한국음향학회지 제24권 제4호 pp. 217-223 (2005)

분산 음성 인식 시스템을 위한 특징 계수 양자화 방식 설계

Design of a Quantization Algorithm of the Speech Feature Parameters for the Distributed Speech Recognition

이 준 석 *, 강 상 원* , 윤 병 식**
(Joonseok Lee*, Sangwon Kang*, Byungsik Yoon**)

*한양대학교 전자컴퓨터공학부, **한양대학교 전자전기제어계측공학과 대학원 박사과정
(접수일자: 2005년 2월 28일; 수정일자:2005년 5월 3일; 채택일자: 2005년 5월 6일)

본 논문에서는 분산 음성 인식 시스템에서 사용되는 멜켑스트럼 계수를 양자화 하기 위하여 예측 구조를 갖는 BC-TCQ 양자화기를 제안하였다. 분산 음성 인식 시스템을 위한 효율적인 멜켑스트럼 계수 양자화기를 설계하기 위하여, 인접 프레임간의 높은 상관도를 이용한 1차 AR 예측 필터를 적용하였다. 그리고 예측 필터에 의해서 구해지는 예측 에러 벡터는 BC-TCQ를 사용하여 양자화를 수행하였다. 본 연구에서 제안된 예측 BC-TCQ 멜켑스트럼 계수 양자화기는 분산 음성 인식 시스템을 위해 ETSI 규격에서 사용되는 split VQ 멜켑스트럼 계수 양자화 방식보다 cepstral distortion (CD) 측면에서 훨씬 좋은 성능을 보이며, 인코딩 연산 복잡도 및 메모리 요구량에서도 더 유리하다.

핵심용어: 멜켑스트럼, 분산 음성 인식, BC-TCQ, 벡터 양자화

투고분야: 음성처리 분야 (2,5)

In this paper, we propose a predictive block constrained trellis coded quantization (BC-TCQ) to quantize cepstral coefficients for the distributed speech recognition. For prediction of the cepstral coefficients, the 1st order auto-regressive (AR) predictor is used. To quantize the prediction error signal effectively, we use a BC-TCQ. The performance is compared to the split vector quantizers used in the ETSI standard, demonstrating reduction in the cepstral distance and computational complexity.

Keywords: Mel-cepstrum, Distributed speech recognition, BC-TCQ, Vector quantization

ASK subject classification: Speech Signal Processing (2.5)

I. 서 론

현재 분산 음성 인식 시스템에서 가장 널리 사용되는 특징 벡터는 멜켑스트럼 계수이다. 멜켑스트럼 계수는 일정한 대역내에서 여러 대역의 소리가 합쳐진 소리는 그 대역의 중간 주파수 대역에서의 소리와 동일하게 들리는 특성을 이용한 것이다. 즉, 일반적인 주파수 단위를 특징에 맞게 매핑시켜 특징 벡터를 구하고 켑스트럼 필터를 적용하여 그 특징을 보다 강인하게 표현하는 방법이다. 분산 음성 인식 시스템 에서는 위의 특징벡터를 양자화하여 모바일 채널 환경에서 송수신하게 되는데, 송신단에서는 특징벡터를 기존 모델인 코드북과 매핑시켜 대표값으로 부호화하고, 수신단에서는 부호화된 특징벡터를 이용해서 유사도 측정 및 인식과정을 거친다.

본 연구에서 특징 벡터 양자화를 위해 적용한 trellis coded quantization (TCQ)[1]은 일종의 벡터 양자화 방식으로 부호화에 요구되는 벡터 코드북을 벡터를 이루는 각 요소에 대응하는 스칼라 코드북으로 구성하며, 길쌈 부호화기로 trellis 구조를 표현한다. Viterbi 알고리즘 [2]을 이용하여 최적 인코딩을 위한 trellis 경로를 찾으며, trellis 부호화 양자화 기법은 연산 복잡도가 unstructured VQ에 비하여 훨씬 작은 특성을 보인다. 전통적인 trellis 부호화 양자화 기법은 viterbi 알고리즘 탐색 후

책임저자: 강 상 원 (swkang@hanyang.ac.kr)
경기도 안산시 사1동 1271번지 한양대학교 전자컴퓨터공학부
(전화: 031)400-5215; 팩스: 031)418-8476)

