'89 韓國自動制御學術會議論文集 1989. 10. 27~28

A Rule-Based Recognition System for Korean Spoken Place Names

O* Won Kyu Choi, ** Fi Hyol Lee and *Kageo Akizuki

* Faculty of Science and Engineering, Waseda University
** Faculty of Engineering, Tokai University

A rule-based recognition system for Korean spoken place names using anti-formants which is analyzed by ARMA model is presented. The recognition system is composed of three parts; the extraction, the recognition and the recognition support. As a result of experiment, the recognition rates of city place names was 90.9%.

1. Introduction

Recently, expert systems are constructed in a wide field of industrial problems. Some researches on the expert system for speech recognition have been reported [1][2].

This paper describes a recognition system for Korean spoken place names based on rule-base. As Korean speech includes a lot of nasal sounds, which makes it rather difficult to recognize by AR(Auto Regressive) model [3]. Therefore we implemented anti-formant frequencies using ARMA (Auto Regressive Moving Average) model.

The recognition system is composed of three parts; the extraction, the recognition and the support. The rule-base which is used in recognition part is constructed on the basis of formant and anti-formant frequencies, cepstrum pitches, logarithmic energies and zero crossing rates.

Furthermore, when we construct the recognition system, it is used that forward reasoning by True-False tree structure and production of recognition mode by a combination of zero crossing rates at beginning syllable.

2. The recognition system of Korean spoken place names

The recognition system, KOSRES (Korean Speech REcognition System), composed of three parts; the feature extraction part, the recognition part and the recognition support part.

The extraction part extracts feature parameters of Korean spoken place names. The recognition part is composed of the syllable segmentation and recognition by the rules. The rules are obtained from the feature parameters by 2 males and 1 female based on human expert's knowledge. The recognition support part supports recognition, for example, a modification of rules, draw feature parameters and process of recognition, etc. Figure 1 illustrates the composition of the recognition system for Korean spoken place names, KOSRES.

2.1 The ARMA model

A speech production process is expressed as an inverse filter which input is a sound source signal and the output is a speech signal[4]. G(Z) is transfer function of speech production process, the output is known only and the input is unknown. Therefore we used an inverse filter for estimation of input. Transfer function H(Z) of the inverse filter takes the form of equation (1),

$$G(Z) = \frac{B(Z)}{A(Z)} = \frac{1}{H(Z)} \qquad (1)$$

where,

$$A(Z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}$$

$$B(Z) = 1 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m}$$

Further, the output of the inverse filter is assumed to be an unit impulse.

$$U_k = S_k + h_1 S_{k-1} + \cdots + h_p S_{k-p} + V_k \qquad (2)$$

where, $h_1, h_2, \ldots, h_p$ are impulse responses of the inverse filter and V_k is sample value of error.

The matrix expression of equation (2) is as follows:

$$\mathbf{U} = \mathbf{s}_0 + \mathbf{s}\,\mathbf{h} + \mathbf{V} \qquad (3)$$

A least square estimate of h is given by

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} = \begin{bmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ r_1 & r_0 & \cdots & r_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & \cdots & r_0 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix} \qquad (4)$$

where, r_i is auto-correlation function of speech signal.

Furthermore, the relation between the transfer function H(Z) and the impulse responses is as follows:

$$H(Z) = \frac{A(Z)}{B(Z)} = 1 + h_1 z^{-1} + h_2 z^{-2} + \cdots + h_p z^{-p} \qquad (5)$$

That is to say,

$$\sum_{j=0}^{i} b_j h_{i-j} = \begin{cases} a_i, & i = 0, 1, \cdots, n \\ 0, & i > n \end{cases} \qquad (6)$$

where

$a_0 = b_0 = h_0 = 1 \quad b_i = 0, \; i > m$

Thus the parameters a_n, b_m of equation (1) are estimated using equations (7),(8) and spectral envelope is obtained from the frequency response of the transfer function G(Z).

