

EVRC의 고정 코드북 탐색 과정에서의 계산량 감소 알고리즘

A Complexity Reduction Algorithm for Fixed Codebook Search in EVRC

정성교, 최용수, 박상욱, 윤대희, 이충용
연세대학교 전자공학과

Sung Kyo Jung, Yong-Soo Choi, Sang-Wook Park, Dae Hee Youn, Chung-Yong Lee
Department of Electronic Engineering, Yonsei University
e-mail : skjung@radar.yonsei.ac.kr

요약문

본 논문에서는 최근 우리 나라 CDMA 디지털 셀룰라 시스템에서 채택한 EVRC(Enhanced Variable Rate Codec)의 고정 코드북 탐색 과정에서의 계산량 감소 알고리즘을 제안한다. 고정 코드북 탐색 과정에서 펄스 위치에 제한을 주어 기존의 방법보다 가능한 펄스 위치의 조합을 줄인다. 또한 ETIR(Extremely Truncated Impulse Response) 방법을 적용하여 고정 코드북 연산량을 개선시킨다.

주/객관적 음성 평가 방법을 수행한 결과, 제안된 방법이 기존의 방법에 비해 음질의 저하가 없음을 확인하였다.

1. 서론

낮은 전송율 부호화 연구의 비약적인 발전을 가져온 CELP(Code Excited Linear Prediction)[1]는 선형 예측 방법을 이용하여 4.8 kbps 이하에서도 우수한 음질을 갖는 분석-합성(analysis-by-synthesis)[2] 부호화 방법이다. EVRC[3]는 기존의 CELP와는 다른 RCELP(Relaxation CELP)[6] 알고리즘에 기반하고 입력 신호에 따라 전송율을 가변하며 CDMA 환경에서 강인한 코덱이다. 기존의 CELP는 원래의 음성 신호에 대하여 분석-합성 방법을 사용한다. 반면에 RCELP는 보간된 피치에 의해 수정된 음성 신호에 대하여 분석-합성 방법을 사용한다[4][5][6]. 선형 보간하여 얻어진 피치 진폭을 적용 코드북 이득을 위한 지연 진폭으로 사용하면 피치에 한달된 미드를 주어 고정 코드북의 이득 신호와 채널 손상 방지에 미드를 더 할당할 수 있다는 장점이 있지만 계산량이 증가한다. EVRC의 고정 코드북 구조는 각장과 계산량에서 장점을 가지고 있는 ACELP(Algebraic CELP) 코드북에 기반하고 있다.

본 논문에서는 EVRC의 계산량을 줄이기 위해 CELP 계열 코더에서 많은 계산량이 요구되는 고정 코드북 탐색 과정의 고속화 방법을 제안한다. 고정 코드북 탐색 과정에서의 계산량 감소를 위해 우선적으로 펄스 위치에 제한을 주어 가능한 펄스 위치의 조합을 줄인다. 또한 ETIR(Extremely Truncated Impulse Response)[7] 방법을 적용하여 고정 코드북 탐색 과정에서의 역방향 필터링과 임펄스 응답의 상관 행렬 연산량을 줄인다.

컴퓨터 모의 실험 후 주/객관적 음성 평가를 수행한 결과 제안된 방법이 적은 계산량으로도 기존의 방법에 비해 음질의 저하가 없음을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 EVRC의 고정 코드북 과정을 간단히 설명하고, 3장에서는 고정 코드북 탐색 과정에서의 계산량을 감소를 위한 방법을 제안한다. 4장에서는 주/객관적 음성 평가를 통해 제안된 방법과 기존의 방법을 비교하고, 마지막으로 5장에서 결론을 맺는다.

2. EVRC의 고정 코드북 탐색

EVRC의 고정 코드북은 대수(algebraic) 코드북으로 그 구조는 ISPP(Interleaved Single Pulse Permutation) 디자인에 기반한다. 최대 전송률인 경우 한 서브프레임은 다섯 트랙으로 나뉘어지며 각 트랙은 11개의 펄스 위치를 포함한다.

코드북 탐색을 서브프레임 단위로 행하며 원신호와 합성 신호 사이의 가중된 평균 차승 오차를 최소화하는 코드북 인자와 코드북 이득을 찾는다. k번째 대수 코드북 벡터를 g_k 라 하면 대수 코드북 탐색은 다음 성분을 최대화하는 코드벡터를 찾는 과정이다.

