

논문 2004-41SP-4-15

# AMR과 G.723.1 음성부호화기를 위한 효율적인 상호부호화 알고리즘 및 성능평가

(An efficient transcoding algorithm for AMR and G.723.1 speech coders and performance evaluation)

최진규\*, 윤성완\*\*, 강홍구\*\*, 윤대희\*\*

(Jin-Kyu Choi, Sung-Wan Yoon, Hong-Goo Kang, and Dae-Hee Youn)

## 요약

무선망과 VoIP 같은 서로 다른 음성 통신 네트워크간의 통신을 할 경우, 서로 다른 구조를 갖는 두 음성부호화기간의 효율적인 연동이 필요하다. 이런 경우, 가장 간단한 방법으로 두 음성부호화기의 복호화기와 부호화기를 직렬로 연결시키는 tandem방식을 사용할 수 있다. 하지만, tandem방식은 긴 지연시간과 많은 연산량, 그리고 음질저하의 문제점들을 갖는데, 이는 상호부호화 방법을 통해서 해결할 수 있다. 상호부호화 알고리즘은 송신단과 수신단의 음성 부호화기의 구조에 의해 결정되고, 본 논문에서는 연산량을 감소시키고, 음질은 향상 시킬 수 있는 LSP 변환, 개선된 고속 피치 검색, 상호부호화기를 위한 새로운 지각가중 필터 알고리즘을 제안한다. 제안된 알고리즘은 AMR과 G.723.1간의 상호부호화에 적용하였다. 제안된 상호부호화 알고리즘을 사용함으로써 tandem 방식에 비하여 연산량은 약 20%-58% 감소되는 반면, 음질은 향상된다.

## Abstract

In the application requiring the interoperability of different networks such as VoIP and wireless communication system, two speech codecs must work together with the structure of cascaded connection, tandem. Tandem has several problems such as long delay, high complexity and quality degradation due to twice complete encoding/decoding process. Transcoding is one of the best solutions to solve these problems. Transcoding algorithm is varied with the structure of source and target coder. In this paper, transcoding algorithm including the LSP conversion, the pitch estimation and new perceptual weighting filter for reducing complexity and improving quality is proposed. These algorithms are applied to the pair of AMR and G.723.1. By employing the proposed algorithms in the transcoder, the complexity is reduced by about 20%-58% and quality is improved compared to tandem.

**Keywords:** 음성부호화, 상호부호화, AMR, G.723.1, 상호운용

## I. 서론

통신기술이 발달함에 따라, 다양한 형태의 통신서비스에 대한 요구가 증대되고 있다. 현재 음성을 전달 또

는 저장하는 통신서비스에서는 응용분야에 따라 각각의 특성에 적합한 음성부호화기를 표준으로 사용하고 있다. 이러한 이유로 인하여 서로 다른 네트워크간에 음성통신을 할 경우, 완벽한 통신을 위해 송/수신단에선 두 음성부호화기를 모두 지원할 수 있어야 한다. 이렇게 두 음성부호화기의 연동이 필요한 경우, 가장 간단한 방법은 송신단의 복호화기와 수신단의 부호화기를 직렬 연결시키는 tandem 방식이다<sup>[3-5]</sup>. 그러나 tandem 방식은 두 번의 부호화 과정과 복호화 과정을 거치기 때문에 음질이 많이 저하되고, 긴 지연시간과 많은 연산량을 필요로 하는 단점들을 갖고 있다. Tandem 방식

\* 정희원, LG전자, 단말연구소  
(MOBILE HANDSET R&D CENTER, LG Electronics)

\*\* 정희원, 연세대학교 전기전자공학과  
(Department of Electrical and Electronic Engineering, Yonsei University)

※ 이 연구는 한국전자통신연구원(ETRI)의 지원에 의해 이루어졌습니다.

접수일자: 2004년2월16일, 수정완료일: 2004년5월27일

과는 달리 상호부호화 알고리즘은 최대한 송신단의 비트열을 수신단의 비트열로 직접 변환하는 방식을 사용하기 때문에 이러한 문제를 해결할 수 있다<sup>[3-5]</sup>.

