삼각필터를 이용한 Spectral 포락변경에 관한 연구

최 성 은, 김 동 현, 홍 광 석
성균관대학교 정보통신공학부
전화 : 031-290-7196

A Study on Spectral Envelope Modification using Triangular Filter

Seong-Eun Choi, Dong-Hyun Kim, Kwang-Seok Hong
School. of Information and Communication Engineering, Sungkyunkwan University
E-mail : kkcse@hanmail.net

Abstract

In this paper, we present a new filter to adjust formant information. Spectral envelope in speech analysis shows information about characteristics of speech and formant information determines speech timbre. So, if formant position is adjusted, we can verify adjusted speech timbre. A presented filter is to adjust this formant. This filter is composed of triangular filters. Using this filter we could locate the formant frequency at target position.

I. 서론

기존의 음색 변환을 위한 방법으로는 시간영역과 주파수 영역의 피치 변경법과 주파수 영역의 포만트 변경을 이용한 음색 변환 기법을 사용하였다. 본 논문에서는 주파수 영역의 포만트 변경 방법중 Multirate 기법과 삼각필터 를 이용하여 Formant의 위치를 변경하여 합성음의 음색을 조절하는 방법을 제안한다. 음성의 인식과 합성에서 지금까지 가장 유효하게 사용되고 있는 기술의 하나로 spectral 포락정보를 들수 있다. 포먼트 정보는 음성의 개별 화자의 특징을 잘 나타내어주는 파라미터로 지금까지 여러가지 방법으로 연구되고 있다. 특히 Abe 는 각 분석 구간의 포먼트 특성에 대해 VQ를 이용하여 양자화한 후 대상 화자와 기준화자 사이의 포만트 특성에 대한 DTW를 통해 대응관계를 구한 포만트 특성을 변환하는 방법을 제안 하였다. 그러나 VQ 와 DTW 에 의한 포만트의 변환은 합성음의 음질 저하를 야기하였다. 그후에 포만트 포락선의 변환후에 변환된 포락선에 근접하므로 포만트 값을 근사화시키는 포만트 변환 기법을 다시 제안하기도 했다[6]. 이런 기법은 포만트의 폭과 크기를 고려하지만, 반복적인 계산과 합성음의 음질을 보장하기 어려웠다. 본 논문에 제시한 포먼트 변경을 위한 필터로서 삼각 필터를 사용하였으며, 주파수 band 별로 병렬로 각각 필요한 만큼 필터를 씌워 포먼트의 위치를 변경하였다. 이 방법의 또한 가지 특징은 원하는 주파수대역의 포락을 원하는 위치에 거의 근접하게 이동시킬수 있다는 것이다. 앞으로 이러한 필터의 구성 방식은 많은 응용 분야에 적용될수 있을 것으로 보인다.

II. 음성합성

합성음의 음색 변환에 관한 연구가 다년간 수행되어 왔지만, 좋은 합성음을 내기에는 여전히 부족하다. 그 이유는 합성음의 생성 방식에 있다. 음성 합성의 방식은 크게 세 부류로 나눌수 있다. 첫 번째는 직접합성(Direct Synthesis), 두 번째는 음성 생성 모델을 사용한 합성(Synthesis using a Speech Production Model)

