

한국어의 LSP 분석에 관한 연구

A study on the analysis of Korean vowels by the Line Spectrum Pair method

* 李 應 俊 (Lee, E. J.)

** 金 炯 來 (Kim, H. L.)

요 약

LSP 방식은 음성의 주파수 특성을 포함하는 공진 주파수를 낮은(Low) 부분과 높은(High) 부분의 주파수로 표시되는 선스펙트럼쌍(LSP) 계수를 구하는 방법이다.

본 논문은 LSP 방식을 사용하여 한국어의 기본 모음 7 개를 대상으로 하여 분석하고 LSP 계수를 구하는 Algorithm 을 개발하였으며 PARCOR 방식과 비교하였다.

실험 결과 LSP 방식의 연산량이 PARCO 방식의 연산량보다 약 1/2 정도로 적응을 알 수 있었고 Hardware 구성 시에 있어서도 경제적임을 알 수 있었다. 그리고 LSP는 계수 모음의 종류에 따라 각기 다른 공진 주파수, 대역폭을 나타내기 때문에 음성 합성이나 음성 인식 분야에 있어 기초 자료로 이용할 수 있을 것으로 사료된다.

ABSTRACT

The LSP(Line Spectrum Pair) method is to find out the LSP coefficients which are represented by the formant frequency of the speech signals as the low frequency and the high frequency.

In this paper, it was analyzed the seven vowels of Korean speech using the LSP method and was compared the LSP method with the PARCOR method.

As the result of this experiment, it was found that the amount of calculation with LSP method could decrease a half times comparing with PARCOR method.

And it was also turned out that the LSP method had the cost-down for the hardware.

*, ** 건국대학교 전자공학과 정희원

According to the kinds of Korean vowel, it was shown that the LSP coefficient had the distinctive frequency and bandwidth.

For that reason, the LSP method will be easily adapted and used basically to synthesize and to recognize the speech signals than other methods.

I. 서 론

디지털 신호 처리 기술의 발달되면서 음성 정보 처리 기술의 비약적인 발전을 보게 되었다. 특히 LPC(Linear Predictive Coding) 방법에서 발전된 PARCOR(Partial auto-correlation) 방식이나 LSP(Line Spectrum Pair) 방식등이 음성의 분석, 합성 방법으로 연구되어 왔다.

현재 가장 널리 사용되고 있는 음성의 분석과 합성 방식은 PARCOR 방식인데 PARCOR 방식은 LPC 방식에 비해 매초당 음성의 정보량을 9.6~4.8 Kbit 정도로 압축할 수 있는 장점을 가지고 있으나 정보량을 4.8Kbit 이하로 압축하려면 frame 간격을 보통 20ms 이상으로 하여야 하며 그 사이에 각 파라메타의 변화를 부드럽게 하기 위해 약 frame 간격의 $\frac{1}{4}$ 이나 $\frac{1}{2}$ 정도의 선형 보간을 필요로 한다. 그러나 이 선형 보간은 꼭 필요한 것이나 적절한 것이 아니며 이로 인하여 스펙트럼의 왜곡이 증대하고 음질의 저하를 가져오는 요인으로 지적되고 있다. 이것은 PARCOR 계수가 단일의 명확한 볼리랑에 대응하지 않고 복잡한 복잡 파라메타로 구성되어 있는 것이 원인으로 해석된다⁽³⁾.

본 논문에서는 음성의 특징 파라메타의 추출을 위하여 PARCOR 방식에 주파수 특성을 고려한 LSP(선스펙트럼대) 방식을 이용하여 한국어의 기본 다음을 대상으로 하여 주파수 스펙트럼대를 추출하였고 LSP 분석에 대한 Algorithm을 개발하였다. 또한 각 모음간의 주파수 특성을 고찰함으로써 모음에 대한 음성 합성, 인식의 기초, 자료로 삼고자 하

였다.