$$\mathbf{h} = \mathbf{H}\,\boldsymbol{\theta} \qquad (7)$$

$$\mathbf{h} = (h_1, h_2, \cdots, h_p)^T$$

$$\boldsymbol{\theta} = (-b_1, -b_2, \cdots, -b_m, a_1, a_2, \cdots, a_n)^T$$

$$\boldsymbol{\theta} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{h} \qquad (8)$$

Using the solution of equation (9), formant and anti-formant frequencies and bandwidths are given as equation (10).

$$1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n} = 0$$
$$1 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m} = 0 \qquad (9)$$

$$F_i = |\lambda_i| / 2\pi T$$
$$B_i = -\log \gamma_i / \pi T \qquad (10)$$

where T is sampling period.

Formant and anti-formant frequencies obtained by this method are shown in Figure 2.

2.2 T - F tree structure

When we construct a recognition system based on rule-base, we used to make on AND-OR tree for reasoning. But this method takes time too long.

In this paper, True-False (T-F) tree structure is used, which is represented by Prolog language as shown in figure 3.

2.3 The feature parameter extraction part

Formant and anti-formant frequencies, cepstrum pitches, logarithmic energies and zero crossing rates are feature parameters in this system as shown in Figure 4.

We extract several noticeable features characterizing each spoken place names among which the following three features are the most important:

<1> Intensities of the pitch in some frames.
<2> Intensities of the logarithmic energy and the number of zero crossing rates.
<3> A change of formant frequencies,for example, an increase, a decrease, a distance of first and second frequencies, and the existence of anti-formant frequencies.

2.4 The recognition part

We can find remarkable changes of logarithmic energies, the formant frequencies and zero crossing rates on each boundaries of the syllable.
Using the knowledge of the noticeable feature of the parameters, speech signals are divided into the syllable units. The rules for segmentation are defined as follows:

if__ < logarithmic energy more than or equal to 10[dB] exists in two consecutive frames >
then__ < S1 is start frame at 1st syllable >

if__ < the frame is between (S1+9) and end frame >
and__ < logarithmic energy less than 10[dB] exists in two consecutive frames >
or__ < logarithmic energy is minimum >
then__ < S2 is end frame at 1st syllable >

if__ < the frame is between S2 and end frame >
and__ < logarithmic energy more than or equal to 10[dB] exists in three consecutive frames >
and__ < zero crossing rates less than 20 exists in two consecutive frames >
then__ < S3 is start frame at 2nd syllable >

if__ < the frame is between S3 and end frame >
and__ < logarithmic energy less than 10[dB] exists in two consecutive frames >
then__ < S4 is end frame at 2nd syllable >

The result of the segmentation rules, for example, Korean city place name [CHUNG - JU] is S1(1),S2(42),S3(43),S4(73).
The rule-base for the recognition is defined as follows:

if__ < beginning syllable by zero crossing rates is un voiced >
and__ < 1st formant has not F1DOWN >
and__ < anti-formant exists between 500 and 1500Hz continuously, and so on >
and__ < 2th formant is not constant >
then__ [CHUNG]

if__ < beginning syllable by zero crossing rates is un voiced >
and__ < 1st formant has not F1DOWN >
and__ < 2nd formant is decreasing >
then__ [JU]

Figure 5 shows the tree-structure for above rules. Its recognition rule-base applied in the recognition system are obtained from human experience knowledge. There are 64 pieces of rules in the KOSRES.

3. Experiment

The Korean spoken place names are shown in table 1.
Experimental data is spoken 30 place names by 8 speakers, 2 males and 1 female for training, 4 males and 1 female for testing. The input speech is sampled at 12.8[KHz] with 12[bit] A/D converter.

Each data length and frame length are 1[sec] and 20[msec], respectively. And input speech is analyzed by 18th order of AR and 9th order MA parts, and impulses are 40th with every frame. For the comparison, Table 2 shows the recognition results between AR model and ARMA model.

As a result of the experiment, the recognition rate of 90.9 % has been obtained. We could obtain 4.1% higher than AR model.

