한국어 인식을 위한 알고리즘의 개발

°안 창 , 진 상 현 , 이 상 범
단국대학교 전자공학과

Development of the algorithm for Korean vowel recognition

Chang Ahn , Sang-Hyun Chin , Sang-Burm Rhee
Dankook University

ABSTRACT

A vowel is based on the recognition of a phoneme. Thus it is necessary for the programming of an algorithm to achieve the speech recognition in that case.

In this paper, cepstrum is used for a voiced-unvoiced decision and speech parameters are extracted by linear prediction coding. Using these parameters, a speech understanding algorithm has been developed.

I. 서 론

현재 음성 인식은 화자, 어휘, 발성 형태에 중점을 두고 연구되고 있으며, 독립단어나 연속단어의 발성 형태에 대한 파라미터 추출 방법과 음성 인식 장치의 개발도 활발히 진행되고 있다.[1)]

특히 음성 인식을 하기위한 파라미터 추출 방법에는 크게 시간 영역에서 음성 신호의 변이를 추적하여 음성간의 특징을 추출하는 방법이 있으며, 또 다른 방법은 시간영역 데이터를 주파수 영역으로 변환하여 시간 영역에서는 관찰할 수 없는 특징 파라미터를 찾아내기 위한 자기 상관 함수, 영교차율, 에너지비, FFT, LPC 등의 방법이 사용되고 있다.[2)]

본 논문에서는 특정 화자인 경우에 대하여 음소 단위의 음성 인식에 기초가 되는 알고리즘을 개발하고 단모음의 경우에 대하여 이를 적용하고자 하였다. 이 알고리즘은 몇개의 표본 신호에 대해 LPF, A/D, Window를 통과시켜 1 프레임 간의 스펙트럼을 분석하고, 이때 산출된 각각의 데이타를 표준 패턴으로하여 음성을 인식하고자 한다.

II. 음성 분석

음성의 분석은 음성에 포함된 특성, 즉 특징 파라미터 (parameter)를 추출하여 패턴에 따라 음성 신호를 인식하는 과정이다. 특징 파라미터 추출법에는 관측된 음성 신호의 시간적 변화를 그대로 이용하는 방법과 음성신호가 포함하고 있는 주파수 성분을 이용하는 방법이 있다.

시간 영역의 분석에는 상관함수에서 선형 예측 계수 (linear prediction coefficient)나 편자기 상관 계수 (partial autocorrelation coefficient)를 추출하거나 피치(pitch), 영교차율(zero crossing rate)[2,3)] 등을 구하는 것이고 주파수 영역의 분석으로는 스펙트럼 (spectrum)분석과 포먼트(formant)와 기본 주파수 추출[3,4)]등이 있으며 또한 예측 계수에서 스펙트럼을 구하는 방법등이 있다.

II-1. 선형 예측 분석

음성 신호의 파형을 관찰하면 음성파형의 이웃한 표본사이에 상관관계가 높음을 알 수 있다. 이와같은 관계를 간단한 선형 예측 형태로 표시하면 식(1)과 같다.

$$x_n = \alpha_1 x_{n-1} + \alpha_2 x_{n-2} + \cdots + \alpha_p x_{n-p} \quad \cdots\cdots (1)$$

식(1)은 음성 파형의 한 표본값을 과거 p 개의

표본값들의 선형 결합으로 예측할 수 있다는 것을 가정하고 있으며, 이때 각 표본에 곱하여지는 가중치 $\{\alpha_i, i = 1,2,...,p\}$ 를 선형 예측계수라 한다. 이와같이 선형 결합에 의하여 실제의 음성 신호값이 정확히 추정되며 실제의 신호값과 예측된 신호값과의 오차신호 e_n은 다음과같이 주어진다.

$$e_n = x_n - \hat{x}_n \quad \cdots\cdots (2)$$

이때, $\hat{x}_n$은 입력 음성 신호 표본 x_n에 대한 예측값으로

$$\hat{x}_n = -(\alpha_1 x_{n-1} + \alpha_2 x_{n-2} + \ldots + \alpha_p x_{n-p}) = -\sum_{i=1}^{p} \alpha_i x_{n-i} \quad \cdots\cdots (3)$$

와 같다.