결정된 trellis 경로의 초기 스테이트를 추가 정보로서 전송하여야 한다. 이러한 부가 전송 정보는 차수가 큰 소스 백터에 대해서는 큰 영향을 끼치지 않으나, 차수가 작은 소스에 대하여서는 rate-distortion 관점에서 상당이 중요한 걸림돌로 작용한다. 이와 같은 TCQ 방식의 정보 전송에 대한 단점을 해결하기 위하여 Nikneshan과 Khandani이 tail-biting TCQ (TB-TCQ) [3]을 소개하였고, 이들이 제안한 TB-TCQ는 차수가 작은 소스 백터에 대하여 VQ와 비교할 만한 성능을 보인다.

본 연구에서는 멜켑스트럼 계수 양자화를 위하여 TB-TCQ개념을 변형한 예측 block constrained TCQ (BC-TCQ)[4]를 적용하였다. 본 연구에서 적용된 예측 BC-TCQ 멜켑스트럼 계수 양자화기는 ETSI 규격[5] 에서 분산 음성 인식 시스템을 위해 사용되는 split VQ 멜켑 스트럼 계수 양자화 방식 보다 cepstral distortion (CD)[6]측면에서 더 좋은 성능을 보이며, 인코딩 연산복잡도 및 메모리 요구량에서도 이득을 얻었다.

II. Block-Constrained TCQ

2.1. TCQ 기본 구조

본 연구에서 기본적으로 사용하는 Trellis 부호화 양자화 기법은 rate-1/2 길쌈 부호화기로 구성된 trellis 구조를 가지며, feed-back free 인코더 구조를 가진다. 전체 trellis 구조의 스테이트 수는 $N=2^{\nu}$ 개 이며, 각 trellis 스테이트로 들어오는 혹은 나가는 branch는 2개이다.

그림 1은 길쌈 부호화기와 4-스테이트 trellis 구조를 나타낸 것이며 각 branch는 길쌈 부호화기의 출력 값이 할당되어 있다. 모든 trellis 스테이트를 초기 스테이트로 선택 가능한 trellis 부호화 양자화 기법은 모든 초기 경로 메트릭을 '0' 혹은 같은 값으로 초기화 한 후, viterbi 알고리즘 인코딩을 마지막 스테이지까지 실행함으로써 최소 거리 값을 갖는 trellis 경로를 결정할 수 있다. Trellis 스테이트 수를 N 개라 가정했을 때 이와 같은 인코딩은 임의의 스테이트에서 시작하여 임의의 스테이트로 끝나는 N^2 개의 trellis 경로 중 하나의 최적 경로가 결정된다. n 개의 샘플로 구성된 블록 소스를 이러한 trellis 부호화 기법을 이용하여 양자화 하는 경우, trellis 경로 정보를 전송하기 위하여 trellis 경로의 초기 스테이트를 나타내기 위한 ν 비트 및 전체 n 스테이지의 transition 정보를 나타내기 위한 n 비트가 필요함으로 전체 $n+\nu$ 비트가 요구된다.

일반적으로 작은 사이즈의 블록 소스인 경우 각 샘플에 대한 ν/n 비트의 overhead 비트는 rate-distortion 측면에서 무시 할 수 없는 비트이다. 이러한 trellis 부호화 양자화 기법을 TCQ 인코딩 방식이라 한다. 그림 2는 4-스테이트 trellis 구조에서 TCQ 인코딩 방식으로 결정된 trellis 경로 (굵은 점선) 정보를 전송하기 위해 필요한 초기 스테이트 정보와 스테이트 transition 정보를 보였다.

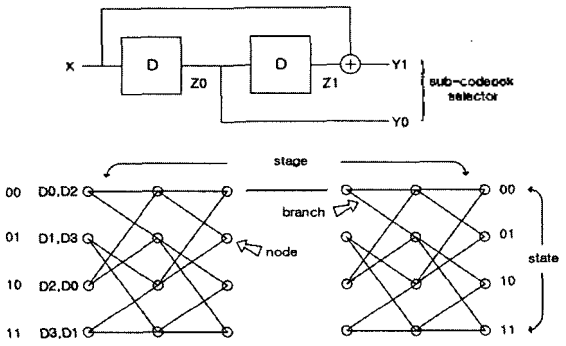

그림 1. 길쌈 부호화기 및 4-스테이트 trellis 구조
Fig.1. Convolutional encoder and four-state trellis.