II. 본 론

1. LSP 방식의 개요

그림 1에서 보는 바와 같이 성대로 부터 입술, 콧구멍에 도달되는 음파의 통로를 성도라 하고, 성도는 성대로부터의 거리에 의해 결정되는 단면적과 같은 음향관이라고 할 수 있다. 성도의 단면적은 조음 기관의 운동에 따라 변화하며 어느 한 순간에 있어서는 정지한 성도 단면적을 갖는다.

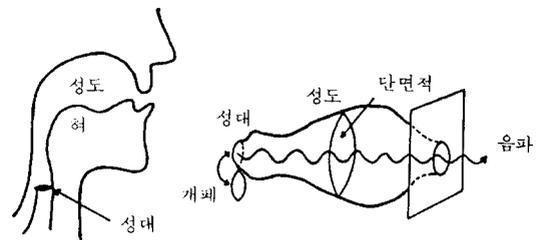


그림 1 성도의 음향관 모델
Sound tube model of the vocal tract.

이 음향관의 입술 측이 개방되어 있을 때 성도 측은 성대의 진동에 따라 열리고 닫힌다. 그러므로 이 음향관에 있어 입술 측과 성대 측의 경계 조건을 단순화하고 성대의 조건을 완전 개방과 완전 폐쇄라는 2개의 경계 조건을 모델화 할 수 있다. 그러나 이 음향관의 단면적이 길이에 따라 일정하지 않으므로 이를 통과하는 음파는 반사가 일어난다. 그러므로 음파의 전달함수는 반사계수로서 표현될 수 있으며 이 반사계수는 성도 단면적과 관계된다. 음파

의 공진 특성은 음파의 전달함수에 의해 결정되므로 음파의 반사계수 즉, 성도 단면적에 의해 결정될 수 있다. 음향관이 성도벽의 진동이나 점성 그리고 열의 전도에 의한 에너지 손실이 없는 무손실계라고 하면 완전개방($K_{P+1} = 1$)과 완전폐쇄($K_{P+1} = -1$)라는 2가지의 경계 조건에 따라 하나의 공진 특성이 그림 2와 같이 나타난다. 즉 공진 특성은 2가지의 경계 조건에 의해 2개의 주파수가 구해지며 이것을 선스펙트럼대(Line Spectrum Pair)라고 한다^{(3), (6)} 또 그림 2에서 보는 바와 같이 Q가 높은 곳에서의 LSP의 대역폭은 좁게 나타나고 Q가 낮은 부분에 있어서는 공진주파수의 대역폭이 넓게 나타난다.

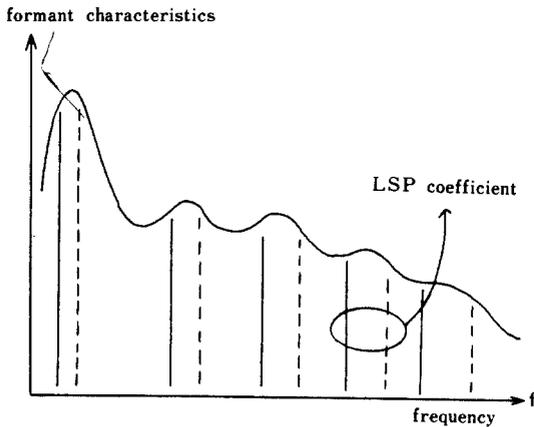


그림 2 LSP와 공진 특성과의 관계
The Relations of LSP and Formant characteristics.

2. LSP 분석

PARCOR 방식은 파형의 샘플치 사이의 선형 예측성에 따라 관정하고 그 예측가능한 부분을 취하여 제거한 잔류 부분간의 (정규화)상관으로 정의되어진 선형 예측 계수 사이에서 정보가 집약된 파라미터를 이용하는 것으로 분석 필터의 구성은 그림 3과 같이 격자형 디지털 필터로 실현된다⁽⁹⁾.

그림 3과 같이 음성의 합성 필터가 P단으로 구성

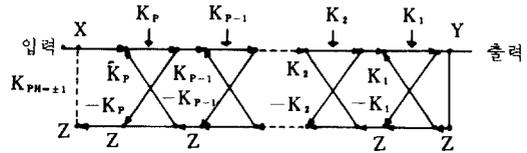


그림 3 PARCOR 합성 필터.

