labeled databases using K-ToBI system

Yong-Ju Lee et al.

This study describes an algorithm for the F0 contour generation system for Korean sentences and its evaluation results. 400 K-ToBI labeled utterances were used which were read by one male and one female announcers. F0 contour generation system uses two classification trees for prediction of K-ToBI labels for input text and 11 regression trees for prediction of F0 values for the labels. Evaluation results of the system showed 77.2% prediction accuracy for prediction of IP boundaries and 72.0% prediction accuracy for AP boundaries. Information of voicing and duration of the segments was not changed for F0 contour generation and its evaluation. Evaluation results showed 23.5Hz RMS error and 0.55 correlation coefficient in F0 generation experiment using labelling information from the original speech data.

* 본 연구는 원광대학교 일반과제 교비지원 및 한국전자통신연구원의 지원 하에 이루어진 것임.
** 원광대학교 컴퓨터공학과, *** 원광대학교 영어영문학과, **** 한국전자통신연구원

1. 서론

음성합성 시스템이 단순히 글자정보를 소리정보로 변환하는 기능만을 가지고선 그 역할을 다 하였다 할 수 없다. 우리가 일상생활에서 사용하는 음성신호 속에서는 텍스트 형태로 표현할 수 있는 정보뿐만 아니라 화자의 감정상태, 화자의 청자에 대한 의도, 요구 등 다양한 의미를 내재하고 있다. 이러한 음성신호의 특성을 많은 언어학/음성학 연구가들은 분절자질과 운율자질로 나눠 고찰한다. 그러므로, 음성 합성 시스템이 인간의 음성통신 수단의 대행자가 되기 위해선 단순히 분절자질의 자동 생성뿐 아니라 인간이 사용하는 운율자질을 최대로 모의할 수 있어야 한다.

본 연구의 최종목적은 한국어에 대한 자연스러운 합성음을 생성할 수 있는 합성기의 개발이다. 이를 위해 한국어에 적합한 분절자질과 운율자질을 모델링과 대량의 음성 DB로부터 최적의 합성 단위를 선택할 수 있는 알고리듬이 필요하다. 입력 문장에 대한 최적의 합성 유닛을 음성DB로부터 선정할 수 있도록 최적 합성음에 대한 템플릿을 제공하는 Target value 생성 모듈이 필수적이며, 합성음의 품질을 좌우하게 된다. Target value 생성 모듈은 정교한 지속시간 예측 모델, F0곡선 예측모델 그리고 에너지 곡선 예측 모델로 구성된다.

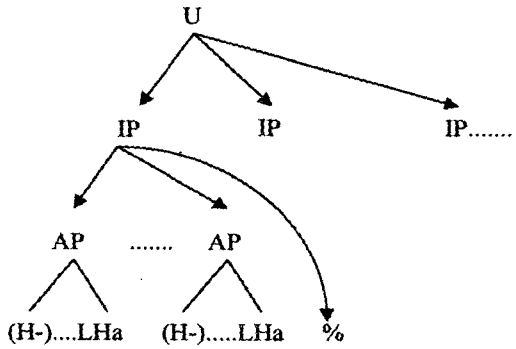
본 연구에서는 Target value 생성 시스템의 일부인 F0 곡선에 대한 Target value 생성 시스템을 개발하기 위해, K-ToBI 레이블링 된 음성DB와 K-ToBI에 기술된 한국어 문장의 운율구조에 기반을 둔 F0곡선 생성 시스템 개발을 시도하였으며 이에 대한 성능을 평가하였다. 개발된 F0곡선 생성 시스템은 합성 시스템의 전반부에서 자연스러운 합성단위를 추출하기 위한 Target F0 생성 모듈로 활용하고, 종단에서 합성음에 대한 F0곡선 수정 모듈로 활용할 예정이다.

2. F0 곡선 예측 모델

2.1 F0 곡선 예측 모델을 위한 기본 가정[3]

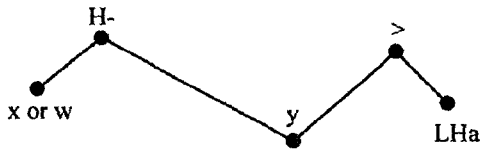
본 연구에서는 한국어 문장의 운율구조에 대해서 [1][2][3]의 기본가정과 [3]의 K-ToBI 레이블링에 관한 연구에서 제시된 계층적 운율구조 표기법에 기반을 두고 입력 문장에 대한 K-ToBI기호의 예측 및 F0 값을 예측할 수 있는 CART 결정 트리를

구현하였다. K-ToBI 기호를 이용한 한국어 문장에 대한 운율구조의 계층적 표현방식은 <그림 1>과 같다.