선형 예측법에서는 예측계수 α_i를 구하는 것이 중요하며 예측 계수를 정확히 구하면 올바른 파형의 재생이 가능하다. 예측계수 α_i를 구하는 방법은 식(2)에서 구한 오차신호 e_n이 최소가 되도록 결정하면 된다. 어떤 구간의 오차신호 e_n의 제곱 E를 도입하여 E가 최소값을 갖도록 하여 구하면 오차신호의 제곱 E는 다음과 같다.

$$E = \sum_n e_n^2 = \sum_n (x_n + \sum_{i=1}^{p} \alpha_i x_{n-i})^2 \quad \cdots\cdots (4)$$

이 값은 다음 식이 만족될 때 최소화 된다.

$$\frac{\partial E}{\partial \alpha_i} = 0 \quad \cdots\cdots (5)$$

(단, $1 \le i \le p$)

선형 예측 부호화 (linear predictive coding, LPC)방식에는 예측오차를 최소화하는 방식에 따라 autocorrelation 방법과 covariance 방법으로 분류된다.[5,6] Autocorrelation 방법은 성도 filter의 안정도를 보장할 수 있는 장점이 있으므로 이를 이용하여 예측오차를 최소화 할 수 있다.

입력 음성 파형의 표본값에 대한 선형 예측 해석은 음성파형을 AR(autoregressive) process로 가정하고 이에 대한 all pole 시스템 모델을 구하는 과정이라 할 수 있으며 all pole 모델은 다음과 같다.

$$H(z) = \frac{1}{1 + \alpha_1 z^{-1} + \ldots + \alpha_p z^{-p}} \quad \cdots\cdots (6)$$

앞에서 기술한 선형 예측 계수를 구하는 방법은 시간축상에서 표본값들 사이의 상관계수에 의하여 시스템의 전달함수를 구하는 것, 즉 주파수 영역에서의 입력 여기신호의 주파수 스펙트럼 포락선 (spectrum envelope)을 결정하여 주는 것이라 할 수 있다.

II-2. 켑스트럼

신호대 잡음비(signal to noise ratio)가 매우 큰 경우에 음성 신호의 에너지는 주위 잡음에 비하여 현저한 차이를 보이지만 실제적으로는 대부분의 경우에 있어서 주변 잡음의 영향을 무시할 수 없다. 따라서 유성음과 무성음을 구별할 수 있는 방법으로 대수 스펙트럼(logarithmic spectrum)의 전력 스펙트럼(power spectrum)인 켑스트럼(cepstrum)[7]을 이용하였으며 켑스트럼의 정의는 식(7)과 같다.

$$C_k(q) \equiv \left| \int_0^{\omega_c} LOG|F_k(\omega)|^2 \cos\omega q dq \right|^2 \quad \cdots\cdots (7)$$

여기서 $F_k(\omega)$는 시간창처리를 거친 신호를 Fourier 변환한 함수이며, 즉

$$F_k(\omega) = \int_{-T_\omega}^{T_\omega} f(t) e^{-j\omega t} dt \quad \cdots\cdots (8)$$

(단, $\pm T_\omega$ 는 시간창의 상한과 하한)

와 같다.

유성음은 segment의 기본적인 주기에 대한 켑스트럼에서 피크(peak)가 존재한다. 그러나 무성음의 경우에는 피크가 나타나지 않으므로 유성음과 무성음의 구별이 간단해진다.

III. 알고리즘의 구성

표준 파라미터를 추출하기 위해서 우선 음성을 받아들여 이를 12 비트 양자화 레벨을 갖는 A/D 변환기로 표본화하였으며 이때 표본 주파수는 8KHz, 즉 1 프레임을 12.5 msec로 정하였다.

선형 예측 부호화에서 선형예측 계수와 반사계수를 autocorrelation을 이용하여 구한 결과를 표준 패턴 (reference pattern)으로 하였다. 그러나 실제 응용에 있어 예측 계수보다 이의 변형인 반사계수는 dynamic range 가 적어 ($|k_i| \leq 1$) 부호화하는데 있어서 비트수를 줄일 수 있다. 여기서 이들 계수를 이용하여 LPC 스펙트럼 포락선 (LPC spectrum envelope)을 구할 수 있고 peak-picking법에 의하여 포먼트(formant)를 추출할 수 있다.

본 논문의 알고리즘은 다음과 같다.