본 논문에서는 G.723.1과 AMR 음성부호화기를 위한 상호부호화 알고리즘을 제안한다. 일반적으로 G.723.1은 VoIP에서, AMR은 3세대 이동통신에서 널리 사용되는 음성부호화기이다<sup>[1][2]</sup>. 두 음성부호화기의 특징을 살펴보면, G.723.1은 30ms, AMR은 20ms의 프레임 길이를 가지게 되기 때문에, 제안된 상호부호화 알고리즘은 G.723.1 3프레임의 비트열과 AMR 2프레임의 비트열 간의 변환을 하게 된다. 제안된 알고리즘은 LSP 변환 과정, 개선된 고속 피치 검색과정, 상호부호화기를 위한 새로운 지각가중 필터로, 크게 3부분으로 구성된다. LSP 변환 과정은 송신단 LSP 파라미터를 기하학적 거리와 look-ahead delay를 고려하여 수신단의 LSP 파라미터로 변환해준다. 개선된 고속 피치 검색과정은 개회로 피치를 적은 연산량으로 음질의 저하없이 검색한다. 마지막으로 상호부호화기를 위한 지각 가중필터를 사용함으로써 연산량을 감소시키고, 음질은 향상시킨다. 위의 3가지 과정을 통하여 제안된 상호부호화 알고리즘은 송신단의 비트열을 적은 음질 왜곡과 적은 연산량으로 수신단의 비트열로 변환한다.

본 논문은 총 V장으로 구성되고, II장에서는 음성부호화 표준안인 AMR과 G.723.1에 대하여 설명하고, III장에서는 제안된 알고리즘에 대하여 상세하게 설명한다. 제안된 알고리즘의 성능평가는 IV장에서 설명하고, 마지막으로 V장에서 결론을 짓는다.

## II. ITU-T G.723.1과 ETSI AMR 음성 부호화기

ITU-T G.723.1은 멀티미디어 통신을 위한 음성부호화 표준안으로서, 5.3 kbps와 6.3 kbps의 전송률을 지원한다<sup>[1]</sup>. G.723.1은 30ms의 음성신호를 받아들여 부호화하고 복호화 과정을 거쳐 같은 길이의 음성신호를 복호화한다. 또한 10차 LPC 분석을 위한 7.5ms의 look ahead 지연시간을 갖고 있고, 총 알고리즘 지연시간은 37.5ms이다. 적응 코드북 검색에선 5차 예측기가 사용되고, 고정 코드북으로는 전송률에 따라 ACELP (Algebraic Code Excited Linear Prediction) 또는 MP-MLQ(Multipulse Maximization Likelihood Quantization)을 사용한다.

ETSI AMR은 3GPP WCDMA(Wideband Code Division Multiple Access)를 위한 음성부호화기이고,

표 1. G.723.1과 AMR 음성부호화기의 특징

Table 1. Characteristics of G.723.1 and AMR.

	G.723.1	AMR
Frame size	30ms/frame	20ms/frame
LPC	10th order	10th order
ACB	5-tap predictor, integer resolution	1-tap predictor, 1/6 or 1/3 resolution
FCB	ACELP/MP-MLQ	ACELP

4.75kbps에서 12.2kbps까지 8개의 전송률을 지원한다<sup>[2]</sup>. AMR은 20ms의 프레임 길이를 가지고, 각 프레임은 4개의 부프레임으로 나누어진다. 또한, LPC분석을 위한 5ms의 look ahead 지연시간을 가지고, 총 알고리즘 지연시간은 25ms이다. 표 1은 두 음성부호화기의 주요 특징들을 보여준다.

상호부호화 알고리즘에서 송신단과 수신단의 음성부호화기간의 공통점과 차이점은 중요한 사항이다<sup>[5]</sup>. 예를 들면 G.723.1은 30ms의 프레임 길이를 가지고, 반면에 AMR은 20ms의 프레임 크기를 갖는 차이로 인하여 상호부호화기의 알고리즘은 G.723.1의 두 프레임의 60ms 비트열을 전송받아, AMR의 세 프레임, 즉 60ms 비트열로 변환하여야 한다. 두 음성부호화기의 LPC의 경우는 모두 10차 선형예측 필터를 사용하기 때문에 프레임 길이를 고려하여 변환가능하다. 하지만, 여기신호의 변환의 경우 두 음성부호화기간의 피치의 분해능의 차이와 프레임의 불일치로 인하여 직접 변환 방식을 사용할 수 없고 추가적인 프로세싱이 필요하게 된다. 이렇듯 두 음성부호화기간의 공통점과 차이점은 상호부호화 알고리즘을 결정하는 중요한 요인이 된다.