4. Conclusion

This paper describes a rule-based recognition system for Korean spoken place names using rule-base. As Korean speech includes a lot of nasal sounds, we implemented anti-formant frequencies by ARMA model. And the rule-base applied in the KOSRES is obtained from human experience knowledge.

As a result of the experiment, the recognition rates of Korean place names uttered by 4 males and 1 female speakers was 90.9%.

Reference

[1] Victor W. Zue, Ronald A.Cole, "Experiment on spectrogram reading", IEEE ICASSP, pp.116-119, 1979.

[2] R.Mizokuchi, K.Tsujino, O.Kakusho, "A continuous recognition system based on knowledge engineering techniques", IEEE ICASSP, pp.1221-1224, 1986.

[3] W.K.Choi, F.H.Lee, F.Ohkoshi, K.Akizuki, "The recognition for spoken words using anti-formant", SICE, pp.177-178, 1988.

[4] T.Fukabayashi, H.Suzuki, "Speech analysis by linear pole-zero model", Trans.IEICE, Vol.58-A, No.5, 1975.

[5] W.K.Choi,F.H.Lee,K.Akizuki, "Reserved formant tracking for Korean speech and application to recognition of digits", Trans.IEE Japan, Vol.108-C, No.10, 1988.

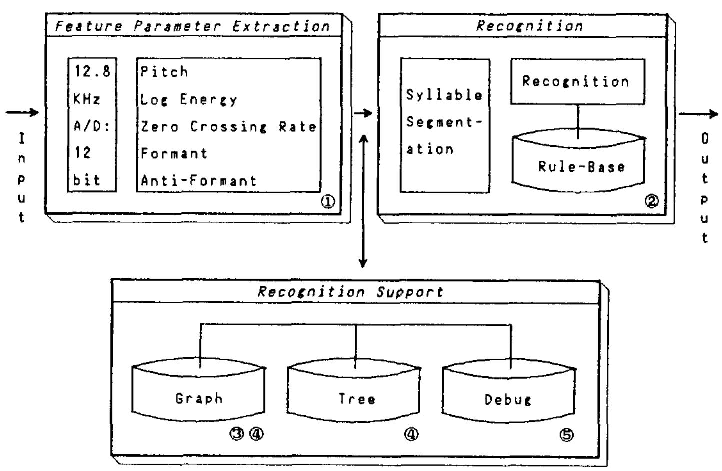


Figure 1. Composition of the KOSRES

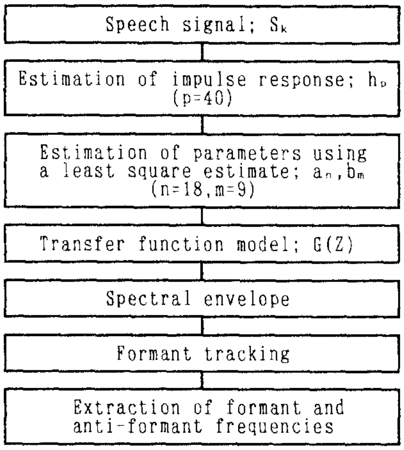
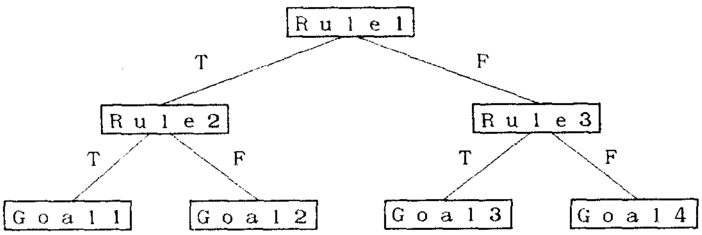


Figure 2.

Analytic method of formant and anti-formant frequencies based on an ARMA model



Rule expression (PROLOG)

```
class:-rule1,!,sub1.
class:-sub2.
sub1 :-rule2,!,goal1.
sub1 :-goal2.
sub2 :-rule3,!,goal3.
sub2 :-goal4.
```

Figure 3.