단계 1 : 음성 데이터를 받음

단계 2 : 시간창 처리

단계 3 : 대수 전력 스펙트럼을 구함

단계 4 : 켑스트럼을 구함

IF not exist 피크

THEN GO TO 단계 1

단계 5 : 선형 예측계수와 반사계수를 구함

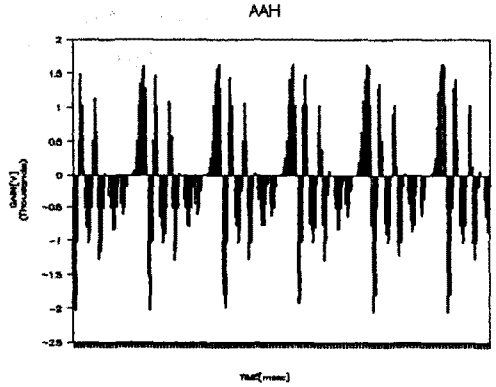
단계 6 : all pole 스펙트럼 포락선을 이용하여 포먼트 추출

단계 7 : IF exist 처리할 데이터

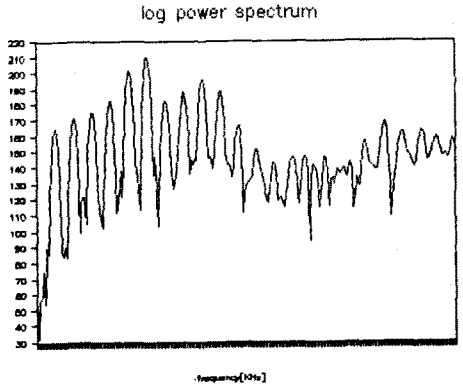
THEN GO TO 단계 1

ELSE GO TO pattern matching 단계

위의 알고리즘은 입력 데이터를 받아들여 Hanning window로 시간창 처리를 하였고 대수 전력 스펙트럼을 구하였다. /아/에 대한 입력 음성 신호와 대수 전력 스펙트럼을 그림 1에 보였다.



(a) 입력 음성 신호



(b) 대수 전력 스펙트럼

그림 1. /아/의 입력 음성 신호와 대수 전력 스펙트럼

켑스트럼을 구하는 단계 4의 상세한 사항을 가상코드(pseudo code)로 나타내면 다음과 같다.

```
procedure cepstrum;
  LN := 512
  DO I = 1,256
     J := LN + 1 - I
     X(I) := LOGPW(I)/20.
     Y(I) := 0.
     X(J) := X(I)
     Y(J) := Y(I)
  ENDDO
  CALL FFTP(X, Y, C, SI, NFFT, 9)
  print X
  return
END;
```

켑스트럼을 구하기 위하여 대수 전력 스펙트럼을 고속 Fourier 변환을 하였으며 이때 Fourier 변환을 위하여 FFT Pruning 프로그램[3,8]을 이용하였다.

단계 4에서 유성음으로 판정되면 단계 5에서는 선형예측계수와 반사계수를 구하기 위하여 subroutine AUTO[3]를 이용하였다. 이때 예측 차수는 10으로 하였다. 또한 단계 6에서는 all pole 스펙트럼 포락선을 구하여 포먼트를 추출하였다.

단계 5,6에 대한 가상코드는 다음과 같다.

```
procedure LPC;
  DO I = 1,LPCP
     A(I) := 0
     RC(I) := 0
  ENDDO
  CALL AUTO(LN2,SIG,LPCP,A,ALPHA,RC)
  DO I = 1,LPCP+1
     print RC,A
  ENDDO
  DO I = 1,LPCP+1
     X(I) := A(I)
     Y(I) := 0
  ENDDO
  DO I = LPCP+2,512
     X(I) := 0.
```

```
    Y(I) := 0.
  ENDDO
  CALL FFTP(X,Y,C,S,NFFT,5)
  SLP := 0
  FLG := 0
  LPCENV(0) := 0.
  DO I = 1,256
  T1 := ALOG(ALPHA)
  T2 := ALOG(X(I)**2+Y(I)**2)
  LPCENV(I) := 10.*(T1-T2)
     IF LPCENV(I) > LPCENV(I-1)
        THEN SLP := 1
        ELSE SLP := 0
     ENDIF
     IF SLP = 0
        THEN F := 3400.*(I-1)/256.
     ENDIF
     IF SLP = 0 AND FLG = 0
        THEN print I-1,F
     ENDIF
     IF SLP = 1
        THEN FLG := 0
        ELSE FLG := 1
     ENDIF
  ENDDO
END;
```

이러한 과정을 고급언어로 작성하였으며 입력 데이터는 한국어 단모음 /아/, /어/, /오/, /우/, /이/ 등으로 하였다. 단모음 /아/,/우/에 대한 단계 5의 결과를 표 1에 보였으며 단계 6에서 구한 all pole 스펙트럼 포락선을 대수 스펙트럼과 함께 그림 2에 보였다.

