제한된 시간적 분해법에 기반한 선스펙트럼 주파수의 효과적인 양자화

Efficient Quantization Method for Line Spectral Frequencies Based on Restricted Temporal Decomposition

김 승 주*, 오 영 환*
(Sung Joo Kim*, Yung Hwan Oh*)

※본 연구는 한국과학기술원 인공지능연구센터의 지원에 의해서 이루어졌습니다.

요 약

본 논문에서는 선스펙트럼 주파수(LSF) 파라미터를 위한 제한된 시간적 분해법을 제안한다. LSF 파라미터는 인접 차수에 대해 의존적이고, 차수간 순차성이 있으나, 기존의 시간적 분해법은 이러한 성질을 보존하지 못한다. 즉, 추정된 사건 벡터가 더 이상 LSF 파라미터로서 해석되지 못하는 문제가 있다. 이를 해결하기 위하여, 본 논문에서는 사건 함수간에 새로운 제약을 두어, 추정된 사건 벡터가 LSF 파라미터의 성질을 유지하도록 한다. 결과적으로 제안된 방법을 이용하여 구해진 사건 벡터는 LSF 파라미터와 동일한 방법을 적용하여 효과적으로 양자화될 수 있고, 실험 결과 평균 752bps의 전송률로 투명한 양자화를 수행할 수 있었다.

ABSTRACT

In this paper, a restricted temporal decomposition method for line spectral frequency (LSF) parameters is presented. LSF parameters are dependent to adjacent orders and have the ordering property, but the original temporal decomposition method cannot preserve such properties. As a result, the estimated event vectors are no longer LSF parameters. To solve this problem, we enforce another restriction on event functions so that every event vectors for LSF parameters retain their properties. Consequently, the event vectors estimated by this method can be quantized efficiently and experimental results show that LSF parameters can be quantized transparently at the average rate of 752bps.

I. 서 론

음성을 선형 예측 모델로 근사하고, 이를 전송하거나 저장하기 위해서 프레임별로 추정된 모델 파라미터의 적절한 양자화가 필요하다. 선형 예측 모델에서 음성의 스펙트럼 포락 정보를 나타내는 선형 예측 계수(LPC)는 그 값의 범위가 일정치 않고, 특정 차수의 양자화 오류가 전체 대역폭에 걸쳐 반영되는 등 양자화에 어려움이 있다. 또한 양자화된 선형 예측 계수로 결정된 전극 필터의 안정성(stability)을 검사할 수 있는 직접적인 방법이 없어 음성부호기에 사용될 파라미터로는 적합하지 못하다. LPC 파라미터 양자화의 이러한 문제점을 회피하기 위한 방법으로 전극 필터를 결정할 수 있는 다른 형태의 파라미터로 변형하여 양자화하는 방법들이 제안되었고, PARCOR (partial correlation) 계수, LAR(log-area-ratio) 계수 등이 양자화에 이용되고 있다. 그러나, 이러한 방법 중에 LSF (Line Spectral Frequency), 즉 선스펙트럼 주파수가 가장 널리 쓰이고 있으며, 본 논문에서도 LSF 파라미터의 양자화에 대해 기술하고자 한다.

LPC 파라미터를 LSF 파라미터로 변환하게 되면 양자화 시에 다음과 같은 장점을 얻게 된다. 첫째, LSF 파라미터가 0과 π사이에 순서대로 존재한다는 사실을 이용하여 필터의 안정성을 쉽게 검사할 수 있다. 둘째, LSF 파라미터는 차수간 상관 관계가 크고, 시간에 따라서도 변화가 적어 인접 차수나 인접 프레임의 파라미터에 대한 예측 및 보간이 용이하다. 셋째, LSF 파라미터는 그 파라미터가 나타내는 주파수를 중심으로 한정된 범위의 스펙트럼 포락에만 영향을 미치므로, 특정 차수의 양자화 오류는 특정 주파수 대역에 대해서 반영되는 특징이 있다. 따라서 주파수 대역별 청각 특성을 이용할 수 있다.

이러한 LSF 파라미터의 장점 때문에 많은 연구자들이 LSF 파라미터의 양자화 방법을 연구해 왔다. 효과적인 LSF 파라미터 양자화를 위해 사용된 성질을 살펴보면,

* 한국과학기술원 전산학과
접수번호 : 1998년 3월 12일

먼저 LSF 파라미터의 차수간 순차성을 이용하는 방법이 있다[4]. 즉, LSF 파라미터의 각 차수간 증분치, dLSF (differential LSF)를 양자화하는 방법이다. 다음으로는 LSF 파라미터의 시간적인 변화가 느리다는데 착안하여 이전 분석 구간의 파라미터로 현재 구간의 값을 예측하고, 잔차값을 양자화하는 방법이 연구되었다[9]. 마지막으로 음성 스펙트럼의 분포 특성을 반영하여 LSF 파라미터의 벡터 양자화하는 방법이 있다[5, 10]. 또한 두개 이상의 성질들을 조합하여 적용한 방법도 다수 연구되었다[6, 7, 8].

이상의 방법들은 모두 일정 분석 구간별로 구해진 LSF 파라미터를 양자화하므로, 고정 주파수 표본 방식으로 볼 수 있다. 즉, 양자화된 정보량이 시간에 따라 변하지 않는 특징이 있다. 그러나 음성 신호의 스펙트럼 포락 정보는 시간에 따라 정보량이 변하는 가변성을 가지고 있다. 모음과 같은 안정된 구간에서는 신호의 스펙트럼 포락이 일정하게 유지되는 반면, 자음이나 천이 구간에서는 빠르게 변화하는 차이가 있다. 한편 표본 이론에 따라 이렇게 가변적인 대역폭을 갖는 신호에 대해서는 최대 대역폭의 두 배에 해당하는 주파수로 표본을 수행해야 한다. 따라서 고정 주파수 표본 방식의 양자화는 입력 신호 대역폭의 가변성을 수용하지 못하는 비효율적인 측면을 내포하게 된다. 결국 음성의 스펙트럼 포락 정보를 효과적으로 양자화하기 위해서는 그 천이 정도에 따라 양자화 주기를 바꿀 수 있는 가변 주파수 표본 방식이 채택되어야 한다.