## III. 제안된 상호부호화 알고리즘

제안된 G.723.1과 AMR간의 상호부호화 알고리즘은 크게 LSP 변환과정, 개선된 개회로 피치 변환과정, 고속 적응 코드북 검색과정, 상호부호화기를 위한 새로운 필터로 구성된다. 제안된 상호부호화기는 두 부호화기의 프레임 길이를 고려하여 30ms 프레임길이를 갖는 G.723.1의 2개의 프레임과 20ms 프레임길이를 갖는 AMR 프레임 3개 간의 변환이 이루어진다. 그림 1은 제안된 상호부호화기의 구조를 보여준다.

제안된 상호부호화 구조를 간략히 설명하면, 우선 G.723.1의 두 프레임에 해당하는 LSP를 선형 보간을 이용하여 AMR의 세 프레임에 해당하는 LSP로 변환한다.

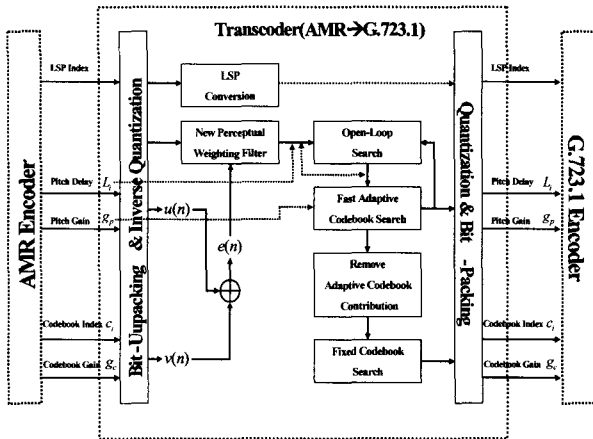


그림 1. 제안된 상호부호화기의 구조  
Fig. 1. Structure of the proposed transcoder.

표 2. LSP 변환의 객관적 음질(PESQ)  
Table 2. Objective quality(PESQ) of LSP conversion.

AMR(Kbps)		4.75	5.15	5.9	6.7	7.4	7.95	10.2	12.2
G(5.3) → A	Conventional	3.047	3.096	3.168	3.241	3.287	3.309	3.390	3.420
	Proposed	3.073	3.125	3.184	3.255	3.301	3.323	3.399	3.424
G(6.3) → A	Conventional	3.104	3.150	3.214	3.299	3.358	3.375	3.490	3.538
	Proposed	3.118	3.161	3.240	3.314	3.376	3.400	3.499	3.539
A → G(5.3)	Conventional	3.133	3.161	3.215	3.252	3.281	3.315	3.339	3.365
	Proposed	3.149	3.169	3.242	3.283	3.301	3.315	3.355	3.362
A → G(6.3)	Conventional	3.196	3.239	3.310	3.359	3.399	3.421	3.468	3.485
	Proposed	3.227	3.255	3.323	3.374	3.414	3.433	3.483	3.492

변환된 LSP는 AMR의 부호화 과정을 거쳐 양자화되어 복호화단으로 전송되고, 다시 LPC 계수로 변환되어 각 부프레임에 해당하는 지각 가중 합성 필터를 구성한다. 그리고 G.723.1의 적응 코드북과 고정 코드북 성분으로 여기 신호를 만든 후 상호부호화기를 위한 제안된 지각 가중 합성 필터에 통과시키면 AMR의 개회로 피치 추정을 위한 목적 신호가 만들어진다.

G.723.1에서 AMR로의 상호부호화의 개회로 피치 추정과정에 송신단의 피치값을 이용하는 개선된 빠른 피치검색(modified fast pitch search) 방법을 적용하여 적은 계산량으로 이중 부/복호화 방법과 동등한 성능으로 개회로 피치 추정과정을 구현하였다. 개회로 피치를 결한 후, 적응 코드북과 고정 코드북 검색을 통해 각각의 파라미터를 추출하게 된다. 변환된 LSP와 적응 코드북과 고정 코드북 파라미터를 부호화하여 비트열을 생성하면 G.723.1에서 AMR로의 상호부호화가 완료된다. 반대 방향인 AMR에서 G.723.1로의 상호부호화 알고리즘은 G.723.1의 적응 코드북 검색과정의 특성을 이용하여 음질의 저하 없이 연산량을 줄이는 고속 적응 코드북 검색(fast adaptive codebook search)방법<sup>[3]</sup>이 추가되고 나머지 과정은 G.723.1에서 AMR로의 상호부호화기의