True-False tree structure

안양 (A-NYANG)	부천 (BU-CHEON)	창원 (CHANG-WON)
제천 (CHE-CHEON)	제주 (CHE-JU)	천안 (CHEO-NAN)
청주 (CHEONG-JU)	춘천 (CHUN-CHEON)	충주 (CHUNG-JU)
충무 (CHUNG-MU)	경주 (GYEONG-JU)	인천 (IN-CHEON)
이리 (I-RI)	전주 (JEON-JU)	김천 (KIM-CHEON)
구미 (KU-MI)	군산 (KUN-SAN)	광주 (KWANG-JU)
마산 (MA-SAN)	목포 (MOK-PO)	포항 (PO-HANG)
부산 (PU-SAN)	성남 (SEONG-NAM)	서울 (SEO-UL)
속초 (SOK-CHO)	수원 (SU-WON)	대구 (TAE-GU)
대전 (TAE-JEON)	원주 (WON-JU)	울산 (UL-SAN)

Table 1.

Korean spoken place names

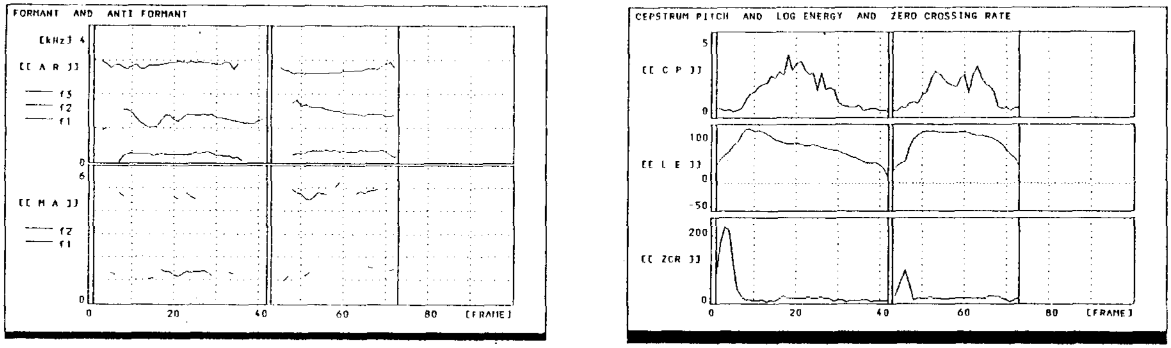


Figure 4. Analytic results of parameters

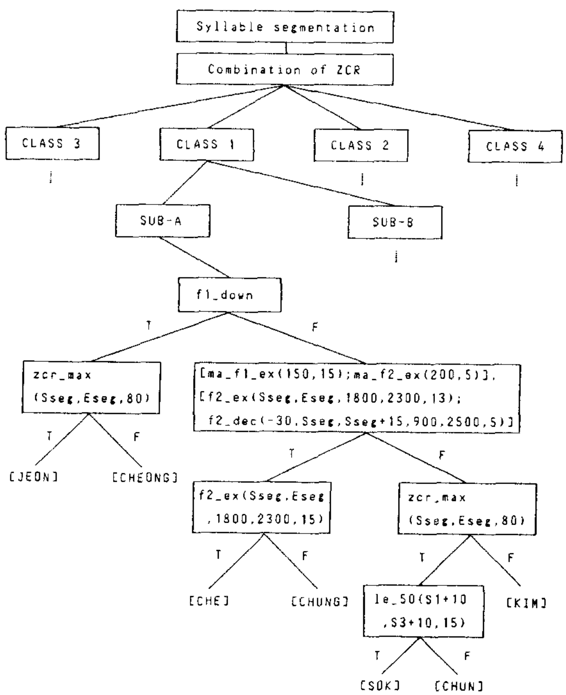


Figure 5.

A tree structure of the recognizing rule-base

Table 2. Recognition results

	MALES				FEMALE	TOTAL
	C	L	J	LS	P	
AR	100%	83.3%	84.6%	85.7%	80.4%	86.8%
MODEL	88.4%				80.4%	
ARMA	100%	94.4%	89.3%	88.5%	82.1%	90.9%
MODEL	93.1%				82.1%	