표 1. 선형예측계수와 반사계수

아		우	
RC	A	RC	A
-.56597270	1.00000000	-.68665740	1.00000000
.66595100	-.65534070	.19100530	-.70678730
.38587300	.19191070	.32659420	-.28883290
-.07781281	.86567760	.15887230	.35270290
.35391070	-.22485270	-.43782200	.51318040
.44216020	-.16286940	.24129880	-.42619420
-.19244220	.60597920	.45000860	-.14907960
-.11054930	-.02465672	.12117610	.25072510
.12716460	-.17785320	.20897560	-.00542119
.07488619	.07737549	-.07885462	.26340960
.00000000	.07488619	.00000000	-.07885462

A : 예측계수 RC : 반사계수

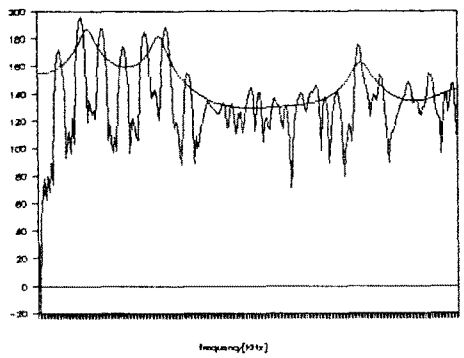


그림 2. /우/에 대한 스펙트럼과 포락선

IV. 결 론

음성을 인식하기 위한 파라미터 추출 방법은 시간영역과 주파수 영역에서 이루어질 수 있다. 이러한 음성 파라미터를 추출하기 위하여 켑스트럼으로 유성음과 무성음 판단을 하고 LPC를 이용하여 그 특징을 구별한다.

본 논문에서는 한국어 인식에 적당한 파라미터를 추출하기 위한 알고리즘을 구성하여 그 특정음에 대한 표준패턴을 구하였고 몇가지 유성음에 대하여 이를 적용하였다.

앞으로 불특정 화자인 경우에 대하여 피치와 포먼트사이의 상관성, 연령과 성별에 대한 변화를 고려하여 적합한 알고리즘을 구성해야 한다.

참고 문헌

[1] W. A. Lea, Trends in Speech Recognition, Englewood Cliffs, N. J., Prentice - Hall. Inc. 1980.

[2] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Englewood Cliffs, N. J., Prentice - Hall. Inc., 1978.

[3] J. D. Markel and A. H. Gray, Jr., Linear Prediction of Speech, Springer-Verlag, New York, 1976.

[4] S. S. Mc-Candless,"An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra," IEEE Trans. ASSP, vol.ASSP-22, No.2, pp.135-141, April 1974.

[5] J. D. Markel and A. H. Gray, Jr., "A Linear Prediction Vocoder Simulation Based upon the Autocorrelation Method," IEEE Trans. ASSP, vol.ASSP-22, No.2, pp.124-134, April 1974.

[6] J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, vol.63, No.4, pp.561-580, Apr. 1975.

[7] A. M. Noll, "Cepstrum Pitch Determination," J. Acoust. Soc. Amer., vol.41, pp.293-309, Feb. 1967.

[8] J. D. Markel, "FFT Pruning," IEEE Trans. on Audio and Electroacoustics, vol.AU-19, No.4, pp.305-311, Dec. 1971.

[9] J. L. Flanagan, et al., "Speech Coding," IEEE Trans. Comm., vol.COM-27, No.4, pp.710-737, Apr. 1979.

[10] J. Makhoul, "Spectral Linear Prediction: Properties and Applications," IEEE Trans. ASSP, vol.ASSP-23, No.3, pp.283-296, June 1975.

[11] J. D. Markel and A. H. Gray, Jr., "On Autocorrelation Equations as Applied to Speech Analysis," IEEE Trans. on Audio and Electroacoustics, vol.AU-21, No.2, pp.69-79, Apr. 1973.

[12] L. R. Rabiner, "The Acoustics, Speech, and Signal Processing Society - A Historic Perspective," IEEE ASSP Magazine, pp.4- Jan. 1984.

[13] B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Amer., vol.50, pp.637-655, Feb. 1971.