스펙트럼 포락 정보의 가변 주파수 표본 방식의 양자화는 극저전송률 음성부호기에 적용하기 위해 개발된 사례가 많다. 분절 양자화 방법이나 시간적 분해법이 대표적인 경우다. 그러나 분절 양자화의 경우, 기본적으로 벡터 양자화 수법을 사용하여, 분절 내에 벡터 궤적 자체를 양자화하므로 코드북을 설계 및 학습하고, 분석시 코드북 탐색 과정에 많은 연산량과 메모리가 요구되어 실용화에 문제가 있다. 한편 시간적 분해법은 주어진 벡터 궤적을 몇 개의 목적 벡터로 순차적으로 변해 가는 과정으로 보고, 목적 벡터와 변화 과정을 각각 사건 벡터와 사건 함수로 분리하여 양자화한다. 따라서 고차원의 벡터 양자화 과정이 불필요하고, 연산량이나 메모리 측면에서 분절 양자화에 비해 구현이 용이하다. 그러나, 지금까지 시간적 분해법은 LAR 파라미터나, 필터 뱅크(filter bank) 파라미터, 또는 켑스트럼(cepstrum) 파라미터와 같은 차수별 독립성이 보장된 특징 파라미터에 대해서만 적용이 가능하였고, LSF 파라미터와 같이 차수별 파라미터 값이 연관성이 있는 경우에는 적용한 예가 없다. 본 연구에서는 양자화 특성이 우수하고, 쉽게 안정성을 검사할 수 있는 LSF 파라미터에 대해 시간적 분해법을 적용하는 방법을 제안한다.

II. 기존의 시간적 분해법

음성 스펙트럼 포락 정보의 양자화를 위해서는 특징 파라미터의 시간적 변화를 일정 주기로 표본하는 방법이 일반적이다. 그러나 모음의 안정된 구간과 자음의 빠른 변화구간을 고려한다면 일정 주기로 표본하는 방법은 효과적이지 못함을 알 수 있다. 즉, 특징 변화의 정도에 따라 표본 주기를 조절하는 방법이 요구된다. 시간적 분해법은 특징 파라미터의 변화를 몇 개의 사건 벡터들의 사건 함수에 의한 선형 결합으로 표현한다. 따라서 비주기적인 음성 생성 사건(speech event)에 대응하도록 사건 벡터와 사건 함수를 설정한다면 효과적으로 음성을 표현할 수 있다[1]. 다음에 음성의 시간적 분해법을 개략적으로 설명한다.

2.1 음성의 시간적 분해

시점 n에서의 스펙트럼 포락 정보를 벡터 $\vec{\psi}(n)$라 하면, 이런 벡터의 시계열을 벡터 궤적(vector trajectory)이라고 한다. 벡터 궤적 Ψ가 J개의 사건으로 구성된다고 가정하면, 대응하는 사건 벡터 $\vec{\omega}_j$ 및 사건 함수 $\vec{\phi}_j$는 다음의 (1)을 만족한다. 단, 사건 벡터의 차수는 주어진 파라미터 벡터의 차수와 동일하다.

$$\Psi = \Omega\Phi \tag{1}$$

단,

$$\Psi = [\vec{\psi}(1), \ldots, \vec{\psi}(n), \ldots, \vec{\psi}(N)]$$
$$\Omega = [\vec{\omega}_1, \ldots, \vec{\omega}_j, \ldots, \vec{\omega}_J]$$
$$\Phi = [\vec{\phi}_1, \ldots, \vec{\phi}_j, \ldots, \vec{\phi}_J]^T$$
$$\vec{\phi}_j = [\phi_j(1), \ldots, \phi_j(n), \ldots, \phi_j(N)]^T$$

시간적 분해법은 (1)에서 주어진 벡터 궤적에 대해 사건 벡터 및 사건 함수를 결정하는 방법이다. 일반적으로 이러한 문제는 여러 개의 해를 가질 수 있으므로 구하려는 사건 함수에 대해 다음과 같은 제약을 둔다.

$$0 \le \phi_j(n) \le 1, \quad \text{for } 1 \le j \le J,\ 1 \le n \le N \tag{2}$$

$$\max_n \phi_j(n) = 1, \quad \text{for } 1 \le j \le J \tag{3}$$

$$\sum_n \phi_i(n)\,\phi_j(n) = 0, \quad \text{if } |i-j| > I \tag{4}$$

위의 (2)는 사건 벡터의 가중치 범위를 제한한다. (3)은 각각의 사건 벡터가 특정 시점에서는 최대 가중치로 기여함을 의미한다. (4)는 I만큼 떨어진 사건 함수간에는 서로 영향을 미치지 못하도록 직교성을 부여한다. 이밖에 기본적으로 각각의 사건은 음성 생성 과정의 제어 신호에 대응하므로 사건 함수는 시간축 상에서 제한된 짧은 단일 구간 내에 분포한다는 가정을 이용하여 적절한 사건 함수가 추정된다.

시간적 분해법의 사건 함수와 사건 벡터를 계산하는 방법은 크게 두 가지가 있다. 첫 번째는 벡터 궤적의

SVD(Singular Values Decomposition)를 취한 후, 이를 이용하여 사건 함수를 추정하는 방법이다[2]. 두 번째는 대략적인 사건 함수의 위치 및 모양을 추정하고, 변형된 Gauss-Seidel iteration 방법을 이용하여 반복적으로 사건 벡터 및 사건 함수를 계산하는 방법이다[3]. 첫 번째 방법은 최초에 제안된 시간적 분해법의 원리에 충실하게 사건 함수 및 사건 벡터를 추정해 낸다는 장점이 있으나, 많은 연산을 필요로 하며, 반복 분석 시 분석 구간의 재설정으로 인해 실시간으로 구현하기에 어려운 점이 있다. 두 번째 방법은 실제 음성부호기에 시간적 분해법을 적용하기 위해 분석 구간 설정에 제한을 두어 구현상의 문제점을 해결하였으나, 경험적인 초기 사건 함수의 설정 방법에 따라 성능이 좌우될 수 있다는 문제가 있다.

음성의 시간적 분해법은 입력 음성의 변화에 따라 특징 파라미터를 가변 주기로 표본화하는 가변 주파수 표본(variable rate sampling)으로 볼 수 있다. 또한 표본화된 파라미터간에 상호 영향을 인정하며, 그 영향의 정도를 정량화하고 있다. 시간적 분해법의 표본된 사건들에 대한 이러한 가정은 음성학에서 거론되는 음소의 성질에 비교되는 것이고, 이미 과거의 연구에서 시간적 분해법의 결과가 입력 음성의 음소열에 잘 대응됨이 보고된 바 있다[2]. 따라서 기존의 고정 주파수 표본 방식(fixed rate sampling)으로 선택된 파라미터들에 비하여 시간적 분해법으로 얻은 파라미터들은 각각 독립적인 음성학적 의미를 지니게 된다. 한편 시간적 분해법은 또 다른 가변 주파수 표본 방법의 하나인 분절 부호기에서와 같은 템플릿을 필요로 하지 않으므로, 사전에 템플릿을 위한 학습이나, 분석시에 입력 음성과 템플릿간의 정합 과정 등이 생략되어 음성부호기를 구현하는데 유리하다.