2.2. Tail-Biting TCQ 구조

Nikneshan과 Khandani는 TCQ 방식에 비하여 좋은 성능을 보이는 TB-TCQ 방식[3]을 제안하였다. 일반적으로, trellis 경로의 초기 스테이트 마다 하나의 마지막 스테이트가 결정되도록 제약을 주는 것을 tail-biting이라 한다. TB-TCQ의 인코딩 과정은 N-스테이트 trellis의 각 스테이트를 초기 스테이트로 가정하고 viterbi 인코딩을 적용함으로서 전체적으로 N 번의 viterbi 알고리즘 인코딩이 필요하며, 각 초기 스테이트에서 시작되는 N 개의 trellis 경로 중 최소 distortion을 나타내는 초기 스테이트와 trellis 경로가 결정된다. TB-TCQ에서

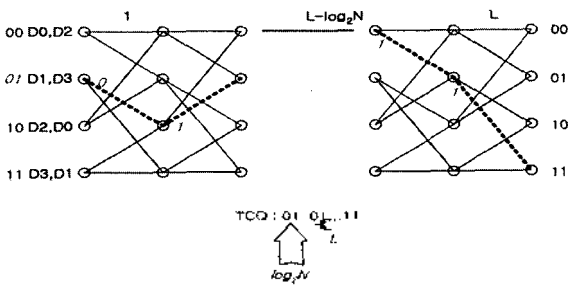

그림 2. 4-스테이트 TCQ 인코딩 방식에서의 trellis 경로 정보
Fig.2. Trellis path information in the 4-state TCQ encoding.

각 초기 스테이트에 따른 마지막 스테이트는 결정되어 있으므로 $n-\nu$ 스테이지에서 마지막 스테이지의 특정 스테이트로의 trellis 경로는 결정된다. 따라서 마지막 ν 스테이지에 대한 스테이트 transition 정보는 전송 할 필요가 없다. 이러한 TB-TCQ 인코딩 방식의 큰 단점은 viterbi 알고리즘 인코딩을 trellis 스테이트 수만큼 실행하여야 하므로 TCQ에 비하여 N 배의 연산 복잡도가 요구된다.

그림 3은 4-스테이트 trellis구조의 TB-TCQ 인코딩 방식으로 결정된 trellis 경로 정보를 이루는 초기 스테이트 정보와 $L-\nu$ 스테이지까지의 스테이트 transition 정보를 나타내었으며, 그림 3에 제시된 TB-TCQ는 시작 스테이트와 마지막 스테이트가 같다는 제약조건이 적용되었다. 따라서 마지막 2개 스테이지의 스테이트 transition정보는 초기 스테이트와 마지막 스테이트의 제약조건 때문에 unique하므로 전송할 필요가 없다.

2.3. Block-Constrained TCQ 구조

본 연구에서 적용된 BC-TCQ는 N -trellis 스테이트 중 $2^k, 0 \le k \le \nu$ 개의 초기 스테이트를 가지며, 각각의 초기 스테이트마다 $2^{\nu-k}$ 개의 마지막 스테이트를 가지도록 제약된 trellis 구조를 가진다. 인코딩 과정은 다음과 같이 진행된다.

Step (1) 2^k 개의 초기 스테이트에 대한 경로 메트릭은 '0'로 초기화 시키며, 나머지 스테이트의 경로 메트릭은 '∞'로 초기화 시킨다.

Step (2) 한번의 viterbi 알고리즘 탐색이 $n-\nu$ 스테이지까지 진행되며, viterbi 알고리즘 탐색에 의해 $n-\nu$ 스테이지까지 결정된 trellis 경로의 초기 스테이트 정보에 따라 마지막 ν 스테이지의 trellis 경로는 결정 지어 진다.