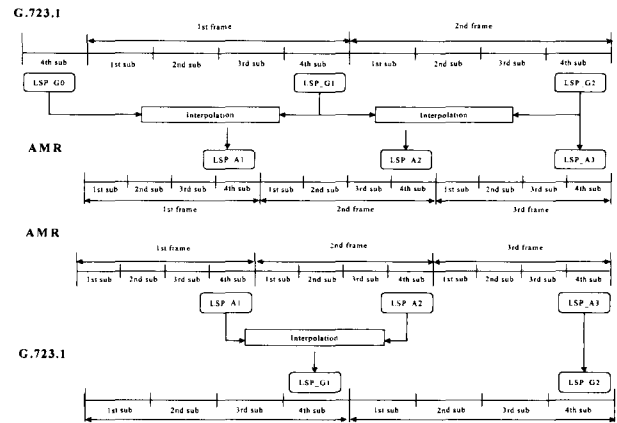


그림 2. 선형 보간법을 이용한 LSP 변환  
Fig. 2. LSP conversion using linear interpolation.

경우와 동일하다.

### 1. 개선된 LSP 변환

제안된 LSP 변환과정에선 G.723.1 프레임의 LSP 정보를 그림 2와 같은 선형보간법을 통해 AMR 3개의 프레임간의 해당하는 LSP 값을 구해준다. 송신단 부프레임의 LSP는 각각 기하학적 거리를 고려하여 수신단 부프레임의 LSP 정보로 사용된다. 이로 인해 LPC 계수를 구하기 위한 자기상관계수, Durbin Recursion과 LPC를 LSP로 변환하기 위해 필요한 5차 다항식의 근을 구하는 과정이 필요 없게 되어 이중 부/복호화를 하는 경우에 비해 계산량이 크게 줄어든다. 또한, 방향에 따라서 G.723.1과 AMR의 LPC 계산에 필요한 7.5ms 또는 5ms의 예견구간으로 인한 지연이 생기지 않기 때문에 이중 부/복호화에서 일어나는 지연시간 증가 문제를 해결할 수 있다<sup>[3-5]</sup>.

일반적인 선형보간 방법은 선형보간에 사용되는 가중치 값을 결정하는데 있어서 기하학적 거리만을 고려한다<sup>[3-4]</sup>. 하지만, 제안된 선형보간 방법을 가중치 값을 결정하는데 있어서, 기하학적 거리뿐만 아니라, 송신단의 예견구간으로 인한 지연시간을 고려하여 가중치 값

을 결정한다. 즉, 선형보간에 사용되는 상수인 가중치 값을 기하학적 거리와 예견구간을 고려하여 결정함으로써, 실질적으로는 예견구간으로 인한 지연시간은 없애는 장점을 가지고, smoothing effect는 존재하게 된다. 이러한 smoothing effect로 인하여 음성의 음질을 향상시킬 수 있다.

표 2는 제안된 선형 보간 방법과 기존의 선형보간 방법간의 객관적 음질을 비교하여 보여준다. 여기서, LPC 이외의 다른 파라미터는 tandem 방식을 사용한다. 표 2에서 보여 주듯이, 제안된 선형 보간 방법을 사용함으로써 객관적 음질을 향상시킬 수 있다.

## 2. 개선된 고속 피치 검색

상호부호화 알고리즘에서는 수신단에 해당하는 피치 변환과정을 통하여 해당하는 개회로 피치를 음질의 저하 없이 적은 연산량으로 검색하게 된다<sup>[3-5]</sup>. 기존 방법으로는 피치 완만화 방법(pitch smoothing)이 있다<sup>[3]</sup>. 피치 완만화 방법은 음질 저하는 거의 없이 적은 연산량으로 개회로 피치를 검색할 수 있는 방법이다. 하지만, 피치 완만화 방법의 연산량은 송신단의 피치와 이전 프레임에 해당하는 수신단의 피치 간격 차이에 의존하게 된다. 하지만, 많은 실험을 통해 확인해본 결과, 일반적으로 피치의 중요성이 상대적으로 적은 무성음구간에서 그 차이가 큰 것을 알 수 있었다. 즉, 음성 부호화 과정에서 피치가 음질의 영향을 거의 끼치지 않은 무성음 구간에서 많은 연산량을 필요로 한다는 단점을 가지게 된다<sup>[6]</sup>.