2.2 LSF 파라미터의 시간적 분해의 문제점

1983년 Atal에 의해 처음 제안된 시간적 분해법은 벡터 궤적을 몇 개의 사건 함수와 대응하는 사건 벡터의 쌍으로 근사하였다. 그리고 사건 함수와 사건 벡터를 구하는 방법에 있어서, 전적으로 사건 함수의 시간적인 응집성을 이용하였다[1, 2]. 즉, 구하고자 하는 사건 함수를 주어진 벡터 궤적의 선형 조합으로 상정한 후, 그 시간축상의 제한점을 이용하여 적합한 추정 값을 얻어낸다. 다음으로 사건 벡터는 사건 함수가 정해진 후에 원래의 벡터 궤적이 근사되도록 최적화 된다. 결과적으로 기존의 시간적 분해법은 사건 벡터에 대한 별도의 제약은 고려하지 않으며, 다만 사건 벡터는 대응되는 벡터 궤적의 목표 벡터로서 의미를 갖는다. 이러한 사건 벡터의 집합은 음성 생성 과정의 사건(event)의 집합과 대응하여 입력 벡터와는 별개로 벡터 공간상에 분포하게 된다.

사건 벡터에 대한 제약점을 고려하지 않기 때문에 기존의 시간적 분해법은 LAR 파라미터나, 켑스트럼 파라미터와 같이 각 차수의 파라미터 값이 독립적이며, 그 값의 범위가 제한을 받지 않는 특징 파라미터에 대해서만 적용되어 왔다. 만일 LSF 파라미터와 같이 차수간에 값이 순차적이고, 값의 범위가 제한된 경우에는 기본적으로 다음의 두 가지 문제점이 발생된다. 첫째, 분석된 사건 벡터가 입력 벡터 궤적의 특징을 유지하지 못하므로 입력 벡터 궤적의 장점을 이용할 수 없다. 둘째, 입력 벡터로 결정되는 조음 필터의 안정성을 검증하는 방법을 분석된 사건 벡터에는 적용하지 못하므로 부호화에 이용할 수 없게 된다.

기존의 시간적 분해법은 사건 벡터에 대한 문제점 외에 구현상 어려운 점도 있다. 사건 함수를 추정하기 위해서는 구간 밖의 벡터 궤적 값이 분석 결과에 영향을 주지 않는 적절한 분석 구간을 설정할 필요가 있다. Atal의 경우, 전체 신호에 대해 분석 후, 다시 일정 개수의 사건이 포함된 분석 구간을 잡아 재분석할 것을 권장했지만, 실시간 음성부호기에서 전체 신호를 분석할 수는 없다[1]. Dijk-Kappers는 이러한 문제를 해결하기 위해서 임의의 초기 분석 구간으로 사건 함수를 찾고, 구해진 사건 함수의 분포에 따라 분석 구간을 재조정하는 적응 분석 구간(modified analysis window)을 제안하였다[2]. 그러나 이 방법 역시 재분석에 따라 반복되는 연산이 많아 실제 구현에는 어려움이 많다. 마지막으로 Cheng은 완벽한 사건 함수를 찾는 대신에 경험적인 지식을 바탕으로 사건 함수의 후보 위치를 찾고, 각 위치에 직사각형 창 형태의 초기 사건 함수를 설정한 후, 반복적인 최적화 과정으로 사건 함수와 사건 벡터를 동시에 추정하는 STTD(Short-Term Temporal Decomposition) 방법을 제안하였다[3]. STTD는 평균 160ms에서 210ms가량의 지연 시간만으로 시간적 분해법을 수행하여 음성부호화에 적합한 알고리즘이나, 초기 사건 함수의 위치와 모양을 설정하는 방법이 LSF 파라미터에 적합하지 않다. STTD는 초기 사건 함수의 위치를 입력 벡터의 크기(norm)의 지역 평균(local mean) 값이 극대값을 갖는 지점으로 결정하는데, 필터의 극 위치에 해당되는 LSF 파라미터의 경우, 그 벡터 크기는 음성학적으로 의미를 갖지 못한다. 또 초기 사건 함수가 겹쳐지는 구간에서는 사건 함수의 시간별 합이 1보다 크게 되어, 다음에 설명하는 사건 벡터의 범위 문제를 유발시킨다. 따라서 STTD도 LSF 파라미터의 시간적 분해에는 사용될 수 없다.

III. 제한된 시간적 분해법

이 장에서는 기존의 시간적 분해법의 문제점을 해결하기 위해 고안된 제한된 시간적 분해법에 대해 설명한다. 시점 n의 LSF 파라미터 벡터 $\vec{\phi}(n)$을 시간적 분해법의 벡터 식으로 나타내면 (5)와 같이 사건 벡터 $\vec{\omega}_i$들의 가중치 합으로 표현된다.

$$\vec{\phi}(n) = \sum_{i=1}^{I} \vec{\omega}_i \phi_i(n) \tag{5}$$

기존의 시간적 분해법에서는 사건 함수가 결정되면, 사건 벡터를 최소자승오류법을 이용하여 추정하는데, 이때 $\sum_{i=1}^{I} \phi_i(n) \neq 1$이라면, LSF 파라미터 $\vec{\phi}(n)$의 각 차

수가 0에서 π사이에서 순차적임을 만족시키기 위해 사건 벡터 $\vec{\omega}_j$들은 LSF 파라미터의 범위를 벗어나게 된다. 즉, 추정된 사건 벡터들은 벡터 공간 내에서 LSF 파라미터와는 다른 별개의 분포를 형성하고, 따라서 LSF 파라미터의 양자화 특성들이 반영되지 않는다. 이러한 고찰을 바탕으로, 기존의 사건 함수에 대한 제약 외에 매 시점에서 모든 사건 함수의 합이 1이라는 새로운 제약을 추가하여, 제한된 시간적 분해법을 제안한다.