되었을 때 P차 전극형(allpole model) 디지털 필터에 대한 전달함수는 다음과 같이 표현된다.

$$H(Z) = \frac{1}{A_P(Z)}, \quad Z = \exp(-j\lambda),$$

$$\lambda = 2\pi \Delta T f \tag{1}$$

식(1)에서 ΔT 는 표본화 주기이고 $A_P(Z)$ 는 다음과 같이 Z에 관한 P차의 다항식으로 표현된다.

$$A_P(Z) = 1 + \alpha Z + a_2 Z^2 + \dots + a_P Z^P \tag{2}$$

다항식 $A_P(Z)$ 의 계수 순서를 반대로 하고 Z를 역함수인 함수를 다음과 같이 정의한다.

$$B_P(Z) = Z(\alpha^P + \alpha_{P-1} Z + \dots + \alpha_1 Z^{P-1} + Z^P) \tag{3}$$

이때 다항식 $A_P(Z)$, $B_P(Z)$ 에 관해서 다음의 점화식이 성립한다.

$$\left. \begin{aligned} A_0(Z) &= 1, \quad B_0(Z) = Z \\ A_n(Z) &= A_{n-1}(Z) - K_n A_{n-1}(Z) \\ B_n(Z) &= Z(B_{n-1}(Z) - K_n A_{n-1}(Z)) \end{aligned} \right\} \tag{4}$$

식(4)에서 파라미터 $\{K_n\}_{n=0}^P$ 은 PARCOR 계수, 또는 반사계수라 한다.

K_1, K_2, \dots, K_P 가 주어지면 식(1)의 성도 공진특성 $H(z)$ 가 정해진다. 점화식(4)에서 $H(z)$ 는 PARCOR 합성 필터에서 실현됨을 알 수 있다. 그림 3에서 보

는 바와 같이 X에서 Y간의 전달함수는 $1/A_p(Z)$ Y에서 Z간의 전달함수 $B_p(Z)$ 로 해석되고 상대경계 조건이 무반사로 되기 위해서 상대 경계 조건을 완전 개방($K_{p+1} = 1$) 또는 완전 폐쇄($K_{p+1} = -1$)로 하면 계는 무손실계로 되고 공진 Q값이 무한대의 선스펙트럼과 같게 된다.

그림3에서 Z와 X사이에 K_{p+1} 을 삽입할 때 X-Y간의 전달 특성 $H(z)$ 는 다음과 같다.

$$H(z) = 1 / (A_p(Z) K_{p+1} B_p(Z)) \quad (5)$$

따라서 $K_{p+1} = +1$ 의 조건에 대한 공진 주파수는 각각 다음의 P+1차의 다항식에 근으로 구해진다.

$$\left. \begin{aligned} P(Z) - A_p(Z) - B_p(Z) \\ Q(Z) - A_p(Z) + B_p(Z) \end{aligned} \right\} \quad (6)$$

식(6)에서 $P(1)$ 과 $Q(-1)$ 의 경우의 $P(Z)$ 과 $Q(Z)$ 의 값이 0이 됨을 식(2), (3)에 의해 알 수 있다. $\hat{H}(Z)$ 는 무손실계로 있기 때문에 $P(Z)$, $Q(Z)$ 의 근은 Z평면의 단위 원주 상에 있다. 따라서 $P(Z)$, $Q(Z)$ 는 다음과 같이 인수분해된다.

$$\left. \begin{aligned} P(Z) &= (1 - Z) (1 - 2 \cos \omega_1 Z + Z^2) \\ &\quad \dots (1 - 2 \cos \omega_{p/2} Z + Z^2) \\ Q(Z) &= (1 + Z) (1 - 2 \cos \theta_1 Z + Z^2) \\ &\quad \dots (1 - 2 \cos \theta_{p/2} Z + Z^2) \end{aligned} \right\} \quad (7)$$

식(7)에서 $\{\omega_i\}$ 는 상대 개방, $\{\theta_i\}$ 는 상대 폐쇄 조건에 대하여 공진 주파수로 되고 양자를 포함한 것을 LSP파라메타라 한다. 또한 LSP 파라메타는 다음의 순서 관계가 있다.