<그림 1> 한국어 문장의 계층적 운율구조[3]

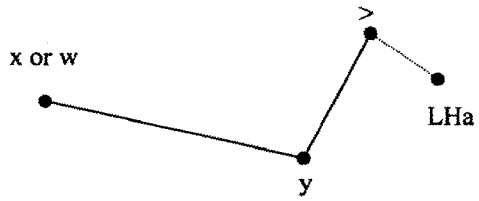
하나의 발화문(U)은 하나 이상의 억양구(IP)로 구성되며 하나의 억양구는 하나 이상의 악센트구(AP)로 구성된다. 억양구는 오른쪽 끝에 경계성조(예:H%)가 오며 악센트구는 마지막 음절에 오름조(LHa)가 나타난다.



<그림 2> 한국어 악센트 구 표준패턴.

또한 한국어 악센트구에 대한 표준적인 톤 이벤트 구조는 <그림 2>와 같이 가정하였으며, 그 특성은 [3][4]의 분석결과를 활용하였다. 여기서 'x'는 IP 처음에 나타나는 AP의 F0시작점을, 'w'는 IP 처음을 제외한 나머지 AP의 F0시작점, 'y'는 LHa의 F0시작점을 의미한다.

그러나 악센트구를 구성하는 음절수에 따라 'H-'는 선택적으로 나타나므로 <그림 2>는 <그림 3>과 같은 변이형으로 실현될 수 있다.



<그림 3> 한국어 악센트 구 표준패턴의 변이형

2.2 음성 데이터

본 연구에 사용된 음성 데이터는 아마추어 아나운서 남녀 각 1명이 발성한 총 2000문장중 K-ToBI 레이블링된 400문장(남녀 각 200 문장)을 이용하였다. 음성 데이터의 문장구성은 라디오 뉴스문 등에서 추출된 것이며, 문장의 길이와 문법적 구조는 별도로 고려되지 않았으며 복잡도는 다양하다.

2.3 F0 곡선 예측 모델링 절차

입력문장에 대한 F0곡선 생성 과정은 다음과 같은 절차를 따른다.

① 입력문장에 대한 IP 경계 어절 예측
② IP 경계 어절 정보를 이용한 AP 경계 어절 예측
③ AP 내 x or w, H-, y> 음절 예측
④ 예측된 K-ToBI기호에 대한 음절 내 실현위치 예측
⑤ 각 기호별 F0값 예측
⑥ Spline알고리즘을 이용한 F0곡선 생성

F0곡선 생성 알고리즘은 입력된 문장에 대한 K-ToBI 기호를 음절단위로 할당하는 기호예측모델과 기호에 대한 실제 F0값을 예측하는 모델로 구성된다. 억양구와 엑센트구 기호는 어절말 음절단위로 예측하였으며, 악센트구내 기호에 대해서는 어절내 음절단위로 예측하였다.

2.4 K-ToBI 기호 예측 모델

2.4.1 억양구 기호 예측

억양구 경계 어절을 예측하기 위해 다음과 같은 10개의 특징 파라미터를 선정하였으며, 이를 이용해 CART 훈련 과정을 거쳐 억양구 경계 어절을 예측할 수 있는 Classification Tree를 생성하였다. 억양구 기호는 억양구 경계 어절의 마지막 어절에 할당하였다. 억양구 예측에 사용된 특징 파라미터와 각 특징파라미터의 억양구 경계 예측에 대한 기여도는 표 1과 같다.