3.1 제한된 시간적 분해법의 사건 추정법

제한된 시간적 분해법은 시간에 따라 변하는 벡터 열 $\vec{\phi}(n)$을 근사하기 위해서 순차적으로 출현되는 사건 벡터와 사건 함수 쌍 $(\vec{\omega}_j, \phi_j(n))$을 구한다. 단, 구해진 사건 벡터와 함수 쌍은 아래의 (6)과 같은 자승오류합을 최소로 하며, (7)에서 (10)에 이르는 사건 함수의 제약을 만족시킨다.

$$E=\sum_{n=1}^{N}\left\| \vec{\phi}(n)-\sum_{j=1}^{J}\vec{\omega}_j\phi_j(n)\right\|^2 \tag{6}$$

단,

$$0\le\phi_j(n)\le 1,\quad \text{for } 1\le j\le J,\ 1\le n\le N \tag{7}$$

$$\phi_j(c_j)=1,\quad \text{for } 1\le j\le J \tag{8}$$

$$\sum_{n=1}^{N}\phi_i(n)\phi_j(n)=0,\quad \text{if } |i-j|>1 \tag{9}$$

$$\sum_{j=1}^{J}\phi_j(n)=1,\quad \text{for all } n \tag{10}$$

제한된 시간적 분해법의 사건 함수는 (10)의 새로운 제약에 의해서 몇 가지 새로운 성질들이 추가된다. 먼저 임의 사건 함수 $\phi_j(n)$가 1이 되는 고유 위치 c_j에서 $i\neq j$인 사건 함수 $\phi_i(n)$는 모두 0이 되어야 한다. 결과적으로 그 시점의 벡터 궤적 $\vec{\phi}(c_j)$은 해당 사건 벡터 $\vec{\omega}_j$만으로 표시된다. 따라서 제한된 시간적 분해법에서는 최소한 그 고유 위치에서는 보간에 따른 오류가 없도록 사건 함수 고유 위치의 벡터 궤적 $\vec{\phi}(c_j)$을 해당 사건 벡터 $\vec{\omega}_j$의 초기 값으로 취한다.

한편, 각 사건 함수가 단일 구간 내에서만 0이 아닌 값을 갖는 성질을 사건 함수의 시간적인 응집성이라고 한다. 제한된 시간적 분해법에 사건 함수의 시간적인 응집성과 각 사건의 고유 위치의 순차성이 가정되면, j번째 사건 함수가 1이 되는 시점 c_j 이후에는 j 이전의 사건 함수들은 모두 0이 되어야 한다. 마찬가지로 시점 c_j 이전에는 j 이후의 사건 함수들이 모두 0이 되므로, (9)에서 $I=1$이 된다.

(9)의 I를 1로 정하면, 임의 시점의 벡터 궤적은 유효한 두개의 사건 벡터의 보간으로 근사된다. 다음에 벡터 궤적이 사건 벡터 $\vec{\omega}_j$와 $\vec{\omega}_{j+1}$만으로 근사되는 시간 구간 $[c_j, c_{j+1}]$에 대해서, 두 사건 벡터와 대응되는 사건 함수는 (11)을 최소로 한다.

$$\begin{aligned} E &= \sum_{n=c_j}^{c_{j+1}}\left\| \vec{\phi}(n)-\vec{\omega}_j\phi_j(n)-\vec{\omega}_{j+1}\phi_{j+1}(n)\right\|^2 \\ &= \sum_{n=c_j}^{c_{j+1}}\left\| (\vec{\phi}(n)-\vec{\omega}_{j+1})-(\vec{\omega}_j-\vec{\omega}_{j+1})\phi_j(n)\right\|^2 \end{aligned} \tag{11}$$

(11)에서 E의 $\phi_j(n)$에 대한 미분 값을 0으로 하여 방정식을 풀면, 주어진 구간 $[c_j, c_{j+1}]$내의 두 사건 벡터에 대한 최적 사건 함수 값을 쉽게 구할 수 있다. 다만 사건 함수의 범위가 0에서 1사이로 제한되므로, 구해진 $\phi_j(n)$이 0보다 작으면 0으로, 1보다 크면 1로 근사시킨다.

3.2 제한된 시간적 분해 알고리즘

이제 제한된 시간적 분해법을 이용하여 입력 벡터 궤적을 사건 벡터와 사건 함수로 근사하기 위해서는 초기 사건 벡터를 결정할 적절한 고유 위치 설정이 필요하다. 이때 사건 벡터가 음성의 목표 신호가 되기 위해서는 음성의 안정된 구간의 스펙트럼 포락이 표본 되야 한다. 여기서 음성의 안정된 구간이란 주변에 비해 스펙트럼 포락의 변화가 작은 구간으로, LSF 파라미터를 사용하는 경우에는, 그 시계열의 1차 미분치 벡터 크기(norm)가 극소값을 갖는 지점에 해당된다. LSF 파라미터 시계열의 1차 미분치 벡터 크기는 다음의 (12)와 같이 계산된다[11]. 본 연구에서 $M=2$를 사용한다.

$$STM_{LSF}(t)=\left\| \frac{\sum_{t=-M}^{M} t\cdot\vec{\phi}(t+t)}{\sum_{t=-M}^{M} t^2}\right\| \tag{12}$$

$STM_{LSF}(t)$[1]이 극소값이 되는, 시점 t' 위치의 벡터 궤적 $\vec{\phi}(t')$을 이용하여 연속되는 두개 이상의 초기 사건 벡터가 결정되면, 앞서 설명한 방법으로 대응하는 사건 함수들을 계산할 수 있다. 이때 j번째 사건 함수는 j-1번째 사건 벡터의 위치에서부터 j+1번째 사건 벡터의 위치 사이에서만 0과 1사이의 값을 갖고, 자신의 위치에서는 1,양끝 지점에서는 0이 된다.

그러나, $STM_{LSF}(t)$만을 이용해 찾아낸 사건 벡터의 수는 LSF 파라미터 벡터 궤적을 만족할 만큼 근사시키기에는 충분하지 못하다. 이것은 $STM_{LSF}(t)$이 각 차수의 미분치를 한꺼번에 고려하기 때문에 각 차수별 변화 특성을 반영하지 못한 결과이다. 따라서 $STM_{LSF}(t)$만을 이용하여 결정된 사건 벡터만으로 주어진 LSF 벡터 궤적을 근사하면, 몇몇 지점에서 간과할 수 없는 오류가 발생

1) Spectral Transition Measure의 약자임[15].

된다. 이러한 현상을 없애기 위해서는 보간 오류가 큰 지점에 새로운 사건 벡터를 추가할 필요가 있다. 즉, $STM_{LSF}(l)$으로 구한 두 사건 벡터 사이의 벡터 궤적을 근사시킨 후, 경험적인 수치 δ 이상의 보간 오류가 발생된 지점을 새로운 사건 벡터의 위치로 삼는다. 실험적으로 이 방법은 사건 벡터의 수를 크게 늘리지 않으면서 제안된 시간적 분해법의 성능을 크게 향상시킴을 확인하였다.