$$0 = \omega_0 < \theta_1 < \omega_1 < \dots < \theta_{p/2} < \omega_{p/2} < \theta_{p/2+1} = \pi \quad (1/2 \Delta THz)$$

그림 4는 한국어 단음중 "오"음에 대한 LSP 파라메타 관계를 묘사한 것으로 스펙트럼의 선폭에 Q가 높으면 선스펙트럼대가 좁게 나타나고 Q가 낮으면 선스펙트럼대가 넓게 나타난다. 따라서 선스펙트럼대의 값은 원래의 스펙트럼형의 정보가 보존된다고 할 수 있다.

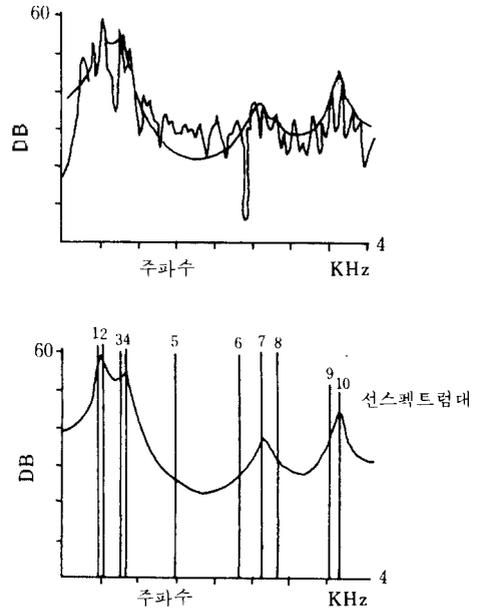


그림 4 스펙트럼의 선스펙트럼대 Modeling of the Line Spectrum Pair.

LSP계수를 구하는 전체 Algorithm의 순서도는 그림 5와 같다. 주어진 LPC계수들에 의해 전달함수 $H(Z)$ 의 분모 다항식 $A_p(Z)$ 와 $B_p(Z)$ 을 구한다. 그리고 주어진 반사계수 조건에 따라 $P(Z)$ 와 $Q(Z)$ 을 구한 후 $P(Z) = 0$ 와 $Q(Z) = 0$ 인 근을 구한다. 그리고 구해진 각 근들을 Z와 주파수의 관계를 이용하여 ω_i 와 θ_i 를 변환시킨 후 이 값들을 작은 순서에 따라 배열한다. 추출된 LSP 파라메타 θ_i, ω_i 로부터 식(1)의 전달함수 $H(Z)$ 으로 구성되는 합성필터는 그림 6과 같이 실현할 수 있다. 필요한 연산량은 1 sample 당 승산은 P회, 가감산은

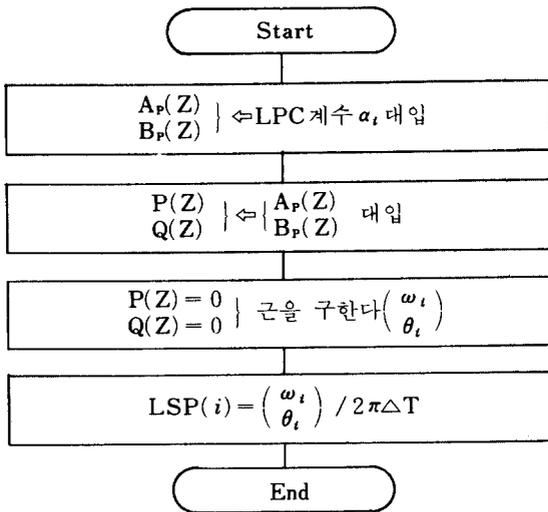


그림 5 LSP 계수를 구하는 순서도
The flowchart computing LSP coefficient.

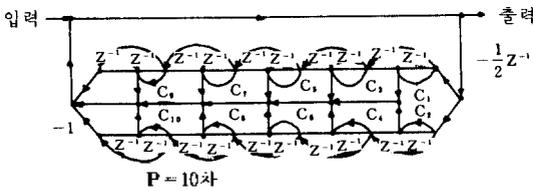


그림 6 합성 필터(LSP)
Synthesis filter(LSP).