그리고 세번째는 성도를 시뮬레이션한 합성(Synthesis by Vocal Tract Simulation) 으로 나눌수 있다. 각각의 합성 방식은 다음과 같다. 대표적인 합성방법으로 파형을 직접 연결한 방식(Wave form concatenation)을 들수 있다. 몇 개의 기본 주파수로 구성된 음소(Phonemes)열의 직접 합성으로 단어나 문장을 합성하는 것을 이 방식의 예로 들수 있다. 이러한 직접합성 방식은 계산량이 적고 메모리 공간도 적게 차지한다. 그러나 각 음소의 연결부위를 고려하지 않았기 때문에 명료성이 떨어진다. 즉, 상대적으로 음질이 좋지 않다는 단점이 있다. 이 기술의 적용 분야는 넓지 않고, 아이들을 위한 장난감 등에 적합한 기술이다. 또 다른 기술로 채널 보코더(channel vocoder) 이 기술은 여러 개의 채널 주파수에 대한 신호를 발생시켜 합성음을 만드는 기술이다. 여기서 채널 주파수는 음성의 스펙트럼에서 포먼트 에너지의 분포에 따라 만들어 진다. 이 기술의 이점은 스펙트럼 분석 결과에 대해 직접 합성음을 생성 한다는 것이다. 그러나 채널수가 제한 되기 때문에 그만큼 음질은 떨어진다 또한 많은 양의 스펙트럼 정보를 가지고 있어야 된다는 단점이 있다. 따라서 현재는 이 기술이 더 이상 사용되지 않고 있다. 음성이 생성되는 기관을 시뮬레이션 하는 방법은 주로 소스 필터 이론(Source-Filter Theory) 에 따른다. 이것에는 두가지 대표적인 기술이 있다. 첫 번째는 포먼트 합성기(The format synthesiser) (Klatt,1980) 이다. 이 방법은 음성의 포먼트 특성을 인공적으로 재결합 시켜서 합성음을 만드는 것이다. 이것은 발성음의 스펙트럼에 따라 voice source generator 또는 noise generator 에 의해 각각 모음과 무음부의 음성을 발생시킨다. 이 방법의 잇점은 여기서 사용된 파라미터(parameter)가 구강(The oral tract)에서 생성 또는 전달되는 소리와 직접적으로 많은 연관 관계를 가진다는 것이다. 그러나, 이 방법의 경우에 포먼트에 대한 파라미터를 자동적으로 생성해 내는 것이 만족스럽지 못하다는 단점이 있다. 두 번째는 LPC (Aral and Hanauer,1971)를 사용한 다이폰단위 연결(diphone concatenation) (O'Shaughness,1988) 기술 이다. 현재는 상업적으로 다양한 연결 합성 기술을 사용하고 있다. 이 기술은 음성의 적절한 세그멘테이션을 통한 수작업된 음성 데이터가 필요하다. 일단 세그먼트된 음성 데이터가 준비되면 합성음을 위해서 각 단위별 데이터에 계산적인 파워를 조절하고 연결하여 합성하면 된다. 음소(Phoneme)단위의 연결합성의 경우에는 음소와 음소 사이의 조음현상을 고려하기 어려운 문제(Coarticulation problem)가 있다. 만약, 좀더 큰 단위로 연결 합성을 할 경우에는 이러한 조음현상을 고려하는 부분이 유리하기 때문에 합성음의 음질이 상대적으로 좋아진다. 다이폰 단위의 경우 이러한 조음현상의 문제를 잘 고려한 최소 단위로 볼수 있다. LPC 합성의 경우는 분석과 재합성을 통해 제한된 저장공간에서 상대적으로 양질의 합성음을 만들 수 있는 합성방법이다. LPC 모델의 신호는 예측샘플의 선형결합에 의한 샘플신호로 볼수 있다. 이것의 알고리즘은 원 신호와 예측신호의 MSE(Mean Square Error)를 최소화 하는 계수(Coefficient)를 구하는 것이다. 양질의 합성음을 만들기 위해서는 5 ~ 20 ms 분석 신호에 대해 적어도 10 ~ 16 개의 예측 계수(Coefficients)가 필요하다. 이 시스템의 장점으로는 자동으로 원신호에 대해 분석할 수 있고, 합성 또한 원 신호에 가까운 합성음을 낼 수 있다는 것이다. 그러나 다이폰 단위의 LPC 합성의 경우 합성시의 두 모음간의 경계에서 불연속적인 합성음이 생성 될 수 있다는 것이다. 포먼트의 불연속의 경우에 다이폰의 경계 부분에서 이중적인 소리나 불연속적인 소리가 발생 될 수 있다. 또한 비음이나 마찰음에 대해서는 좋은 합성 결과를 얻기 어렵다. 왜냐하면, 성도 모델이 폴(Poles) 만을 포함하는 반면 비음이나 마찰음의 경우 폴(Poles)뿐 아니라 제로(Zero)를 가지기 때문이다. 앞선 기술들은 관찰을 통하여 계산적으로 최적의 소리를 만드는 방법을 사용하였다. 또한, 조음 연결에 자연성의 문제가 있었다. 1990년대 이후로 음성 생성 기관의 물리적인 움직임을 시뮬레이션함 (Scully, 1990 ; Maeda, 1990)으로서 조음의 문제를 직접적으로 해결하기 위한 많은 시도가 이루어져 왔다. 이러한 조음 모델(Articulatory Model)은 발성기관(입술, 턱, 혀, 연구개)의 위치를 함수화 하여 성도의 모형을 재구성하는 것이다. 발성되는 신호는 성도로부터 내보내어 지는 공기를 수학적으로 시뮬레이션 하여 계산된다. 이러한 합성기를 제어하기 위한 파라미터로는 성문하압(Sub-glottal pressure), 성대장력(Vocal cord tention), 그리고 조음기관(Articulatory organ)의 상대적인 위치 등이 있다[3].