EVRC의 고정 코드북 탐색 과정에서의 계산량 감소 알고리즘

$$T_k = \frac{C_k}{E_k} = \frac{(d' c_k)^2}{c_k' \Phi c_k} \quad (1)$$

여기서 d 는 역방향 필터링된 신호이며 $\Phi = H' H$ 는 임펄스 응답의 상관 행렬이고 H 는 임펄스 응답으로 이루어진 lower triangular Toeplitz 행렬이다.

최적의 코드북 벡터를 결정하기 위해서는 펄스 위치와 부호의 가능한 모든 조합에 대하여 식 (1)을 계산하여야 한다. 이는 너무나 많은 계산량을 요구하므로 EVRC에서는 코드북 탐색을 간단하게 하기 위해 두 가지 방법을 사용한다. 우선 코드북 탐색 과정을 간단히 하기 위해서 적절한 참조 신호의 부호를 고려하여 펄스의 부호를 미리 정한다. EVRC에서 사용하는 참조신호 $e(i)$ 는 다음과 같이 구한다.

$$e(i) = \sqrt{\frac{\sum_{j=0}^{54} d^2(j)}{\sum_{j=0}^{54} x^2(j)}} x(i) + 2d(i) ; 0 \leq j \leq 54 \quad (2)$$

여기서 $x(i)$ 는 산차 신호이고 $d(i)$ 는 역방향 필터링된 신호이다. i 번째 위치의 펄스의 크기를 $e(i)$ 의 부호와 함께 놓음으로써 펄스의 부호를 미리 정한다.

그리고 펄스의 위치 탐색에서도 펄스 위치의 가능한 모든 조합에 대해서 탐색을 하는 nested loop 탐색 방법을 사용하지 않고 depth-first tree 탐색 방법을 반복적으로 사용하여 한정된 펄스 위치의 조합에 대해 탐색한다. 즉, 여덟 개의 펄스를 네 쌍의 펄스로 분류하고 펄스의 위치는 한 쌍씩 순차적으로 결정한다. 이 과정에서 다섯 개의 트랙에 대해 여덟 개의 펄스를 찾으므로 두 개의 트랙은 하나의 펄스가 존재하는 단일-펄스 트랙(single-pulse track)이 되고 나머지 세 개의 트랙은 두 개의 펄스가 존재하는 이중-펄스 트랙(double-pulse track)이 된다. 단일-펄스 트랙은 (T_3, T_4) , (T_4, T_0) , (T_0, T_1) , (T_1, T_2) 중에서 한 쌍이 선택되고 이 코드워드는 2비트로 부호화된다[3].

합성 신호의 음질을 향상시키기 위하여 하모닉 성분을 강화시키도록 설계된 적응 프리필터를 사용한다. 프리필터는 다음과 같다.

$$F(z) = \frac{1}{1 - g_p z^{-1}} \quad (3)$$

여기서 g_p 는 서브프레임의 평균 피치이며 g_p 는 [0.2, 0.9] 사이의 값으로 제한된 적응 코드북 이득이다. 이 프리필터를 고정 코드북 탐색 과정에서의 코드 벡터와 임펄스 응답에 포함시킬 수 있다.

3. 제안한 고정 코드북 탐색 방법

3.1. 펄스 위치 제한에 의한 펄스 위치 조합의 감소

EVRC의 고정 코드북 탐색 과정은 다섯 개의 트랙에 대해 여덟 개의 펄스를 찾기 때문에 단일-펄스 트랙(single-pulse track)과 이중-펄스 트랙(double-pulse track)의 여러 가지 조합 중에서 네 가지의 경우로 제한된다. 네 가지의 코드워드에 대해 코드북 탐색을 수행하는 경우에 가능한 펄스 위치 조합은 1936이다.

기존의 방법에서는 $e(i)$ 의 절대값의 크기를 이용하여 홀수번째 펄스 위치에 대하여 펄스의 가능한 위치를 11개에서 6개로 줄였다. 이때 가능한 펄스 위치 조합은 1056이 된다. 코드북 탐색 과정에서 계산량 감소를 위해 가능한 펄스 위치 조합을 줄이는 6가지의 방법을 제안하였고 각 방법에 대한 가능한 펄스 위치의 조합을 표 1에 정리하였다.