<표 1> 억양구 경계 예측 특징 파라미터의 기여도

특징 파라미터	기여도
억양구 후보 어절의 품사	100.000
억양구 후보 이후 음절 수	93.171
억양구 후보 이후 어절 수	90.814
억양구 후보 어절 품사의 출현빈도 유형	82.164
억양구 후보 앞 어절의 품사	28.429
억양구 후보 앞 앞 어절의 품사	12.468
문장의 음절 수	2.226
억양구 후보 다음 어절의 품사	2.137
억양구 후보 이전 어절 수	1.887
문장의 어절 수	1.858
억양구 이전의 음절 수	1.232

억양구 타입은 디폴트로 HL% 타입으로 설정하였다. HL% 유형은 전체 훈련데이터에서 출현한 억양구중 48.1%를 차지하였으며 이로 인해 CART훈련을 통해 생성한 결정 트리가 HL%이외의 다른 억양구 타입은 예측하지 못하였기 때문이다. 이는 현재의 음성DB가 라디오 뉴스문에서 추출된 문장을 대상으로 하였으므로 모두 평서형이었기 때문이다. 이러한 단점을 보완하기 위해 이 부분은 추후 K-ToBI 훈련 데이터의 확장 시 다양한 문장 타입을 고려하여 보강하도록 할 계획이다.

2.4.2 악센트구 기호 예측

악센트구 경계 어절을 예측하기 위해 13개의 특징 파라미터를 선정하였으며, 이를 이용해 CART훈련 과정을 거쳐 악센트구 경계 어절을 예측할 수 있는 Classification Tree를 생성하였다. 악센트구 기호는 악센트구 경계 어절의 마지막 어절에 할당하였다. 악센트구 예측에 사용된 특징 파리미터와 각 파라미터별 기여도는 <표 2>와 같다.

<표 2> 악센트 구 경계 결정 특징파라미터의 기여도

특징 파라미터	기여도
AP후보가 속한 IP내 선행 AP수	100.000
선행 AP로부터 AP후보까지의 음절 수	87.885
선행 AP로부터 AP후보까지의 어절 수	24.797
악센트 구 후보 어절의 품사	16.263
악센트 구 후보 앞 어절의 품사	14.921
악센트 구 후보 뒤 어절의 품사	12.121
악센트 구 후보 다음 다음 어절의 품사	10.786
악센트 구 후보 앞 앞 어절의 품사	6.568

악센트구 타입은 한국어의 계층적 운율구조에 준하여 디폴트로 'LHa'로 선정하였다.

2.4.3 악센트구내 x or w, H-,y,> 기호 예측

'x' or 'w', 'H-', 'y', '>' 의 악센트구내 음절위치는 [4]의 연구의 통계적 결과를 바탕으로 설정하였다. 'x'기호는 억양구의 처음 악센트구의 첫 음절에 할당하였으며, 억양구의 처음 악센트구를 제외한 나머지 악센트구의 첫 음절에는 'w'를 할당하였다. 악센트구의 'H-' 유무는 악센트구의 음절수에 따라 결정하였다. 악센트구의 음절수가 4개 이상인 경우, 96.9%가 'H-'를 가지는 악센트구로 나타났으며, 음절수가 4개 이하인 악센트구의 86.2%가 'H-'를 가지지 않는 악센트구로 나타났다. 그러므로, 기호 예측 시 악센트구가 4개 이상인 경우에는 'H-'가 있는 악센트구로, 4개 이하인 악센트구는 'H-'를 가지지 않는 악센트구로 결정하였다. 'H-'가 있는 악센트 구에서 'H-'의 위치는 해

당 규칙을 만족하는 악센트 구의 첫 번째 음절이나 두 번째 음절에 할당하였다.

- o 만일 첫째 음절이 super heavy이거나 heavy이면 'H-'를 그 음절에 실현시켜라.
- o 만일 첫째 음절이 light이면, 'H-'를 둘째 음절에 실현시켜라.
- o 만일 첫째 음절이 heavy이고 둘째 음절이 superheavy이면, 둘째 음절에 'H-'를 실현시켜라.

- o Super heavy:CV(C), onset C=+asp(ㅍ,ㅌ,ㅋ,ㅅ) 또는 +fortis
- o Heavy:(C)VC, onset C=-asp, -fortis
- o Light:(C)V, onset C=-asp, -fortis

'y'는 'H-'를 가지는 악센트구에 대해선 악센트구의 뒤에서 2번째 음절에 할당하였으며, 'H-'를 가지지 않는 악센트구에 대해선 뒤에서 3번째 음절에 할당하였다. '>'는 악센트구의 마지막 음절에 할당하였다.