스펙트럼 포락의 안정성을 바탕으로 결정된 초기 사건 벡터는 STTD에서와 마찬가지로 재추정될 수 있으며, 이러한 재추정 결과는 벡터 궤적의 근사 오류를 줄일 수 있다[1, 2, 3]. 사건 벡터의 재추정 값은 다음과 같이 사건 함수들로 구성된 행렬의 가상 역행렬(pseudo inversion)을 써서 구할 수 있다.

$$\Omega = \Psi\Phi^T(\Phi\Phi^T)^{-1} \tag{13}$$

위의 식에서 $(\Phi\Phi^T)$는 $J\times J$ 정방 행렬로, 사건 함수의 제약 (9)에 의해 대각선 위아래의 값만 0이 아닌 삼차 대각 행렬(tridiagonal matrix)이고, 그 역함수는 O(J)의 연산만에 쉽게 구할 수 있다[12]. 다만 사건 함수의 재추정 과정에서는 LSF 파라미터의 순차성이 위배되는 결과가 발생될 수 있으며, 이런 경우에는 바로 전 차수의 LSF값으로 강제적으로 추정치를 조정한다. 실험 결과 사건 벡터의 재추정 및 그에 따른 사건 함수의 재추정 과정은 대부분 5회 미만으로 결과가 수렴되며, 시간적 분해 결과의 성능을 크게 향상시켰다.

이상의 설명을 정리하여, 본 연구에서 제안하는 제한된 시간적 분해 알고리즘을 보인다. 본 연구에서는 시간적 분해 결과가 전후 사건의 영향을 받지 않도록 임의 사건 벡터 및 함수는 전후 두개의 사건과 함께 최적화 된다.

1단계:시점 0 위치에 첫 번째 사건 벡터를 설정한다.

2단계: $STM_{LSF}(l)$의 극소값을 찾아 다음 사건 벡터를 설정한다.

3단계:첫 번째 사건 벡터의 위치로부터 마지막 사건 벡터의 위치까지 설정된 사건 벡터들에 대응하는 사건 함수를 추정한다.

4단계:3단계 결과 오류가 한계치를 넘은 시점 있으면, 그 위치에 새로운 사건 벡터를 설정하고 3단계를 반복한다. 초기 사건 벡터가 모두 설정되면 5단계로 진행한다.

5단계:추정된 사건 함수를 이용하여 사건 벡터를 재추정한다. 단, 첫 번째 사건 벡터가 이전 분석의 결과인 경우는, 재추정 결과를 사용하지 않고, 그 값을 유지시킨다.

6단계:재추정된 사건 벡터를 이용하여 3단계와 같은 방법으로 사건 함수를 재추정한다. 재추정 결과가 수렴하거나, 미리 정해 놓은 횟수만큼 반복될 때까지 제5단계와 6단계를 반복 수행한다.

7단계:최종 결정된 사건 중 마지막 두개의 사건 벡터는 다음 분석을 위해 각각 첫 번째와 두 번째 사건 벡터로 기억해 두고, 나머지 결과들을 양자화하여 전송 혹은 저장한다.

8단계:입력 벡터 궤적의 종점에 이르기까지 제2단계에서 7단계를 반복한다.

한편, LSF 파라미터를 시간적 분해법으로 근사할 때 생기는 보간 오류를 계산함에 있어, LSF 파라미터의 차수간 차이가 적은 곳에 주파수 응답 값이 커지는 경향을 이용하여, (15)와 같이 가중치를 두는 방법이 널리 이용되고 있다[6, 8, 10]. 본 연구에서도 LSF 파라미터의 시간적 분해에는 LSF 파라미터의 가중치를 적용한 보간 오류를 사용하도록 (14)와 같이 최소화 기준을 변경한다. 이때 $\phi_0(n)=0$이고, $\phi_{p+1}(n)=\pi$로 정한다.

$$E_{LSF} = \sum_{n=1}^{N}\sum_{k=1}^{p} W(k,n)[\phi_k(n) - \sum_{j=1}^{J}\omega_{k,j}\phi_j(n)]^2 \tag{14}$$

단,

$$W(k,n) = \frac{1}{\phi_k(n)-\phi_{k-1}(n)} + \frac{1}{\phi_{k+1}(n)-\phi_k(n)} \tag{15}$$

IV. 실험 및 결과

이 장에서는 제한된 시간적 분해법을 이용하여 LSF 파라미터 시계열을 양자화할 때의 성능을 평가한다. 본 실험에 사용된 음성 자료로는 미국 NIST(National Institute of Standard and Technology)에서 연속 음성 인식 시스템 개발을 위해 수집한 TIMIT 음성 자료의 일부를 사용한다. 다음 표 1에 사용된 TIMIT 음성 자료의 내역을 정리하였다. 표에서 문장 종류 SI는 발음성 다양한 문장(phonetically-diverse sentence)을 의미한다.

표 1. 실험에 사용한 음성 자료의 내역

	문장 형태	화자 수	화자 당 문장 수	총 문장 수
학습 데이터	Diverse (SI)	462	3	1386
평가 데이터	Diverse (SI)	168	3	504
총 계	.	630	.	1890

TIMIT 음성 자료는 16Khz로 표본되어 있으나, 본 실험에서는 이를 50탭 FIR 필터를 이용한 저주파 대역 필터링 후에 8Khz로 다운 샘플링 하여 사용하였다. 음성 자료 중에 학습 자료는 시간적 분해의 결과로 구해진 사 파라미터에 대한 양자화기(quantizer)를 학습시키는데 사용된다.

LSF 파라미터 계산 방법은 다음과 같다. 8Khz로 샘플링 되고, 16비트로 양자화된 입력 음성에 30ms Hamming 창을 씌운 후, 자기 상관 계수법을 이용하여 10차 LPC

계수를 계산한다. 다음으로 LPC 계수는 10차 LSF 파라미터로 변경된다. 이때 분석 창은 20ms씩 이동되며, 결과적으로 LSF 파라미터의 갱신율(updating rate)은 50Hz가 된다.

제한된 시간적 분해법을 이용한 LSF 파라미터의 양자화 실험은 크게 두 단계로 나뉜다. 먼저 입력 음성의 LSF 파라미터 벡터 열을 시간적 분해법으로 분석된 사건 벡터와 사건 함수들로 근사시키는 보간 결과를 평가한다. 다음으로는 보간에 쓰인 모든 파라미터들을 양자화한 후, 그 오차를 측정하는 양자화 성능 평가를 수행한다.