표 1. 제안한 방법에 대한 펄스 위치 조합.

Method	Complexity		
	Possible Position Combination		Percentage (%)
Source	{(6x11)x4}x4	1056	100.00
Method 1-1	{(5x11)x4}x4	880	83.33
Method 1-2	{(3x11)x4}x4	528	50.00
Method 1-3	{(5x11)x2+(5x5)x2}x4	640	60.61
Method 1-4	{(3x11)x2+(3x3)x2}x4	336	31.82
Method 1-5	{5x11+(5x5)x3}x4	520	49.24
Method 1-6	{3x11+(3x3)x3}x4	240	22.73

3.2. 코드워드 제한에 의한 펄스 위치 조합 감소

앞에서 언급한 바와 같이 코드북 탐색에서 다섯 개의 트랙에 대해 여덟 개의 펄스를 찾기 때문에 단일-펄스 트랙과 이중-펄스 트랙이 존재하게 되는데 그 경우의 수를 EVRC에서는 네 가지의 코드워드로 정하였다. 이 점에서 제안하는 방법들은 네 가지로 정의된 코드워드에 대해서 하나의 코드워드를 제외하여 세 가지의 코드워드로 줄인 방법들이다. 이 경우에 가능한 펄스 위치의 조합은 792이다.

단일-펄스 트랙의 쌍 중에서 Method 2-1은 (T_3, T_4) 를, Method 2-2는 (T_4, T_0) 를, Method 2-3은 (T_0, T_1) 를, 그리고 Method 2-4는 (T_1, T_2) 를 각각 제외시켜 3가지의 코드워드를 사용하였을 때의 방법들이다.

3.3. ETIR(Extremely Truncated Impulse Response) 방법에 기반한 계산량 감소

대수 코드북을 탐색할 때 서브프레임의 길이와 같은 임펄스 응답을 필요로 하게 된다. 이 절에서는 임펄스 응답의 탭 수를 줄이기 위해 ETIR 방법을 적용한다. 펄스 코드벡터에 대한 피치 동기화 과정에서 식 (3)에서와 같은 피치 침예화 필터를 사용한다. 그러므로 이 필터의 영향을 고려하지 않고 임펄스 응답의 탭 수를 줄인다면 음질의 저하가 크게 일어난다. 그러므로 코드북 탐색 과정에 앞서 잔차 신호를 피치 제거 필터 (pitch-removing filter)에 통과시킨다[7].

임펄스 응답의 탭 수를 줄이면 아래의 식 (4)와 같이 구해지는 인지 도메인의 목적 신호 $x_n(n)$ 과 임펄스 응답 $h_{ua}(n)$ 사이의 상호상관(cross-correlation)을 나타내는 역방향 필터링된 신호 $d(n)$ 의 계산량을 줄일 수 있다.

$$d(n) = \sum_{i=n}^{54} x_n(n) h_{ua}(j-n); \quad 0 \leq n \leq 54. \quad (4)$$

또한 임펄스 응답의 상관 행렬 Φ 를 구할 때의 연산량도 감소시킬 수 있다. 대칭 행렬 Φ 의 (i, j) 번째 원소는 다음과 같이 계산된다.

$$\phi(i, j) = \sum_{n=\max(i, j)}^{54} h_{ua}(n-i) h_{ua}(j-n); \quad 0 \leq n \leq 54. \quad (5)$$

임펄스 응답의 탭 수를 n 이라고 하면 역방향 필터링된 벡터 d 와 임펄스 응답의 상관행렬 Φ 를 구할 때 필요한 계산량은 각각 $\frac{n(n+1)}{2}$ [MAC (multiply and accumulation)]이다.

술여진 임펄스 응답의 탭 수에 대해 개선된 MAC 연산량을 표 2에 정리하였다.

표 2. 제안된 방법에 대한 임펄스 응답의 탭 수.