상호부호화에서의 개회로 피치 검색의 목적 신호는 송신단에서 전송된 파라미터들로 복원된 신호이다. 즉, 송신단에서 전송된 폐회로 피치를 그 주기로 갖는 신호이다. 만약, 송신단의 부호화기와 수신단의 부호화기가 같은 프레임 크기를 갖는다면, 송신단의 폐회로 피치를 수신단의 개회로 피치로 아무런 변환 과정 없이 사용할 수 있다. 하지만, 두 음성부호화기의 프레임 길이가 다르기 때문에 송신단의 폐회로 피치를 수신단의 개회로 피치로 사용하기 위해서는 이러한 프레임간의 불일치를 보상해 주는 과정이 필요로 한다. 제안된 고속 피치 검색 방법(modified fast pitch search method)은 다음과 같은 식 (1)로 표현된다.

$$C_{oz}(j) = \frac{\sum_{n=0}^N S_w(n) * S_w(n-j)^2}{\sum_{n=0}^N S_w(n-j) * S_w(n-j)}, P_{min} \leq j \leq P_{max} \quad (1)$$

여기서,  $S_w$ ,  $N$ ,  $P_{min}$ 와  $P_{max}$ 는 각각 지각 가중된 신호, 부프레임 사이즈, 검색 범위를 나타낸다. 즉, 제안된 빠른 개회로 피치 검색 방법은  $C_{oz}$ 을 최대화하는 인덱스  $j$ 를 찾게 되고, 이때의  $j$ 를 개회로 피치로 선택하게 된다. 제안된 방법의 연산량은  $P_{min}$ 와  $P_{max}$ 로 표현된 검색 범위에 의하여 결정되고, 이 때의 검색범위는 각각에 해당하는 송신단과 수신단의 음성부호화기의 특성을 고려하여 결정하였다.

G.723.1은 피치를 검색하는 과정에서 AMR의 피치 검색 과정보다 큰 윈도우를 사용한다. 즉, 더욱 많은 샘플을 사용하여 피치를 결정하기 때문에 G.723.1의 피치가 AMR의 피치보다 신뢰성이 높다<sup>[8]</sup>. 이러한 특징을 이용하여 제안된 피치 검색 방법의 검색 범위를 결정하였다. 그림 3은 각 방향의 상호부호화기의 피치 검색 범위에 따른 객관적 음질의 변화를 보여준다. 그림 3-(a)에서 보여주듯이, G.723.1에서 AMR로의 상호부호화기의 경우 G.723.1의 피치를 AMR의 피치로 아무런 변환 과정 없이 사용하였을 경우, 즉 직접 변환의 경우에는 많은 음질의 저하가 발생함을 알 수 있다. 하지만, G.723.1의 피치의 신뢰성이 높기 때문에 3샘플 검색 즉, 송신단의 피치 주위의 1샘플씩을 검색하는 것에 비하여, 검색 범위를 늘려도 음질의 향상은 거의 없음을 알 수 있다. 그림 3-(b)는 반대 방향, AMR에서 G.723.1로의 상호부호화기의 피치 검색 범위에 따른 객관적 음질의 변화를 보여준다. 이 경우에는 상대적으로 신뢰도가 낮은 AMR의 피치를 사용하기 때문에 검색 범위가 커질수록 객관적 음질은 향상됨을 알 수 있다. 하지만, 검색 범위에 따른 객관적 음질의 향상은 7샘플 이상에서는 거의 없음을 알 수 있다. 이러한 실험을 토대로 객관적 음질과 연산량을 고려하여, G.723.1과 AMR간의 상호부호화 알고리즘의 제안된 피치 검색 방법의 검색 범위를 식 (2)와 같이 결정하였다.

$$\begin{cases} P_{min} = P_G - 1, P_{max} = P_G + 1, case : G.723.1 \rightarrow AMR \\ P_{min} = P_G - 3, P_{max} = P_G + 3, case : AMR \rightarrow G.723.1 \end{cases} \quad (2)$$

그림 4는 피치 검색 방법에 따른 피치 컨투어(pitch contour)를 보여준다. 여기서, "Full", "Pitch smoothing", "Proposed"은 각각 전체 검색 범위를 다 검색하는 방법, 피치 완만화 방법, 그리고 개선된 고속 피치 검색 방법(modified fast pitch search method)을 의미한다. 그림에서 볼 수 있듯이, 제안된 피치 검색 방법은 기존의 피치 완만화 방법을 사용했을 경우보다 더욱 정확한 피치를 찾을 수 있다.