4.1 시간적 분해법을 이용한 보간 결과

시간적 분해법을 이용한 LSF 파라미터의 보간에 앞서, 사건 벡터 및 사건 함수 재추정 과정의 유효성을 타진하기 위한 실험을 수행하였다. 다음의 그림 1은 평가 음성 자료 중의 한 문장에 대하여, 재추정 횟수를 변화시켜 가며, 보간 오류 E_{LSF}를 측정한 결과이다. 사용된 문장은 4.98초분으로 제안한 시간적 분해법을 적용한 결과 95개의 사건을 발생시켰다.

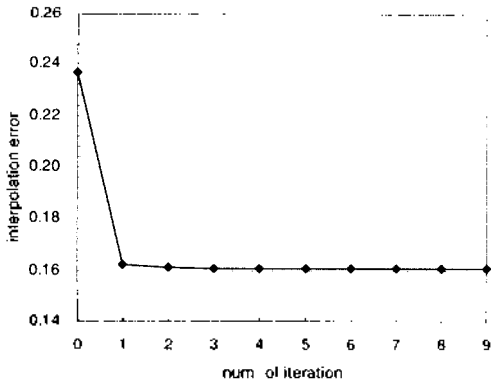

그림 1. 재추정 횟수에 따른 보간 오류의 변화

그림에서 보듯이 재추정 과정은 보간 오류를 줄이는데 매우 효과적이며, 대략 5회 정도의 재추정으로 오류값이 수렴됨을 알 수 있다. 따라서 본 연구에서는 제한된 시간적 분해법의 재추정 과정이 항상 5회 반복되도록 설정하였다.

이제 시간적 분해법을 이용한 LSF 파라미터 벡터 궤적의 보간 결과를 평가하기 위해서 평균 예측 이득(prediction gain)을 비교한다. 평균 예측 이득은 그 값이 클수록, 사용된 LSF 파라미터가 입력 음성 $s(n)$의 스펙트럼 정보를 정확히 표현하고 있음을 의미한다. LSF 파라미터의 예측 이득은 매 프레임마다 구해진 잔차 신호에 대한 입력 음성의 에너지 비율로 (16)과 같이 계산된다[13]. 이때 LPC 계수 a_j는 해당 프레임의 LSF 파라미터를 변환하여 계산된 값이다.

$$PGain = 10\log 10\left(\frac{E\{s^2(n)\}}{E\left\{\left[s(n)-\sum_{j=1}^{p}a_j s(n-j)\right]^2\right\}}\right)\ [dB] \quad (16)$$

보간 결과의 평가에 보간 오류 E_{LSF}를 사용하지 않는 이유는, LSF 파라미터의 보간은 양자화에 앞서는 전처리 과정으로, 반드시 보간 전의 LSF 파라미터가 입력 음성의 스펙트럼을 너 잘 표현하고 있다고는 볼 수 없기 때문이다. 즉, 보간 전후의 예측 이득을 비교하여 제한된 시간적 분해법에 의한 보간이 LSF 파라미터 궤적에 미치는 영향을 판단하기 위함이다.

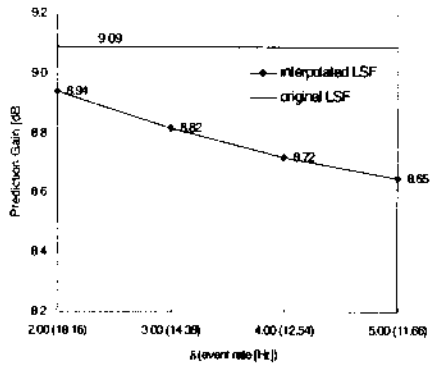

그림 2. 보간된 LSF 파라미터의 평균 예측 이득

그림 2는 제한된 시간적 분해법의 사건 벡터 추가를 위한 문턱치 δ를 변화시켜 가면서 평균 예측 이득을 관찰한 결과이다. 음성 자료로는 TIMIT 평가 음성 자료 504 문장을 사용하였다. 그림에서 괄호 안에 표시된 숫자는 해당 δ값을 사용할 때에 대한 평균적인 사건 벡터의 발생 빈도이다.

실과적으로 문턱치 δ가 2.00일 때, 평균 사건 발생 빈도는 18.16Hz로 보통 속도 음성의 음소 발생 빈도인 15Hz에 근접하고 있으며, 이때의 평균 예측 이득은 8.94dB로 원래 LSF 파라미터의 평균 예측 이득 9.09dB에 비해 0.15dB가 감소되었다. 그러나 이 정도의 예측 이득의 감소는 무시할 수 있으며, 실제로 원래 LSF 파라미터를 이용한 잔차 신호와 보간된 LSF 파라미터를 이용하여 합성한 재생음은 원음과 청각적으로 구분할 수 없을 정도로 좋은 음질을 나타냈다. 다음의 그림 3에는 제한된 시간적 분해법으로 입력 음성 "한국"을 분석한 결과이다. (a)는 입력 음성을, (b)는 각 시점에 대응하는 사건 함수의 모양을, (c)는 각 사건 함수에 대응하는 사건 벡터, 즉 LSF 파라미터에 의해 결정되는 스펙트럼 포락을 각각 나타내었다.

한편 δ가 2.00일 때, 각 사건 벡터의 고유 위치간 거리의 분포는 그림 4와 같다. 이때 거리가 1프레임, 즉 20ms인 경우가 총 7,052번으로 전체 사건의 수 31,098의 22.7%에 이른다. 가장 많은 빈도를 보인 거리는 40ms로 타 연구자의 실험 결과와 일치하고 있다[3].

4.2 제안된 방법의 양자화 성능 평가

시간적 분해법으로 보간된 LSF 파라미터는 전송이나 저장을 위해 모두 이산적인 값으로 변화되어야 한다. 먼저 사건 벡터는 LSF 파라미터의 성질을 유지하므로, 기존의 LSF 파라미터 양자화 기법을 이용하여 양자화할 수 있다. 본 연구에서는 구현이 용이한 매 차수별 dLSF값의 스칼라 양자화(Scalar Quantization)를 적용하였다.