Step (3) Step (2)에서 결정된 trellis 경로 정보는 k 비트의 초기 스테이트 정보와 $n-\nu$ 스테이지까지의 각 trellis 스테이트에서의 스테이트 transition 정보를 나타내기 위한 $n-\nu$ 비트와 $n-\nu$ 스테이지에서 마지막 스테이지의 $2^{\nu-k}$ 개 스테이트 중 선택된 스테이트를 나타내기 위한 $\nu-k$ 비트로 구성된다. 이러한 인코딩 방식은 k 값에 관계없이 한번의 viterbi 알고리즘 탐색으로 이루어 진다.

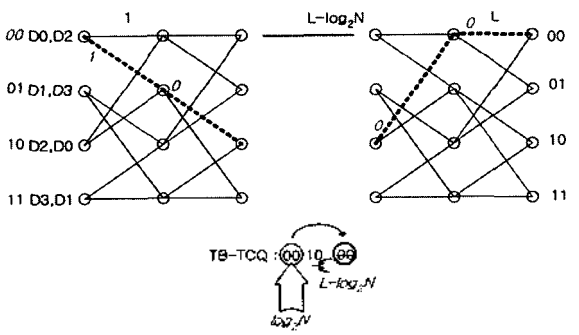

그림 3. Tail-biting TCQ 인코딩 방식에서의 trellis 경로 정보
Fig.3. Trellis path information in the tail-biting TCQ encoding.

TB-TCQ와 BC-TCQ의 연산 복잡도를 다시 한번 비교하면, TB-TCQ의 경우 N-스테이트 trellis의 각 스테이트 마다 $n-\nu$ 스테이지에서 N 개의 survivor 경로가 결정되며, 전체 N trellis 스테이트를 모두 초기 스테이트로 고려 할 경우 전체 N^2 개의 survivor path가 $n-\nu$ 스테이지까지 결정된다. BC-TCQ의 경우, ($k=\nu$ 인 경우), 전체 N trellis 스테이트를 초기 스테이트로 사용할 수 있으며, $n-\nu$ 스테이지까지 결정된 N 개의 survivor 경로는 TB-TCQ에서 고려된 N^2 개의 survivor 경로들 중 일부분이다. 이러한 N 개의 survivor 경로는 $n-\nu$ 스테이지까지 단일 viterbi 알고리즘 탐색에 의해 각 trellis 스테이트까지 선택된 최소 distortion을 나타내는 trellis 경로이다. BC-TCQ는 $n-\nu$ 스테이지의 각 trellis 스테이트에서 나머지 ν 개의 스테이지에서는 유일한 trellis 경로만을 가짐으로서 TB-TCQ의 경우보다 제한적이다. 특히 $k<\nu$ 인 경우, BC-TCQ 인코딩에서는 전체 N 개의 trellis 스테이트 중 일부 trellis 스테이트만을 초기 스테이트로 허용하므로, $n-\nu$ 스테이지까지 결정된 N 개의 survivor 경로는 TB-TCQ의 N^2 개 survivor 경로의 일부분인 $N2^k$ 개의 trellis 경로 중에서 결정된 것이다. 하지만, $k<\nu$ 인 경우 나머지 ν 개의 스테이지에서는 $k=\nu$ 인 경우에 비하여 좀더 많은 trellis 경로들 중에서 선택 가능하다.

그림 4는 4-스테이트 trellis 구조에서 초기 스테이트를 4개중 첫 번째 '00' 과 세 번째 '10' 중에서 선택가능하며, 각 초기 스테이트에 따라 마지막 스테이트는 '00' 에 대하여 '00' 과 '01' 이 선택되며, '10' 에 대하여 '10' 과 '11' 이 선택 가능한 BC-TCQ의 viterbi 알고리즘 탐색 과정을 나타낸 것이다. 그림 4의 $L-\log_2 4$ 스테이지의 스테

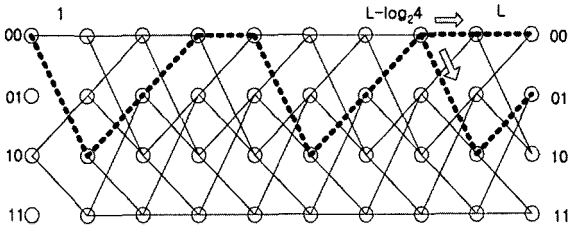

그림 4. BC-TCQ 인코딩 과정
Fig.4. Encoding of the BC-TCQ.