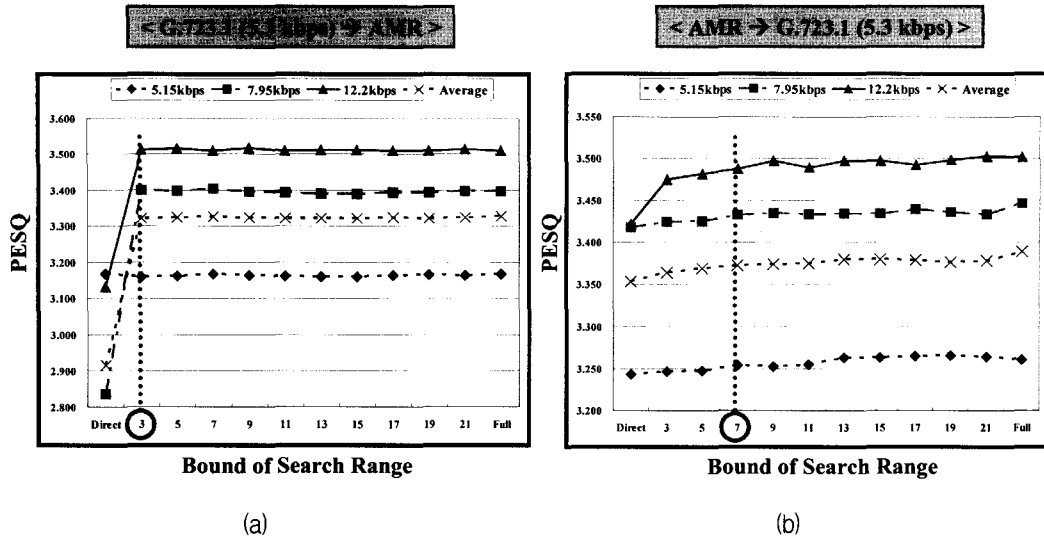


그림 3. 피치 검색 범위에 따른 객관적 음질  
 Fig. 3. Objective quality related to pitch search range.

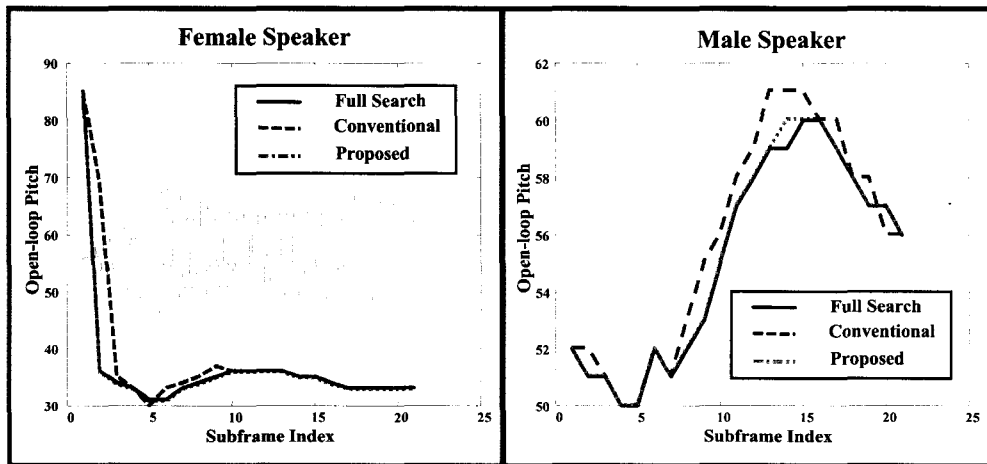


그림 4. 피치 검색 방법 따른 객관적 음질  
 Fig. 4. Objective quality related to pitch search method.

표 3. 피치 검색 방법에 따른 연산량(WMOPS) 비교(AMR→G.723.1)  
 Table 3. Comparison of WMOPS related to pitch search method(AMR→G.723.1).

		G.723.1	5.3 kbps	6.3 kbps
A→G	Full search		1.402	1.545
	Pitch smoothing		0.761	0.761
	proposed		0.066	0.066

표 4. 피치 검색 방법에 따른 연산량(WMOPS) 비교(G.723.1→AMR)  
 Table 4. Comparison of WMOPS related to pitch search method(G.723.1→AMR).

AMR(Kbps)		4.75	5.15	5.9	6.7	7.4	7.95	10.2	12.2
G→A	Full search	1.133	1.133	1.231	1.231	1.231	1.231	1.439	1.251
	Pitch smoothing	0.349	0.355	0.346	0.346	0.346	0.349	1.439	0.349
	Proposed	0.093	0.093	0.120	0.120	0.120	0.120	0.120	0.120

표 5. 피치 검색 방법에 따른 객관적 음질(PESQ) 비교

Table 5. Comparison of objective WMOPS related to pitch search method(AMR→G.723.1).