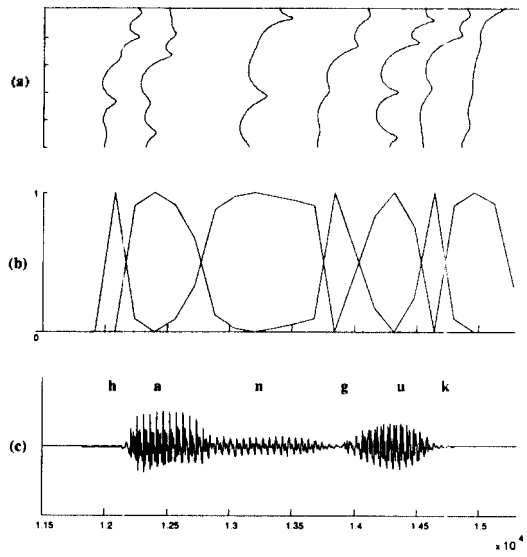

그림 3. 제한된 시간적 분해의 예

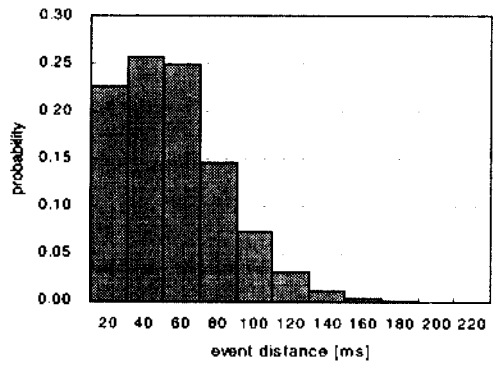

그림 4. 사건 벡터간 거리 분포

사건 함수는 영이 아닌 구간의 위치, 길이, 모양의 세 가지 정보로 나뉘어 양자화된다. 단, 제한된 시간적 분해법에서는 임의의 사건 함수 $\phi_j(n)$에 대하여 다음의 관계가 성립된다.

$$\phi_j(n)=\begin{cases}0 & , \text{ if } n\le c_{j-1}\\ 1-\phi_{j-1}(n) & , \text{ if } c_{j-1}<n<c_j\\ 1 & , \text{ if } n=c_j\\ 1-\phi_{j+1}(n) & , \text{ if } c_j<n<c_{j+1}\\ 0 & , \text{ if } n\ge c_{j+1}\end{cases} \tag{17}$$

따라서 사건 함수 $\phi_j(n)$의 값은 자신의 고유 위치 c_j에서부터 다음 사건 함수의 위치 c_{j+1}까지만 전달하면, (17)에 의해 전체 구간이 복원될 수 있다. 즉, 사건 함수의 모양 정보는 구간 $[c_j, c_{j+1}]$의 함수 값으로 양자화될 수 있다. 본 연구에서 사건 함수의 모양은 구간 $[c_j, c_{j+1}]$의 함수 값을 구간 길이가 10이 되도록 신형 보간을 하고, 10차원 벡터 양자화(Vector Quantization)하였다.

사건 함수 $\phi_j(n)$의 길이 $l(j)$은 다음 사건 함수까지의 거리 $p(j)$를 이용하여 (18)로 나타낼 수 있다.

$$l(j)=c_{j+1}-c_{j-1}=p(j)+p(j-1) \tag{18}$$

단,

$$p(j)=c_{j+1}-c_j \tag{19}$$

결국, 사건 함수의 위치 및 길이 정보는 $p(j)$로 양자화할 수 있다. 더불어 $p(j)=1$인 경우에는 해당 사건 함수의 모양 정보는 전송할 필요가 없게 된다. 실험 결과 그림 4와 같이 두 사건 함수간의 거리는 최대 220ms, 즉 11 프레임을 넘지 않으므로 표 2와 같이 4비트를 써서 양자화할 수 있다. 단, 길이 1인 경우는 상위 3비트만을 사용하여 표현한다. 분석된 음성의 연속적인 재생을 위해서는 두 사건 함수간의 최대 거리만큼의 버퍼링(buffering)이 필요하므로 본 실험에서 구현한 LSF 양자화 방법의 지연시간은 220ms로 볼 수 있다.

표 2. 사건 함수 길이의 양자화 코드

p(i)	1	2	. . .	15
code	000x	0010	. . .	1111

스펙트럼 포락 정보를 양자화하는데 발생되는 오류는 평균 스펙트럼 왜곡(spectral distortion measure)을 사용한다. 임의 프레임의 스펙트럼 포락이 $S(\omega)$로, 양자화된 스펙트럼 포락이 $S'(\omega)$로 표현된다면, 스펙트럼 오차는 다음의 (20)으로 계산된다[14].

$$SD=\left[\frac{1}{\pi}\int_0^{\pi}(10\log_{10}S(\omega)-10\log_{10}S'(\omega))^2 d\omega\right]^{\frac{1}{2}}\ [dB] \tag{20}$$

한편, 스펙트럼 포락 정보의 양자화 방법에 대한 평균적인 스펙트럼 왜곡이 1dB에 가깝고, 2dB를 넘은 프레임의 전체 프레임의 2%미만이며, 4dB를 넘은 프레임이 0%에 가까우면, 이를 스펙트럼 포락의 투명한 양자화(transparent quantization)라 한다. 실험적으로 투명한 양자화에 의한 정보 손실은 무시될 수 있다고 보고되었다[5].

다음 표 3에 사건 벡터에 대한 스칼라 양자화기의 비트 수를 31비트에서 33비트까지 변화시키고, 사건 함수의

모양 정보에 대한 벡터 양자화기의 코드북 크기를 16에서 64까지 변화시킬 때, 평균 스펙트럼 왜곡을 표시하였다. 이때 각 양자화기는 학습 음성 자료를 사용하여 미리 최적화 하였다.

표 3에서 사전 벡터, 즉 LSF 파라미터 양자화에 33비트를, 시간 함수의 모양 정보 양자화에 8비트를 할당할 때, 평균 스펙트럼 왜곡은 0.933dB이었다. 이보다 2dB

표 3. LSF 파라미터 양자화에 따른 평균 스펙트럼 왜곡

SD[dB] (type1[%]) (type2[%])		LSF SQ Bit Allocation		
		31bits (3,3,3,3,4, 3,3,3,3,3,)	32bits (3,3,3,3,4, 3,4,3,3,3)	33bits (3,3,3,3,4, 3,4,3,4,3)
사전 위수 모양 VQ Bit Allocation	4bits	1.126 (4.89) (0.16)	1.082 (3.60) (0.08)	1.054 (2.92) (0.04)
	5bits	1.070 (3.97) (0.16)	1.023 (2.86) (0.07)	0.994 (2.20) (0.04)
	6bits	1.011 (3.25) (0.15)	0.963 (2.17) (0.07)	0.933 (1.57) (0.03)

이상 스펙트럼이 왜곡되는 프레임(type1 error)은 1.57%이고, 4dB 이상인 프레임(type2 error)은 0.03%로 제안한 양자화 방법이 제한된 시간적 분해법으로 보간한 LSF 파라미터를 무방하게 양자화함을 알 수 있다. 또한 사전 벡터의 발생 빈도가 평균 18.16Hz이고, 시간 함수의 위치 정보가 1인 경우가 평균 4.12Hz로 발생하므로, 제안된 양자화 방법의 전송률은 표 4와 같이 계산된다.