이트 '00'까지 결정된 survivor 경로 (굵은 점선)의 초기 스테이트는 '00' 이다. 따라서 마지막 2개 스테이지에서의 선택 가능한 trellis 경로는 마지막 스테이지의 '00' 스테이트로 향하는 경로와 '01' 스테이트로 향하는 경로 뿐이다.

III. 예측 BC-TCQ를 이용한 멜켑스트럼 계수 양자화기

3.1. 예측 BC-TCQ의 설계 및 최적화

멜켑스트럼 계수 양자화기의 설계는 양자화 성능, 요구 비트, 채널 에러에 대한 강인성, 그리고 구현상 연산 복잡도 등을 상호 고려하여야 한다. 또한 대상 신호의 양자화는 최소의 비트로 최대한 정확히 대상 신호를 표현하도록 인코딩 하는 것이 목적이다. 본 연구에서 적용한 멜켑스트럼 계수 양자화기는 제한된 비트에서 보다 더 좋은 양자화 성능을 얻기 위하여 멜켑스트럼 계수의 프레임간 높은 상관도를 이용한 1차 AR 예측 구조를 적용하였다.

그림 5는 프레임간 상관성을 이용한 1차 AR inter-predictive BC-TCQ 양자화기의 구조도이다. 예측 BC-TCQ를 이용한 멜켑스트럼 계수 양자화 과정을 살펴보면 다음과 같다. 먼저 멜켑스트럼 벡터 $\underline{c}(n)$에서 멜켑스트럼 벡터의 DC 성분을 제거함으로써 AR 예측 필터의 입력 벡터 $\underline{x}(n)$을 구한 후, 구해진 입력벡터에 1차 AR 예측 필터를 식 1과 같이 적용하여 에러벡터 $\underline{e}(n)$를 구한다. 구해진 에러벡터는 BC-TCQ 방식을 이용하여 프레임당 총 36 비트를 할당하여 양자화가 이루어진다. 양자화된 DC 제거 멜켑스트럼 계수 $\underline{\hat{x}}(n)$는 식 2와 같이 양자화된 예측 에러값 $\underline{\hat{e}}(n)$에 예측 성분값 $\rho \cdot \underline{\hat{x}}(n-1)$을 더하여 구한다. 최종적으로 양자화된 멜켑스트럼 계수

표 1. 예측 BC-TCQ에 대한 비트 할당
Table 1. Bit allocation for the predictive BC-TCQ.

파라미터	비 트
패스 정보	12
코드워드 정보	4 (stage 1) 3 x 2 (stage 2 to 3) 2 x 5 (stage 4 to 8) 1 x 4 (stage 9 to 12)
총 계	36

$\underline{\hat{c}}(n)$은 양자화된 DC 제거 멜켑스트럼 계수 $\underline{\hat{x}}(n)$에 멜켑스트럼 벡터의 DC 성분을 더함으로써 구해진다.

$$e_k(n) = x_k(n) - \rho_k \cdot \hat{x}_k(n-1) \quad (1)$$

$$\hat{x}_k(n) = \hat{e}_k(n) + \rho_k \cdot \hat{x}_k(n-1) \quad (2)$$

여기서 ρ는 프레임간 상관도를 나타내고, $x_k(n)$은 P 차원 입력 벡터 $\underline{x}(n)$의 k 번째 요소 값을 나타내며, $\hat{x}_k(n-1)$은 $n-1$번째 양자화된 벡터 $\underline{\hat{x}}(n-1)$의 k 번째 요소 값이다.

3.2. 비트할당

본 연구에서 적용된 예측 BC-TCQ 멜켑스트럼 계수 양자화기는 quadrupled-output 레벨을 가지며, 12차 멜켑스트럼 벡터의 각 요소 별로 8개의 부 코드북 및 16-스테이트 trellis 구조를 갖는다. 프레임당 36비트가 할당 되었을 경우 적용 양자화기의 각 요소 별 비트 할당은 표 1과 같다.

비트 할당 시, 멜켑스트럼 벡터의 각 요소 값이 음성 특성에 미치는 중요도가 각각 달라 중요도가 높은 요소 값에 더 많은 비트를 할당할 필요성이 있다. 따라서 본 연구에서는 멜켑스트럼 벡터의 전체 분산과 각 요소들의 분산 비율을 이용하여 식 3과 같이 각 요소 별 최적 비트 [7]를 구하였다.