3P+1 회이며 그림 3의 PARCOR 방식의 합성 필터와 비교하면 PARCOR 방식은 1 sample 당 2P-1 회의 승산과 가감산이 필요하므로⁽²⁾ LSP 방식이 승산에 있어서 1/2 정도가 감소하였고 가감산도 1/2 정도가 증가함을 알 수 있다. 그러나 일반적으로 연산시 승산은 가감산보다 처리 과정이 많기 때문에 LSP 방식의 연산량이 PARCOR 방식에 비해 적음을 알 수 있다.

III. 실험 및 고찰

본 연구에서는 그림 7과 같은 시스템을 구성하여 음성신호를 분석하였다.

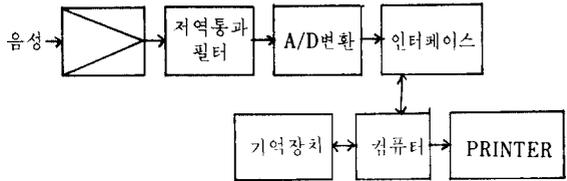


그림 7 실험장치의 구성도
Block diagram of experimental system.

그림 7에서 보는 바와 같이 4 KHz의 저역통과 필터로 음성 신호를 필터링한 후 8KHz의 Sampling 수파수로 12bit A/D 변환을 하였다. A/D 변환된 음성 신호는 보조 기억 장치에 저장되고 분석 구간을 20ms (160 sample) 분석 프레임 주기를 10ms로 하여 LPC (선형예측) 계수와 PARCOR 계수를 板倉法에 의해 구한다⁽²⁾. LSP 계수는 그림 5에 나타난 Algorithm을 사용하여 LPC 계수를 입력 데이터로 추출하였다. 그림 8에 실험에 사용한 전체 장치의 사

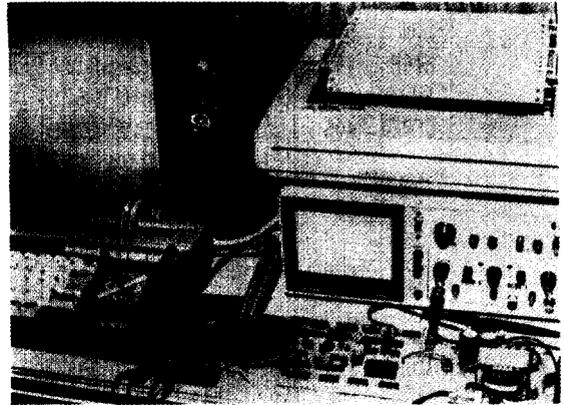


그림 8 전체 실험장치
Whole experimental system.

진을 보였다. 표 1은 본 연구의 분석에 사용한 한국어 기본 모음 “ㅏ”, “ㅑ”, “ㅓ”, “ㅕ”, “ㅡ”, “ㅣ” “ㅞ”, “ㅟ”의 LSP 계수를 주파수 함수에 대하여 낮은(Low) 부분과 높은(High) 부분, 대역폭, 중심 주파수 등으로 분류하여 표시한 실험 결과치 데이터이

표 1 한국어의 기본모음에 대한 LSP 계수.
LSP coefficient of Korean vowels.

“아”					
	[Hz]				
LSP	1	2	3	4	5
LOW	693	901	1675	2901	3450
HIGH	751	979	2498	3276	3647
BW	54	78	823	375	197
CFREQ	720	940	2086	3088	3548

“오”					
	[Hz]				
LSP	1	2	3	4	5
LOW	384	735	1277	2871	3501
HIGH	415	800	2560	3473	3832
BW	31	65	1283	602	331
CFREQ	399	767	1918	3172	3666

“으”					
	[Hz]				
LSP	1	2	3	4	5
LOW	370	965	1756	2811	3446
HIGH	443	1291	2739	3427	4117
BW	73	326	983	616	671
CFREQ	406	1128	2247	3119	3781