III. 음색변경

3.1 음색의 정의

같은 높이, 같은 크기의 소리라도 발음체의 종류에 따라 다른 소리로 듣게 된다. 또 같은 종류의 발음체라도 각각의 발음체에서 나오는 소리에는 그 발음체 고유의 특징이 있다. 이런 소리의 개성을 음색이라고 한다. 이것은 진동에 의해 어떤 부분음이 어느 정도의 강도로 발생하는가에 따라 결정된다. 물리적으로는 그 소리에 포함되어 있는 음의 구성의 차이에 의해 설명된다. 즉 소리의 높이는 주로 그 소리의 기본 주파수

에 의해서 결정되지만, 음의 구성이 다르면 같은 높이라도 음색의 차이로 이를 분간할 수 있게 되는 것이다. 따라서 음색은 음의 높이에 관계없이 주관적으로 음을 다르게 구분 할 수 있는 기준이 된다. 예를 들어 여러 사람이 같은 높이의 같은 발음을 할 경우에 우리가 다른 목소리로 들리는 것은 각 사람의 음높이에 대한 음의 진동 특성이 다르기 때문이다[2].

3.2 포먼트의 조절

음색은 음의 높이가 아니라 음의 진동특성의 차이로 구분한다고 정의 하였다. 음성합성에서 음성의 포먼트의 특성은 이러한 특성을 잘 나타내 준다. 특히, 음높이가 다를 때에도 음높이에 따른 포먼트의 이동이 크지 않다. 또한, 발성하는 화자의 특성에 따라 같은 음이라도 조금씩 다른 포먼트의 특성을 가지는 것은 이러한 특성을 확인할수 있는 예가 되겠다.

포먼트의 변환을 이용한 음색변환에 대한 연구중 Abe는 각 분석구간의 포먼트 특성에 대해 VQ를 이용하여 양자화한 후 대상화자와 기준화자 사이의 포먼트 특성에 대한 DTW를 통해 대응관계를 구한 포먼트 특성을 변환하는 방법을 제안 하였다. 그러나 이 방법의 단점은 계산량이 많고 합성시 합성음의 음질을 저하시키므로, 포먼트의 이동후에 기존의 포먼트의 위치에 유사한 포먼트값에 근사화 시키는 방법을 사용하여 음색을 조절하였다.