AMR(Kbps)		4.75	5.15	5.9	6.7	7.4	7.95	10.2	12.2
G(5.3)→A	Full search	3.086	3.120	3.186	3.256	3.306	3.327	3.397	3.431
	Pitch smoothing	3.054	3.102	3.162	3.241	3.300	3.311	3.397	3.432
	Proposed	3.073	3.125	3.184	3.255	3.301	3.323	3.399	3.424
G(6.3)→A	Full search	3.128	3.169	3.251	3.324	3.384	3.398	3.504	3.548
	Pitch smoothing	3.107	3.149	3.233	3.313	3.380	3.389	3.504	3.538
	Proposed	3.118	3.161	3.240	3.314	3.376	3.400	3.499	3.539
A→G(5.3)	Full search	3.144	3.176	3.238	3.281	3.308	3.324	3.363	3.387
	Pitch smoothing	3.107	3.145	3.206	3.258	3.277	3.307	3.343	3.354
	Proposed	3.133	3.165	3.218	3.258	3.277	3.301	3.334	3.350
A→G(6.3)	Full search	3.212	3.251	3.318	3.371	3.410	3.437	3.478	3.507
	Pitch smoothing	3.193	3.233	3.292	3.350	3.397	3.419	3.463	3.480
	Proposed	3.200	3.240	3.303	3.364	3.386	3.427	3.458	3.492

표 3과 4는 피치 검색방법에 따른 연산량을 비교하여 보여준다. 표에서 볼 수 있듯이 제안된 방법을 사용함으로써 기존의 방법인 피치 완만화 방법보다 연산량을 감소시킬 수 있고, 전체 범위를 검색하는 방법에 비하여 92~96% 이상 연산량을 감소시킬 수 있는 장점을 가짐을 알 수 있다. 표 5는 피치 검색 방법에 따른 객관적 음질의 변화를 보여준다. 제안된 방법을 사용함으로써 기존의 피치 완만화 방법보다 음질이 향상됨을 알 수 있다. 또한, 연산량이 가장 많은 전체 검색 범위를 검색하는 방법에 비하여 음질의 저하는 거의 없음을 알 수 있다.

3. 상호부호화기를 위한 새로운 지각가중 필터

인간의 귀는 주파수 영역에서 주파수 크기응답이 큰 포먼트 주변의 잡음에 대해서는 민감하지 않고, 반대의 경우인 널 주변의 잡음에는 민감한 특성을 갖는다. 이러한 청각적 특성을 이용하여 일반적인 음성부호화기에서는 지각가중 필터를 사용하여 주파수 영역의 포먼트 부분을 약화시키는 역할을 함으로써 평균 오차를 최소화하는 음성부호화 과정의 잡음을 포먼트 주변에 상대적으로 집중시키는 역할을 한다. 복호화 과정에서는 후처리 필터를 사용하여 지각가중 필터에 의해서 약화되었던 포먼트 주변을 다시 강화시키는 역할을 수행한다.

상호부호화기의 지각가중 필터와 후처리 필터는 직렬 연결된 구조를 가지고, 주파수 영역에서 서로 상반되는 역할을 한다. 이러한 필터과정에서 두 필터의 비선형 위상특성으로 인한 음성 왜곡이 발생하게 되고, 또한 많은 연산량을 필요로 하는 단점을 가진다. 특히, 복호화기의 후처리 필터는 상대적으로 많은 연산량을 필요로 한다<sup>[12]</sup>. 본 논문에선, 이러한 문제를 해결하기 위해서 후처리

필터와 지각가중 필터의 역할을 대신 할 수 있는 하나의 새로운 필터를 설계한다. 일반적으로 적용 후처리 필터는 tilt 보상필터를 포함하지만, 객관적 음질 평가와 주관적 음질 평가를 통하여 tilt 보상필터는 음질에 거의 영향을 미치지 않는 것을 확인하였다. 이는 상호부호화기에서는 최종 수신단에 같은 역할을 하는 후처리 필터가 존재하기 때문이다. 따라서 후처리 필터와 지각가중 필터의 직렬 연결된 필터는 아래와 같은 식 (3)으로 나타낼 수 있다.

$$H_R(z) = H_f(z) W(z) = \frac{\hat{A}(z/r_n) * \hat{A}(z/r_1)}{\hat{A}(z/r_d) \hat{A}(z/r_2)} \quad (3)$$