“애”					
	[Hz]				
LSP	1	2	3	4	5
LOW	539	804	2054	2411	3401
HIGH	573	1993	2386	3297	3823
BW	34	1189	332	886	422
CFREQ	556	1398	2220	2853	3612

“어”					
	[Hz]				
LSP	1	2	3	4	5
LOW	603	1011	1299	2923	3568
HIGH	662	1108	2122	3116	3674
BW	59	97	1023	183	106
CFREQ	659	1059	1810	3024	3621

“우”					
	[Hz]				
LSP	1	2	3	4	5
LOW	242	439	1794	2831	3602
HIGH	284	682	2678	3450	4039
BW	42	243	884	619	437
CFREQ	263	560	2236	3140	3820

“이”					
	[Hz]				
LSP	1	2	3	4	5
LOW	112	806	1572	2902	3287
HIGH	177	1232	1881	3072	3709
BW	65	726	309	170	422
CFREQ	144	1169	1726	2987	3498

“에”					
	[Hz]				
LSP	1	2	3	4	5
LOW	431	802	1952	2610	3248
HIGH	476	1674	2007	2771	3503
BW	45	872	55	161	255
CFREQ	453	1238	1979	2690	3375

다. 표 1에서 보는 바와 같이 각 기본 모음의 형태는 본론 부분에서 이미 언급한 바와 같이 내역폭이 넓게 나타난 부분은 공진 특성 즉 Q가 낮음을 볼 수 있으며 좁게 나타난 부분은 공진 특성이 높은 것을 의미한다.

IV. 결 론

LSP 방식은 이제까지 연구되었던 PARCOR 방식과 비교하여 불매 파라미터 추출시 소요되는 연산량이 PARCOR 방식의 약 $\frac{1}{2}$ 정도가 소요됨을 알 수

있었다. 따라서 LSP 방식은 음성 합성시 합성 권터의 구성에 있어서 상당히 경제적인 이점이 있다고 사료된다. 또한 LSP 계수가 주파수 특성 측공진 특성에 의한 파라메타이므로 PARCOR 방식에 있어 가장 커다란 결점으로 지적되었던 주파수 스펙트럼 왜곡 문제를 상당히 보완할 수 있었다. LSP 방식에 의한 한국어의 기본모음 분석 결과 각 모음 부분의 선스펙트럼대 특성이 고유한 성질을 나타낼 수 있었으며 이러한 각 모음 특징을 이용하여 한국어 음성 인식과 합성을 하기 위한 특징 파라메터로 사용될 수 있으리라 사료된다.

참 고 문 헌

1. H. Dudley: "Remaking speech", Jour. Acoust. Soc. Amer. vol. 11pp. 167~177, Oct. 1939.
2. 板倉, 齊藤: 偏自己相関係数による音声分析合成系, "日本音響学会論文集 2-2-6."
3. 板倉, 管村: "LSP音声合成器の原理と構成", 音成研究会. S79~46 (1979~11)
4. R.W. Schafer and L.R. Rabiner, "System for Automatic Formant Analysis of Voiced Speech," J. Acoust. Soc. Am., Vol.47, No.2, pp.634-648, February 1970.
5. J. Olive, "Automatic Formant Tracking in a Newton-Raphson Technique," J. Acoust. Soc. Am., Vol.50, pp.661-670, August 1971.
6. C.R. Patisaul and J.C. Hammett, "Time Frequency Resolution Experiment in Speech Analysis and Synthesis," J. Acoust. Soc. Am., Vol.58, No.6, pp.1296-1307, 1975.
7. L.R. Rabiner & R.W. Schafer: "Digital Processing of Speech Signals," Prentice-Hall, 1978.
8. L.R. Rabiner, R.W. Schafer, and J.L. Flanagan, "Computer Synthesis of Speech by Concatenation of Formant-coded Words," Bell System Tech. J., Vol.50, No.5, pp.1541-1558, May-June 1971.
9. 이용준, "PARCOR 방식을 이용한 한국어의 분석과 합성에 관한 연구", 석사학위논문, 건국대학교 1985. 12.