Method	Complexity		
	Tap number of Impulse response	MAC	
C source	55	3080	100.00 (%)
Method 3-1	10	110	3.57 (%)
Method 3-2	17	306	9.91 (%)
Method 3-3	27	756	24.55 (%)
Method 3-4	35	1260	40.91 (%)
Method 3-5	40	1640	53.25 (%)
Method 3-6	45	2070	67.21 (%)

4. 실험 및 결과 고찰

제안된 방법 각각에 대해 주/객관적인 음질 평가 방법을 수행하였다. 객관적인 음질 평가 방법은 SNR (Signal-to-Noise Ratio)[8]과 segSNR (segmental SNR)[8], 그리고 심리음향을 고려한 BSD(Bark Spectral Distance)[9]를 사용하였다. 입력 음성 신호 대신에 선형 피치 컨투어에 의해 수정된 음성 신호를 합성된 신호에 대한 참조 신호로 사용하였다.

주관적인 음질 평가는 기존의 방법에 의해 복원된 신호와 본 논문에서 제안한 방법에 의해 복원된 신호를 듣고 음질의 우열을 가리는 선호도 평가(preference test)를 사용하였다. 15명의 평가자에게 헤드셋을 통해 두 신호를 임의의 순서대로 두 번씩 들려준 다음 둘 중에서 음질이 나은 하나를 선택하게 하였다. 두 음성 신호에 대하여 음질의 차이를 느끼지 못한 경우에 대한 선택도 첨가하였다.

객관적인 음질 평가 방법에서 음성 데이터는 여성 화자와 남성 화자에 대해 각각 2문장을 사용하였고 주관적인 음질 평가 방법에서는 여성 화자와 남성 화자 각각 1문장씩을 사용하였다.

그림 1과 그림 2는 3.1절에서 제안한 각 방법에 대한 주/객관적 음질 평가의 결과를 보여주고 그림 3과 그림 4는 3.2절에서 제안한 방법에 대한 주/객관적 음질 평가의 결과를 보여준다. 3.3절에서 제안한 방법에 대한 주/객관적 음질 평가의 결과는 그림 5와 그림 6에 나타내었다.

펄스 위치에 제한을 두어 가능한 펄스 위치의 조합을 기존의 방법에 비해 반으로 줄인 경우에 SNR은 여성 화자의 음성에 대해 약 0.11 dB정도가 감소하였고 남성 화자의 음성에 대해서는 약 0.14 dB가 감소하였다. 선호도 평가에서도 여성 화자와 남성 화자의 음성에 대해서는 각각 86.67%와 66.67%정도가 원래의 방법으로 합성된 신호와 차이를 느끼지 못하였다.

ETIR 방법을 적용하여 20개의 탭을 줄여 35탭의 임펄스 응답을 사용하면 약 59%의 계산량을 감소시킬 수 있다. 그 때의 경우 여성 화자의 음성에 대해서는 약 0.54 dB가 감소하였지만 남성 화자의 음성에 대해서는 약 0.19 dB가 감소하였다. 선호도 평가의 결과를 보면 여성 화자의 음성에 대해서는 53.33%가 차이를 느끼지 못하였고 남성 화자의 신호에 대해서는 73.33%가 차이를 느끼지 못한 것으로 나타났다.

5. 결론

본 논문에서는 EVRC의 고정 코드북 탐색 과정을 고속화 할 수 있는 몇 가지 방법을 제안하고 그에 대한 주/객관적 음질 평가 방법으로 성능을 평가하였다.

펄스 위치에 제한을 두어 가능한 펄스 위치의 조합을 기존의 방법에 비해 반으로 줄인 경우, 주/객관적 음질 평가 결과 음질의 차이가 없음을 확인하였다. 그리고 코드북에 제한을 두는 것보다 펄스 위치에 제한을 두는 방법이 주/객관적 평가 방법 모두 우수하게 나타났다. 또한 ETIR 방법을 적용하여 35개 탭의 임펄스

EVRC의 고정 코드북 탐색 과정에서의 계산량 감소 알고리즘

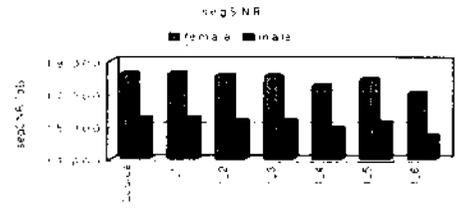
응답을 사용한 경우에는 약 59%의 계산량을 감소하였지만 여성 화자의 음성이 남성 화자의 음성에 비해 왜곡이 크다는 것을 알 수 있다. 그러나 줄어든 임펄스 응답의 영향으로 인한 왜곡이 어느 정도 발생하지만 음질의 저하가 크지 않음을 주/객관적 음질 평가를 통해 확인할 수 있다.