Ⅳ. 필터의 구현

본 논문의 음색 변환 시스템에서 필터(Filter)는 LPC 분석후의 포먼트의 위치를 변환하여 음색을 변경 시키기 위한 새롭게 제안된 방법이다. 필터(Filter)의 기본 아이디어는 Mel-Filter의 삼각 필터(Filter)의 구성에서 얻을 수 있었다. Mel-Filter에서는 주파수 축에서 삼각창이 Mel-Scale에 따라 그 크기와 폭이 다른 각각의 삼각 창으로 이루어져 있다. 그러나, 각 주파수대별로 비교적 자유롭게 원하는 위치로 포먼트의 위치를 옮기기 위해서 모든 주파수대에 걸쳐 일정한 크기와 폭의 필터를 구성하였다. 다음은 이 필터(Filter)에 필터뱅크(Filter Bank)를 통한 다중분할(Multirate) 기법과 주파수 축에서의 리샘플링(Resampling) 기법을 사용하여 필터를 조정 하였고, 이를 통하여 포먼트의 위치또한 조정할 수 있었다. Multirate 기법은 인터폴레이션(Interpolation)과 데시메이션(Decimation)을 통하여 샘플링 비율을 증가 또는 감소 시키는 방법이다. 인터폴레이션(Inerpolation)은 입력신호사이에 L-1 개의 zero를 삽입 하고 이때 주파수축에서 발생하는 이미지(image)성분을 필터링(filtering) 하여 제거한다. 데시메이션(Decimation)은 M배만큼 입력 신호를 감소 시키면, 주파수 축의 대역폭이 M배만큼 증가되어 엘리어싱(Aliasing) 이 발생하고 이것을 보상하기 위해서 일반적으로 데시메이션(Decimination) 하기전에 필터링(filtering)을 한다. Filter는 주파수 대역별로 분리하여 26개의 반 삼각 필터로 구성되어 있다. 각각의 삼각필터의 크기(Magnitude)는 1로 고정되어 있으며, 원하는 구간의 주파수 대역의 필터를 변경하는 것과 필요한 변경 주파수 대역은 프로그램적으로 선택할수 있도록 되어 있다. 다음 그림1 은 Filter Bank와 Multirate 기법을 사용한 주파수 축 스펙트럼의 분석/합성을 보여주고 있다[5].

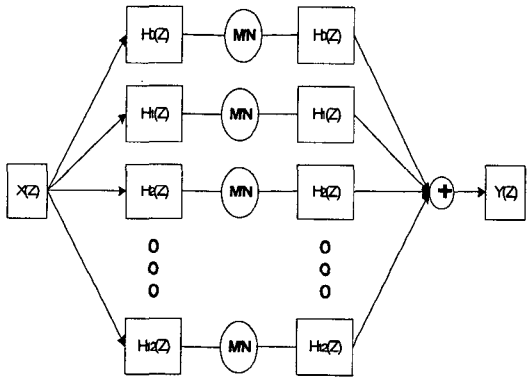

그림1 필터뱅크(Filter Bank)를 통한 다중분할(Multirate) 필터링(Filtering)

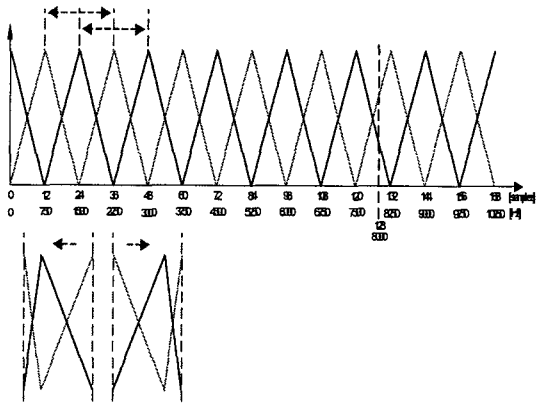

그림2 삼각필터(Triangular Filter)의 구성

표1 삼각필터의 구성

	Filter Number						
UpFilter	1	3	5	7	9	11	13
Sample	6 ~ 17	30 ~ 41	54 ~ 65	78 ~ 89	102 ~ 113	126 ~ 137	150 ~ 161
Frequency	375 ~ 1062.5	1875 ~ 2562.5	3375 ~ 4062.5	4875 ~ 5562.5	6375 ~ 7062.5	7875 ~ 8562.5	9375 ~ 10062.5
DownFilter	2	4	6	8	10	12	
Sample	18 ~ 29	42 ~ 53	66 ~ 77	90 ~ 101	114 ~ 125	138 ~ 149	
Frequency	1125 ~ 1812.5	2625 ~ 3312.5	4125 ~ 4812.5	5625 ~ 6312.5	7125 ~ 7812.5	8625 ~ 9312.5	