표 4. 제한된 시간적 분해법을 이용한 LSF 파라미터 양자화의 전송률

	사전 벡터 LSF SQ	사전 함수		발생 빈도 [Hz]	전송률 [bps]
		위치	모양		
p(j) > 1	33	4	6	14.04	603
p(j) = 1	33	3	0	4.12	149
총 계	752 bps				

V. 결 론

음성의 스펙트럼 정보를 표현하는 여러 방법 중에 선스펙트럼 주파수(LSF) 파라미터는 양자화 특성이 뛰어나고, 안정성 보장이 쉬워, 현재 저전송률 음성부호기에 널리 이용되고 있다. 따라서 선스펙트럼 주파수의 효율적인 양자화 방법은 그 파급 효과가 크고, 활용 가능성이 높다.

본 논문에서는 LSF 파라미터의 효과적인 양자화를 위해 제한된 시간적 분해법을 제안하였다. 기존의 시간적 분해법이 LSF 파라미터의 특수한 성질을 보존하지 못하는 문제점을 해결하기 위하여, 제안된 방법에서는 사전 함수값에 새로운 제약을 두어, 추정된 사전 벡터가 LSF 파라미터의 성질을 유지하도록 하였다. 즉, 추정된 사전 벡터는 LSF 파라미터의 순차성과 0에서 π 사이의 값의 범위를 유지하며, 가장 안정한 스펙트럼 포락을 대표하게 된다. 실험적으로 제안된 방법을 이용하여 구해진 사전 벡터는 일반적인 LSF 파라미터와 동일한 방법으로 양자화될 수 있었다.

실험 결과 음성의 LSF 파라미터 벡터 열은 제한된 시간적 분해법으로 보간하는 전처리를 거쳐, 평균 752bps로 무방한 양자화가 가능하였다. LSF 파라미터의 갱신율을 50Hz로 보면, 제안한 양자화 방법은 프레임마다 15.04비트로 양자화하는 효과를 준다. 한편 사전 벡터의 분포 특성 및 인접 사전과의 연관성을 이용하여 벡터 양자화를 수행한다면 더욱 효과적인 방법이 제안될 것으로 기대된다.

참 고 문 헌

1. B. S. Atal, "Efficient Coding of LPC Parameters by Temporal Decomposition," *Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 81-84, 1983.
2. Astrid M.L. van Dijk-Kappers, Stephen M. Marcus, "Temporal Decomposition of Speech," *Speech Communication*, Vol.8, No.2, pp. 125-135, June 1989.
3. Yan-Ming Cheng, Douglas O'Shaughnessy, "On 450-600 b/s Natural Sounding Speech Coding," *IEEE Trans. on Speech and Audio Processing*, Vol.1, No.2, pp. 207-219, April 1993.
4. Frank K. Soong, Biing-Hwang Juang, "Optimal Quantization of LSP Parameters," *IEEE Trans. on Speech and Audio Processing*, Vol.1, No.1, pp. 15-24, January 1993.
5. Kuldip K. Paliwal, Bishnu S Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame," *IEEE Trans. on Speech and Audio Processing*, Vol.1, No.1, pp. 3-14, January 1993.
6. Ravi P Ramachandran, Man Mohan Sondhi, Nambi Seshadri, Bishnu S. Atal, "A Two Codebook Format for Robust Quantization of Line Spectral Frequencies," *IEEE Trans. on Speech and Audio Processing*, Vol.3, No.3, pp. 157-167, May 1995.
7. Chih-Chung Kuo, Fu-Rong Jean, Hsiao-Chuan Wang, "Low Bit-Rate Quantization of LSP Parameters Using Two-Dimensional Differential Coding," *Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 1-97-100, 1992.
8. Aweke N. Lemma, W. Bastiaan Kleijn, Ed. F. Deprettere, "LPC Quantization Using Wavelet Based Temporal Decomposition of the LSF," *EUROSPEECH'97*, pp. 1259-1262, 1997.
9. M. Yong, G. Davidson, A. Gersho, "Encoding of LPC Spectral Parameters Using Switched-Adaptive Interframe Vector Prediction," *Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 402-405, 1988.
10. Rajiv Laroia, Nam Phamdo, Nariman Favardin, "Robust and Efficient Quantization of Speech LSP Parameters

Using Structured Vector Quantizers," *Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 641-644, 1991.
11. Lawrence Rabiner, Biing-Hwang Juang, *Fundamentals of Speech Recognition*, Prentice Hall International INC., 1993.
12. William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992.
13. Sadaoki Furui, *Digital Speech Processing: Synthesis, and Recognition*, Marcel Dekker, INC., 1991.
14. A. M. Kondoz, *Digital Speech Coding for Low Bit Rate Communication Systems*, John Wiley & Sons, INC, 1994.
15. Sadaoki Furui, "On the Role of Spectral Transition for Speech Perception," Journal of Acoustic Society of America 80(4), pp. 1016-1025, 1986.

▲김 승 주(Sung Joo Kim)

1992년 2월 : 한국과학기술대학 전산학과(학사)
1994년 2월 : 한국과학기술원 전산학과(석사)
1994년 3월 ~ 현재 : 한국과학기술원 전산학과 박사과정 재학중
※ 주관심분야 : 저전송율 음성코딩, 규칙형 음성합성, 실시간 음성통신 시스템

▲오 영 환(Yung Hwan Oh)

1972년 : 서울대학교 공과대학(학사)
1974년 : 서울대학교 교육대학원(석사)
1980년 : Tokyo Institute of Technology 정보공학전공(박사)
1981년 ~ 1985년 : 충북대학교 컴퓨터공학과 조교수
1983년 ~ 1984년 : University of California, Davis 연구교수
1995년 ~ 1996년 : Carnegie-Mellon University 연구교수
1985년 ~ 현재 : 한국과학기술원 전산학과 교수
※ 주관심분야 : 음성인식, 음성합성, 음성코딩, 화자인식, 대화관리, 신경회로망, 전문가 시스템