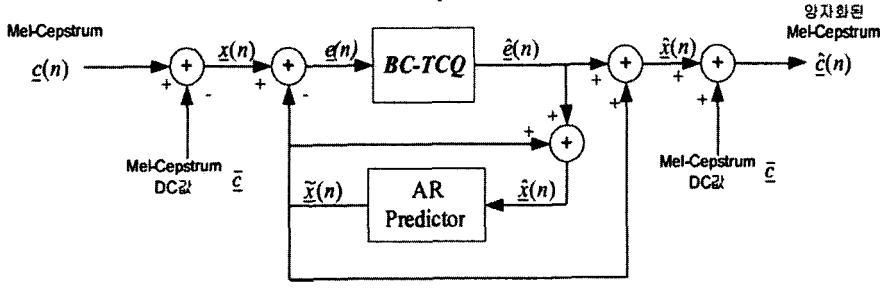

그림 5. 1차 AR inter-predictive BC-TCQ
Fig.5. 1st order AR inter-predictive BC-TCQ.

$$R_k = R + \frac{1}{2}\log_2 \frac{\sigma_k^2}{[\prod_{j=0}^{N-1}\sigma_j^2]^{1/N}} \quad (3)$$

여기서 R_k는 k 번째 요소의 최적 비트이고, R은 전체 벡터에 할당된 비트, 그리고 σ_k는 k 번째 요소의 분산을 나타낸다.

IV. 실험 및 결과

본 연구에서 적용된 예측BC-TCQ를 이용한 멜켑스트럼 계수 양자화기의 성능을 ETSI 규격 분산 음성 인식 표준으로 선택된 멜켑스트럼 계수 양자화기와 비교하였다. 멜켑스트럼 계수 양자화기의 성능 평가를 위해 식 4로 표현되는 cepstral distortion (CD) 값이 사용되었다.

$$CD = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{i=1}^{12}(c(i)-\hat{c}(i))^2\right)^{\frac{1}{2}} \quad (4)$$

여기서 $c(i)$는 원 멜켑스트럼 계수의 i 번째 요소 값이고, $\hat{c}(i)$는 양자화된 멜켑스트럼 계수의 i 번째 요소 값이다. 그리고 N은 전체 프레임 수이다.

제안된 양자화기의 성능을 ETSI 규격 멜켑스트럼 계수 양자화기와 균등한 조건에서 비교하기 위하여 멜켑스트럼 계수 분석 과정을 ETSI규격과 똑 같이 사용하였고, 첫 번째 멜켑스트럼 계수 및 로그 에너지 계수를 제외한 12차 멜켑스트럼 계수에 대해서만 성능 평가를 하였다.

양자화기의 성능 평가를 위해 양자화기가 광대역 (16kHz) 및 협대역 (8kHz) 음성 신호에 대해 동작 할 경우 각각 27분 및 138분 길이의 음성 샘플을 사용하여 예측 BC-TCQ 멜켑스트럼 계수 양자화기를 훈련하였고, 5분 및 7분 길이의 음성 샘플을 사용하여 양자화기 성능 시험을 하였다. 훈련 및 시험용 음성 샘플은 한국어 남자, 여자, 영어 남자, 여자의 음성 순으로 반복 구성되어 있다.

광대역 음성 신호의 12차 멜켑스트럼 계수에 대한 적용 양자화기의 성능은 다음 표 2와 같으며, 프레임당 36비트에서 예측 BC-TCQ는 ETSI 규격 split VQ 방식보다 CD 성능 측면에서 1.6 더 높다. 그리고 프레임당 18비트를 할당하여 예측 BC-TCQ로 양자화를 수행하였을 때 프레임당 36비트를 할당한 split VQ와 성능이 비슷하여 약 18비트의 이득을 얻을 수 있었다.

표 2. 광대역 음성 신호에 대한 ETSI 규격 split VQ 와 예측 BC-TCQ의 CD 성능비교

Table 2. CD comparison of the ETSI standard split-VQ and the predictive BC-TCQ (at 36 bits) for wideband speech.