2.4.4 음절내 톤 이벤트의 실현위치

<표 3> 음절의 유성음 구간내 K-ToBI 기호의 실현위치

K-ToBI 기호	음절내 톤 기호의 실현위치	
x	0.441	0.292
w	0.478	0.319
H-	0.535	0.212
y	0.585	0.343
>	0.510	0.240
LHa	0.715	0.444
(x)H-	0.531	0.188
(w)H-	0.520	0.221
(x)y	0.351	0.263
(w)y	0.593	0.260
HL%	0.752	0.330

음절단위로 K-ToBI 기호가 예측되어 할당되면, 각 기호의 음절의 유성음 구간내 실제 실현위치를 설정하였다. 음절 내 실현위치는 각 기호들의 음절 내 실현위치 정보

를 다음 수식 1과 같은 방식으로 산출한 후 이를 통계처리 하여 <표 3>과 같이 얻었다. 여기서 D1은 음절의 유성음구간의 지속시간을 의미하여, D2는 유성음 구간 내 톤 이벤트의 실현 위치를 의미한다.

$$R = \frac{D_2}{D1} \leq 1.0 \qquad \text{수식 1}$$

2.5 K-ToBI기호에 대한 F0 예측 모델

기호예측모델에 의해 음절단위로 K-ToBI 기호가 예측되면 그 다음은 각각의 기호에 대한 F0를 예측한다. 각 기호에 대한 F0 예측 모델은 14개의 Regression Tree로 표현된다. F0값 예측을 위한 특징 파라미터 및 기여도는 <표 4>와 같다.

<표 4> K-ToBI 기호별 F0값(Hz) 예측 특징 파라미터 및 기여도

	선행IP수	후행IP수	IP내		AP내	
			선행 AP수	후행 AP수	음절수	선행 음절수
x	100.000	16.961	41.698	5.771	27.797	9.069
w	100.000	34.887	97.293	0.000	9.323	29.093
H-	65.615	77.666	100.000	30.158	25.489	1.451
y	90.594	27.325	100.000	86.390	0.683	0.000
>	100.000	28.684	53.248	45.170	13.188	10.847
LHa	100.100	23.174	46.549	1.159	0.151	0.000
(x)H-	100.000	13.194	18.056	2.431	0.000	0.000
(w)H-	82.171	49.096	100.000	0.000	55.814	3.618
(x)y	100.000	0.000	0.000	0.000	0.000	0.000
(w)y	100.000	0.000	0.000	0.000	13.663	0.000
HL%	100.000	0.000	0.000	0.000	0.000	0.000

3. F0 곡선 생성 모듈의 성능평가

F0곡선 생성 시스템의 성능을 평가하기 위해 2가지 방법으로 실험하였다. 첫 번째 실험에서는 기호예측 모듈과 F0값 예측 모듈을 이용하여 F0곡선을 생성한 후 이를 원음의 F0곡선과 비교하였다. 두 번째 실험에서는 F0곡선 예측 모듈만의 성능을 평가하기 위해 레이블링 전문가에 의해 레이블링 된 기호 정보를 이용한 F0값 예측모듈에 의한 F0곡선을 생성하였으며, 이를 원음의 F0곡선과 비교하였다. 그리고, 합성음에 대한 지속시간 및 에너지 곡선은 원음의 것을 그대로 사용하였다.

3.1 실험 I

실험1에서는 기호예측모듈에 의해 K-ToBI 레이블링 기호를 예측하고, 각 기호에 대한 F0값 예측 모듈을 통해 각 기호에 대한 F0값을 예측한 후 Spline곡선을 이용해 곡선화 하여 문장에 대한 F0곡선을 생성하였으며, 이를 원음의 F0곡선과 비교하였다.

기호예측모듈에서 억양구에 대한 경계 어절 예측 결과를 원음에 대한 레이블링 정보와 비교하여 억양구 예측모듈의 성능을 비교하였으며 그 결과는 <표 5>와 같다.

<표 5> 억양구 예측 모듈의 예측성능 평가표

	없다	있다
없다	85.9%	14.1%
있다	22.8%	77.2%

악센트구 경계 예측 모듈의 성능 역시 같은 방법으로 평가되었으며 그 결과는 <표 6>과 같다.