선반적인 결과를 살펴보면 가능한 펄스 위치의 조합이나 임펄스 응답의 탭 수를 감소시켰을 때 남성 화자의 음성보다 여성 화자의 음성이 더 민감하다는 것을 알 수 있다. 그러므로 고속 알고리즘 연구를 할 때 여성 화자의 음성에 대한 고려가 충분히 있어야 한다.

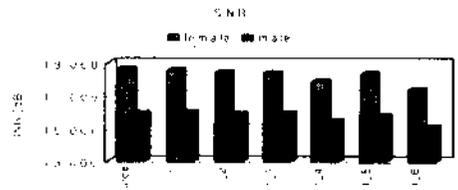
또한 실시간 처리를 위해서는 고정 코드북 탐색 과정뿐만 아니라 다른 부분에 대한 고속 알고리즘 연구가 필요하다.

참 고 문 헌

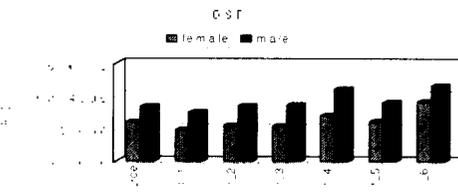
- [1] M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP) High-Quality Speech at Very Low Bit Rates," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, 1985, pp. 937-940.
- [2] C. G. Bell, H. Fujisaki, J. M. Heinz, K. N. Stevens and A. S. House, "Reduction of Speech Spectra by Analysis-by-Synthesis Techniques," *J. Acoust. Soc. Am.* 1961, pp. 1725-1736.
- [3] QUALCOMM Inc., Proposed TIA/EIA/3GPP2 Standard - *Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems*, Official Ballot Version, April 19, 1996.
- [4] W. B. Kleijn, R. P. Ramachandran and P. Kroon, "Generalized Analysis-by-Synthesis Coding and its Application to Pitch Prediction," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, 1992, pp. 1337- 1340.
- [5] W. B. Kleijn, R. P. Ramachandran and P. Kroon, "Interpolation of the Pitch-Predictor Parameters in Analysis-by-Synthesis Speech Coders," *IEEE Trans. on Speech and Audio Processing*, Vol. 2, No. 1, Part 1, January 1994, pp. 12-51.
- [6] W. B. Kleijn, R. P. Ramachandran and P. Kroon, "The RCELP Speech Coding Algorithm," *European Transactions on Telecommunications*, Vol. 5, No. 5, 1994, pp. 573-582.
- [7] H. Ohmura and K. Mano, "A Low Complexity 1.6 kbit/s Speech Coder Based on Extremely Truncated Impulse Response", *Proc. of IEEE Workshop on Speech Coding*, 1997, pp.63-64.
- [8] Sadaaki Furui and M. Mohan Sondhi, *Advance in Speech Signal Processing*, Dekker.
- [9] Shihua Wang, et al, "An Objective Measures for Predicting Subjective Quality of Speech", *IEEE J. Select Areas Commun.*, vol. 10, No.5, pp.819-829, June 1992.



(a) segSNR.

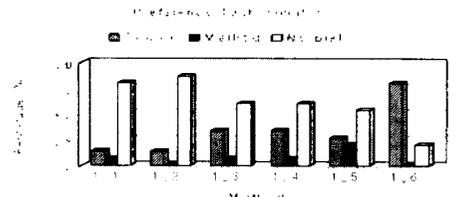


(b) SNR.

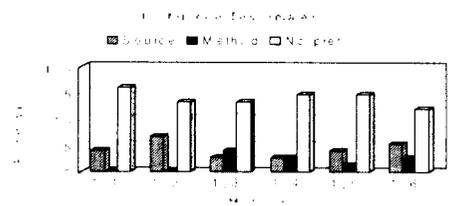


(c) BSD.