	예측 BC-TCQ						ETSI 규격 split VQ
비트	18	22	26	30	34	36	36
CD	2.521	2.035	1.615	1.287	1.022	0.915	2.515

표 3. 협대역 음성 신호에 대한 ETSI 규격 split VQ 와 예측 BC-TCQ의 CD 성능비교

Table 3. CD comparison of the ETSI standard split-VQ and the predictive BC-TCQ (at 36 bits) for narrowband speech.

	예측 BC-TCQ						ETSI 규격 split VQ
비트	18	22	26	30	34	36	36
CD	2.567	2.065	1.635	1.293	1.039	0.937	2.325

협대역 음성 신호의 12차 멜켑스트럼 계수에 대한 적용 양자화기의 성능은 다음 표 3과 같으며, 프레임당 36비트에서 예측 BC-TCQ는 ETSI 규격 split VQ 방식보다 CD 성능 측면에서 1.38 더 높다. 그리고 프레임당 18비트를 할당하여 예측 BC-TCQ로 양자화를 수행하였을 때 프레임당 36비트를 할당한 split VQ와 성능이 비슷하여 약 18비트의 이득을 얻을 수 있었다.

다음 표 4에 예측 BC-TCQ와 ETSI 규격의 멜켑스트럼 계수 양자화기와의 계산량을 비교하였다. 표 4와 같이 예측 BC-TCQ 양자화기는 ETSI 규격 보다 가산, 승산, 비교 연산을 더한 전체 계산량이 5.66% 감소하였다. 또한 예측 BC-TCQ는 스테이지별로 스칼라 양자화 코드북만으로 동작하므로 ETSI 규격의 벡터 양자화 코드북에 비하여 작은 메모리를 요구한다. 다음 표 5는 예측 BC-TCQ 양자화기와 ETSI 규격간의 메모리 요구량을 비교하였다.

ETSI 규격의 경우 12차 멜켑스트럼 벡터를 2개 요소를 갖는 6개의 벡터로 분리하여 각 벡터를 6비트로 양자

표 4. 예측 BC-TCQ와 ETSI 규격 split VQ 양자화기의 연산 복잡도 비교

Table 4. Computational complexity of the predictive BC-TCQ and the ETSI standard split-VQ (at 36bits).

		예측 BC-TCQ	ETSI 규격 split VQ
계산량	가산 연산	1304	1920
	승산 연산	432	768
	비교 연산	1048	384
전체 계산량		2784	3074

표 5. 예측 BC-TCQ 와 ETSI 규격 split VQ의 메모리 요구량 비교
Table 5. ROM requirements of the predictive BC-TCQ and the ETSI standard split-VQ (at 36bits).

	예측 BC-TCQ	ETSI 규격 split VQ
ROM(words)	240	384

화 함으로 $2^6 \times 6 = 384$ words가 필요하다. 본 예측 BC-TCQ 양자화기의 경우 하나의 스테이지에 k 비트가 할당됐을 경우 2^{k+2} 개의 스칼라 양자화 출력 값을 가진다. 따라서 첫 번째 스테이지에 4비트, 두 번째 및 세 번째 스테이지에 3비트, 네 번째부터 여덟 번째 스테이지에 각각 2비트, 나머지 스테이지에 1비트를 할당한 예측 BC-TCQ는 메모리 요구량 측면에서 $64 \times 1 + 32 \times 2 + 16 \times 5 + 8 \times 4 = 240$ words 만큼 필요하다. 따라서 전체 메모리 요구량에 있어서 예측 BC-TCQ 멜켑스트럼 양자화기는 ETSI 규격 split VQ에 비하여 37.5% 감소하였다.

V. 결 론

기존의 TCQ 방식은 프레임 단위로 전송되는 데이터의 양자화시 초기 스테이트 정보를 부 정보로 전송해야 한다. Nikneshan과 Khandani에 의해 제시된 TB-TCQ는 각 블록의 첫 번째 스테이지의 스테이트와 마지막 스테이지의 스테이트가 같다는 제약 조건을 도입함으로써 초기 스테이트 정보를 전송할 필요는 없으나, 스테이트 개수 만큼의 viterbi 알고리즘을 수행해야 함으로서 계산량이 크게 요구된다. 본 논문에 서는 전체 스테이트 수보다 적은 초기 스테이트를 정의해서 TB-TCQ보다 적은 viterbi 알고리즘을 수행하여 연산 복잡도 문제를 크게 개선하고, 각 초기 스테이트에 대해 복수개의 최종 스테이트를 정의 해서 최종 스테이트에 대한 유동성을 늘려 성능 감쇄를 없앤 BC-TCQ를 분산 음성 인식 시스템을 위한 멜켑스트럼 계수 양자화기에 적용하였다.