<표 6> 악센트 구 경계 예측 모듈의 성능 평가표

	없다	있다
없다	29.6%	70.4%
있다	28%	72%

실험 1에서는 예측된 K-ToBI 레이블링 기호의 순서가 원음의 순서와 다르므로 기호별 F0차를 비교할 수 없었다. 대신에, [5][6]에서 사용된 방법처럼, 원음의 F0곡선과 합성된 F0곡선에 대한 RMS에러와 상관계수를 구하였으며 그 결과는 표 7과 같다.

<표 7> F0예측 모델의 예측오차(RMS)

RMS 에러	상관계수
24.96Hz	.44

3.2 F0 예측 실험 2

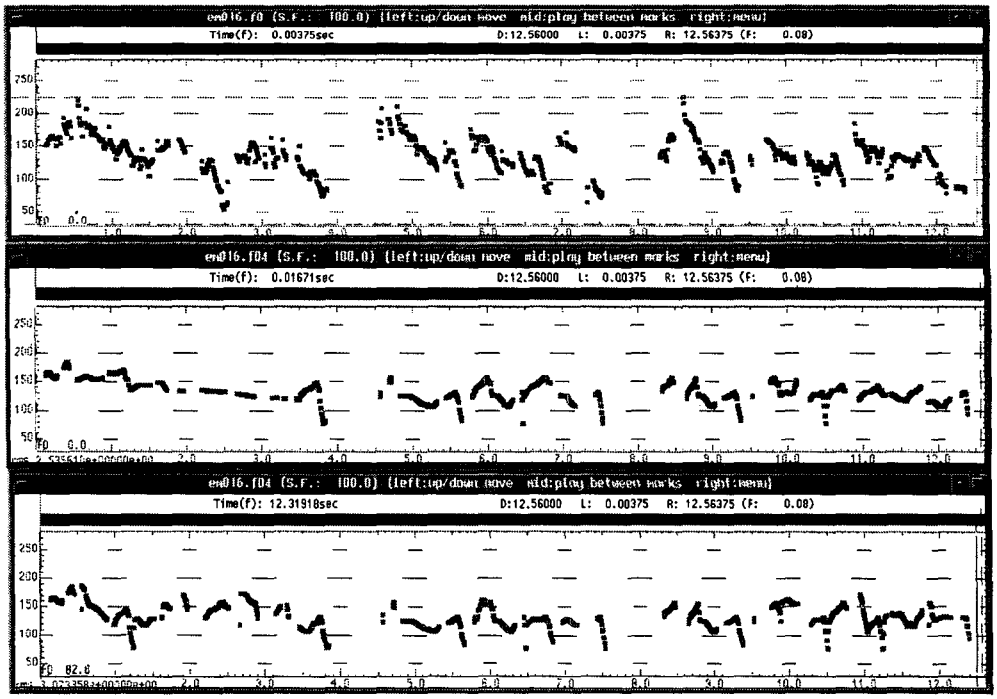
실험2에서는 각 기호에 대한 F0값 예측 성능을 평가하기 위해, K-ToBI 기호 예측 모듈 대신에 원음에 대한 K-ToBI 레이블링 정보를 입력받아 각 기호에 대한 F0값을 예측한 후 이를 곡선화 하여 합성된 F0곡선을 얻었다. 그리고 각 기호별로 원음의 F0값과 합성된 F0곡선에서의 F0값을 비교하여 각 기호별 RMS에러와 상관계수, 그리고 원음의 F0곡선과 합된 F0곡선에 대한 RMS에러와 상관계수를 구하였으며 그 결과는 표 8과 같다.

<표 8> 각 기호별 F0값 예측결과에 대한 평가결과(단위는 Hz이며, ( )는 표준편차를 의미함)

K-ToBI 기호	원음의 F0(Hz)	예측된 F0(Hz)	RMS 에러	상관계수
x	133(29)	130(12)	1.475	0.392
w '	130(22)	127(9)	1.428	0.337
H-	154(21)	152(12)	1.397	0.546
y	122(22)	121(13)	1.401	0.605
>	149(23)	147(15)	1.393	0.641
LHa	148(23)	144(13)	1.411	0.522
(x)H-	168(21)	173(9)	1.419	0.337
(w)H-	162(26)	157(15)	1.492	0.449
(x)y	127(23)	129(10)	1.381	0.446
(w)y	122(18)	121(10)	1.360	0.449
HL%	91(28)	90(0)	1.364	

본 논문의 실험1과 실험 2에 의해 얻어진 F0곡선과 원음의 F0곡선의 예는 <그림

4>와 같다.