그림 1. 펄스 위치 조합의 감소에 따른 객관적 음질 평가



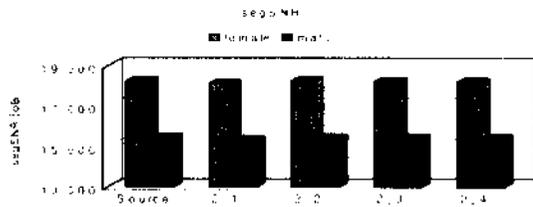
(a) 여성 화자 음성.



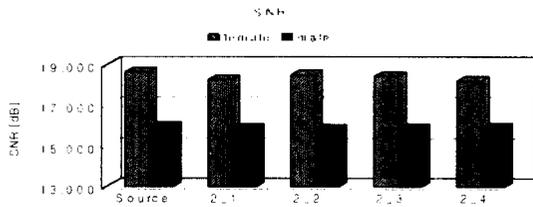
(b) 남성 화자 음성.

그림 2. 펄스 위치 감소에 따른 주관적 음질 평가.

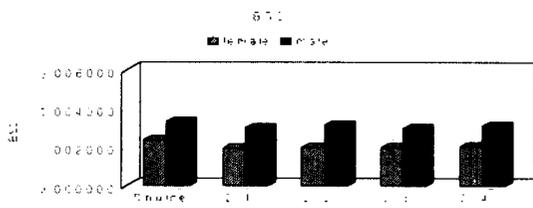
제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)



(a) segSNR.

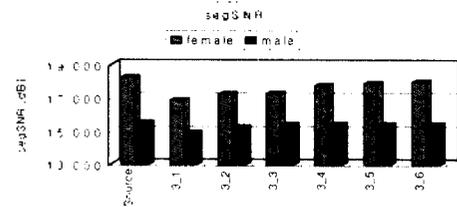


(b) SNR.

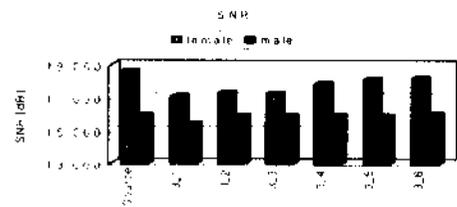


(c) BSD.

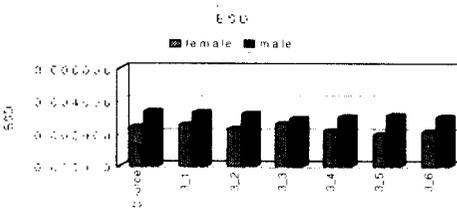
그림 3. 코드워드 감소에 따른 객관적 음질 평가.



(a) segSNR.

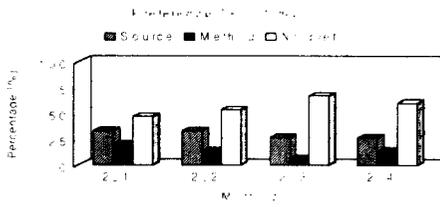


(b) SNR.

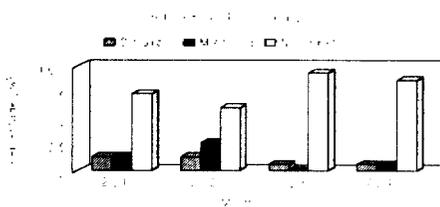


(c) BSD.

그림 5. 탭 수의 감소에 따른 객관적 음질 평가.

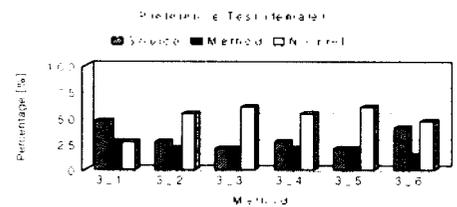


(a) 여성 화자 음성.

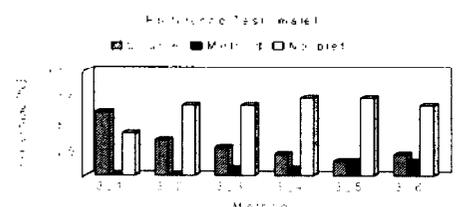


(b) 남성 화자 음성.

그림 4. 코드워드 감소에 따른 주관적 음질 평가.



(a) 여성 화자 음성.



(b) 남성 화자 음성.

그림 6. 임펄스 응답의 탭 수 감소에 따른 주관적 음질 평가.