V. 실험 및 결과

표2 필터링에 따른 주파수 shifting 의 범위

이동폭	sample 개수	sample간 삽입개수	이동폭	sample 개수	sample간 삽입개수
-1	11	10	+1	13	12
-2	10	9	+2	14	13
-3	9	8	+3	15	14
-4	8	7	+4	16	15
-5	7	6	+5	17	16
-6	6	5	+6	18	17

표3 주파수 축의 변화량.

sample개수	1	2	3	4	5	6
변경폭 (frequency)	0 ~ 62.5	63 ~ 125	126 ~ 187.5	188 ~ 250	251 ~ 312.5	313 ~ 375

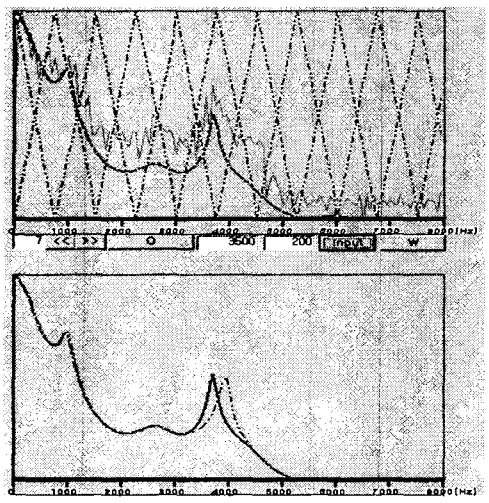

그림3 발성음 " 아 " 의 포락 변경

그림3 은 제안된 필터를 통한 포락변경의 예를 보여주고 있다. 16khz 16bit 발성음에 대하여 LPC 분석하고 spectral 포락의 분포를 확인후 원하는 주파수 대역의 포먼트의 위치(여기서는 3500hz 위치의 포먼트)를 원하는 크기 만큼(여기서는 +200hz) 옮긴 결과를 보여주고 있다.

VI. 결론

본 논문에서는 새로운 필터의 구성을 소개 하였다. 소개된 필터는 기본적으로 삼각 필터로 구성 되었고, 원하는 위치의 정보를 필터링한 후 재 배열하는 방법으로 정보를 변경할 수 있었다. 응용의 예로 음성의 포락 정보를 변경하여 보았다. 향후 이 연구의 결과는 합성 및 인식의 음색 변환 관련 분야에 적용 될수 있을 것으로 생각된다.

< acknowledgement >

본 연구는 한국과학재단 목적기초연구 (R05-2002-000-01007-0) 지원으로 수행되었음.

참고문헌

[1] L. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993.

[2] Thomas D. Rossing , F. Richard Moore and Paul A. Wheeler, "THE SCIENCE OF SOUND", Addision Wesley, 2002.

[3] Styger, T., & Keller, E. "Formant synthesis" , *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges* (pp. 109-128). Chichester :John Wiley.

[4] A. V. Oppenheim, R. W. Schafer ; *Discrete - Time Signal Processing*, Prentice-hall, 1989.

[5] Vetterli, M. "A theory of multirate filter banks" Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on , Volume: 35 Issue: 3 , Page(s): 356 -372, 1987.

[6] 최철민, 전범기, 성굉모, "유성음의 잔류신호 변환을 이용한 음색 변환", 음향학회 학술발표대회 논문집 제16권 제2(s)호, pp.127-130, 1997.