<그림 4> (상) 원음의 F0곡선 (중)기호예측모듈과 F0값 예측 모듈에 의해 생성된 F0곡선
(하) 원음의 레이블링 정보와 F0값 예측 모듈에 의해 생성된 F0곡선.

4. 결론 및 향후 연구 계획

본 논문에서는 자연스러운 한국어 합성음을 생성하기 위해, K-ToBI 기호를 이용한 F0곡선 생성 알고리듬과 현재까지의 그 평가된 성능을 기술하였다. 음성 데이터는 남녀 각각이 200문장씩 발성한 총 400문장의 K-ToBI 레이블링 된 음성 시료를 이용하였으며, 이를 이용해 CART 훈련 과정을 거쳐 K-ToBI 레이블링 기호를 예측할 수 있는 Classification Tree와 기호에 대한 F0값을 예측할 수 있는 11개의 Regression Tree를 생성하였으며 그 성능을 평가하였다.

현재까지 개발된 시스템은 여러 가지 문제점을 가지고 있다. 첫째는 훈련 데이터

의 부족으로 인한 모델의 안정성의 부족이다. 이는 현재 계속 진행되고 있는 800문장의 전문가 레이블링이 종료되면, 이를 도입하여 추가 훈련하고자 한다. 둘째는 보다 정확한 IP경계 어절의 예측을 위해, IP예측모델에 Pause 예측모델을 추가하고자 한다.

<참고문헌>

[1] Beckman, M. E. and Jun, S.-A.(1996), *K-ToBI(Korean ToBI) Labeling Conventions version 2.1,* Revised Nov.

[2] Jun, S.-A(1993), The Phonetics and Phonology of Korean prosody: Ph. D. Dissertation, The Ohio State Univ.

[3] 이용주, 이숙향 외(1997), K-ToBI기호에 준한 F0 contours 생성 알고리듬 연구, 한국전자통신연구원 최종보고서, 원광대학교.

[4] Jong-jin Kim, Sook-hyang Lee, Hyun-ju Ko, Yong-ju Lee, Sang-hun Kim, Jung-cheol Lee(1997), An Analysis of some prosodic aspects of Korean utterances using K-ToBI labelling system, *Proc. ICSP, Vol. 1,* 87-91.

[5] Kenneth N., Ross, Modeling of Intonation for Speech Synthesis(1995), Ph.D. Dissertation, Boston Univ.

[6] Alan W,. Black, Andrew J Hunt(1996), Generating F0 Contours from ToBI lables using linear regression, *Proc. ICSLP , Vol. 3,* 1385-1388.

접수일자: 1998년 11월 26일

게재결정: 1998년 12월 22일

▶ 이용주(Yong-Ju Lee)
주소: 전라북도 익산시 신용동 344-2
소속: 원광대학교 컴퓨터공학과
전화: 0653) 850-6885
e-mail: yjlee@wonnms.wonkwang.ac.kr

▶ 이숙향(Sook-hyang Lee)
주소: 전라북도 익산시 신용동 344-2
소속: 원광대학교 영어영문학과
전화: 0653) 850-6913
e-mail: shlee@wonnms.wonkwang.ac.kr

▶ 김종진
주소: 전라북도 익산시 신용동 344-2
소속: 원광대학교 컴퓨터공학과 휴먼인터페이스연구실
전화: 0653) 850-6885
e-mail: jjkim@speech.wonkwang.ac.kr

▶ 고현주
주소: 전라북도 익산시 신용동 344-2
소속: 원광대학교 영어영문학과
전화: 0653) 850-6913
e-mail: hjko@gaebyok.wonkwang.ac.kr

▶ 김영일
주소: 전라북도 익산시 신용동 344-2
소속: 원광대학교 컴퓨터공학과 휴먼인터페이스연구실
전화: 0653) 850-6885
e-mail: yikim@speech.wonkwang.ac.kr

▶ 김상훈
주소: 대전시 유성구 사서함 106호
소속: 한국전자통신연구원 음성언어연구실
전화: 042) 860-5259
e-mail: ksh@zenith.etri.re.kr

▶ 이정철
주소: 대전시 유성구 사서함 106호
소속: 한국전자통신연구원 음성언어연구실
전화: 042) 860-5259
e-mail: jclee@zenith.etri.re.kr