메모리리스 데이터에 좋은 성능을 갖는 BC-TCQ를 분산 음성 인식 시스템을 위한 멜켑스트럼 계수 양자화기에 적용하기 위해, 멜켑스트럼 계수의 프레임간 높은 상관성을 이용한 1차 AR 예측 필터를 멜켑스트럼 계수에 적용함으로써 구해지는 예측 에러를 BC-TCQ로 양자화하는 예측 BC-TCQ를 제안하였다.

본 연구에서 적용된 예측BC-TCQ 양자화기가 광대역 및 협대역의 음성 신호에 대해서 프레임당 36 비트로 동작할 경우 ETSI 규격 split VQ 멜켑스트럼 계수 양자화기와 비교하여 평균 CD값이 1.6 및 1.38 향상되었다. 또한 프레임당 18비트를 할당하였을 경우, ETSI 규격 split VQ와 CD 성능이 비슷하여 총 18비트의 이득을 얻을 수 있었다. 그리고 가산, 승산, 비교 연산을 더한 전체 계산량 측면에서 5.66% 감소 하였으며, 메모리 사용량 측면에서도 5.66% 감소하였다.

감사의 글

본 연구는 2003년도 한국과학재단의 특정기초연구사업 (과제번호: R01-2003-000-11620-0)의 지원으로 수행되었습니다.

참 고 문 헌

1. M. W. Marcellin, and T. R. Fischer, "Trellis coded qu-antization of memoryless and Gauss-Markov sources," IEEE Trans. Communications, **38**, issue 1, 82-93, Jan. 1990.
2. G. D. Forney Jr., "The Viterbi algorithm," Proc. IEEE, **61**, 268-278, Mar. 1973.
3. S. Nikneshan and A. K. Khandani, "Soft Decision Decoding of Fixed Rate Entropy Constrained Quantizer over a Noisy Channel," 20th Biennial Symposium on Communications, 116-118, Kingston, ON, May 28-May 31, 2000.
4. S. Kang, Y. Shin, and T.R. Fischer, "Low-complexity predictive trellis coded quantization of speech line spectral frequencies," IEEE Trans. Signal Processing, **52** (7), 2070~2079, July 2004.
5. Speech Processing, Transmission and Quality aspects (STQ): Distributed speech recognition: Front-end feature extraction algorithm: Compression algorithms, ETSI ES 201 **108** (V1.1.2), April 2000.
6. R. F. Kubichk, "Mel-Cepstral Distance measure for objective speech quality assessment," Communications, Computers and Signal Processing, IEEE Pacific Rim Conf , **1**, 125 - 128, May 1993.
7. N. S. Jayant, "Digital Coding of Waveforms: Principles and Applications to Speech and Video," Prentice Hall Signal Processing Series, Academic Press, 524-532, 1984.

저자 약력

• 강 상 원 (Sangwon Kang)

한국음향학회지 제20권 제4호 참조

• 윤 병 식 (Byungsik Yoon)

1990년 2월: 경북대학교 공과대학 전자공학과 졸업 (학사)
1992년 2월: 경북대학교 대학원 전자공학과 졸업 (석사)
1992년 1월~현재: 한국전자통신연구원 이동통신연구단 선임연구원
2004년 3월~현재: 한양대학교 전자전기제어계측공학과 박사과정
※ 주관심 분야: 이동통신시스템, 음성 신호처리, 멀티미디어 신호처리

• 이 준 석 (Joonseok Lee)

1990년 2월: 경북대학교 공과대학 전자공학과 졸업 (학사)
1992년 2월: 경북대학교 대학원 전자공학과 졸업 (석사)
1992년 1월~현재: 한국전자통신연구원 이동통신연구단 선임연구원
2004년 3월~현재: 한양대학교 전자전기제어계측공학과 박사과정
※ 주관심 분야: 이동통신시스템, 음성 신호처리, 멀티미디어 신호처리