여기서,  $r_n, r_d, r_1$ 과  $r_2$ 는 각각 후처리 필터와 지각가중 필터의 가중치 값을 나타내고,  $H_R(z)$ 는 후처리 필터와 지각가중 필터의 직렬 연결된 필터를 나타내며, 이후부터는 기준 필터라고 명하기로 한다. 상호부호화기에서는 복호화단의 LSP를 직접 또는 선형 보간하여 부호화단에서 사용한다. 따라서 기준 필터와 같은 주파수 크기응답 특성을 가지는 새로운 필터를 설계할 수 있고 식 (4)와 같이 나타낼 수 있다.

$$H_{NEW}(z) = \frac{\prod_{i=1}^P (1 - n_i e^{jw_i} z^{-1})}{\prod_{i=1}^P (1 - m_i e^{jw_i} z^{-1})} \approx H_f(z) W(z) = H_R(z) \quad (4)$$

여기서  $n_i$ 과  $m_i$ 은 실수이고,  $w_i$ 는 영점과 극점의 각도를 나타내는 값이다.

새로운 필터의 구조는 일반적인 지각가중 필터의 구조와 같도록 구성하였다. 여기서, 제안된 알고리즘의 최종

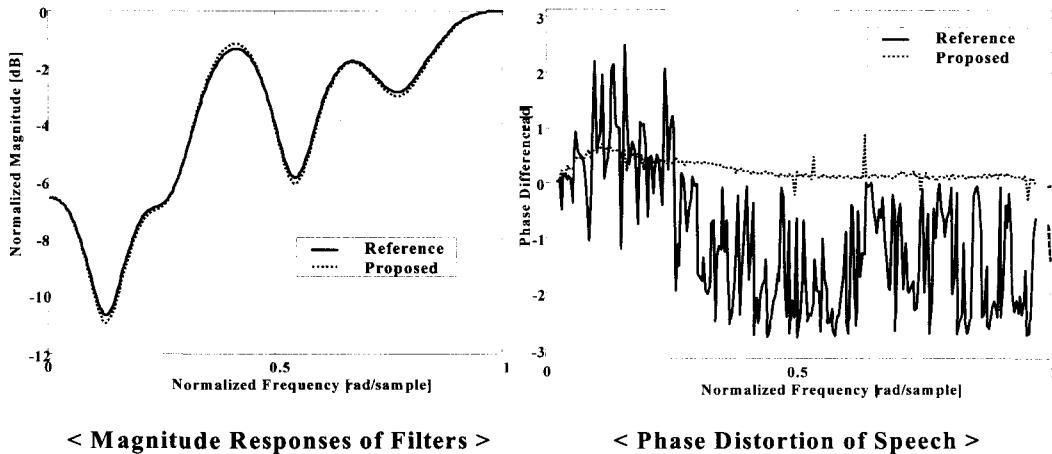


그림 5. (a)주파수 크기응답특성 비교, (b) 위상왜곡 비교  
 Fig. 5. (a) Magnitude Response of Filters, (b) Phase Distortion of speech.

목표는 기준 필터와 같은 주파수 크기응답 특성을 갖도록 새로운 필터의 가중계수를 설정하는 것이다. 두 필터의 차수가 다르기 때문에 식 (4)를 만족하는 새로운 필터를 찾는 것은 상당히 어렵다. 하지만, 두 필터의 극점과 영점의 위치가 같다는 조건을 이용하여 극점과 영점의 위치를 고정시키고, 반복적인 알고리즘으로 기준필터와 거의 유사한 주파수 크기응답 특성을 가지는 새로운 필터를 찾을 수 있다<sup>[9]</sup>. 새로운 필터를 설계하는 과정을 요약하면 아래와 같다.

- ① 기준필터와 새로운 필터의 주파수 크기응답 특성의 차이를 오차로 정의 한다.
- ② 제안된 필터의 가중치 계수를 변화시킨다.
- ③ 최소의 MSE(Mean Square Error)를 가지는 가중 계수를 찾는다.
- ④ 선택된 가중 계수를 새로운 필터에 적용한다.
- ⑤ 설계된 필터로 상호부호화기의 후처리필터와 지각가중 필터를 대신한다.

그림 5-(a)는 기준 필터와 제안된 알고리즘으로 설계된 필터의 주파수 크기응답 특성을 비교하여 보여준다. 제안된 필터의 주파수 크기응답은 기준 필터와 거의 일치함을 알 수 있다. 그림 5-(b)는 제안된 필터("proposed")와 상호부호화기("conventional")의 후처리 필터와 지각가중 필터를 통과하기 전과 통과한 후의 음성의 주파수 위상응답의 차이를 비교하여 보여준다. 제안된 필터를 사용한 경우가 주파수 위상 왜곡이 더욱 적음을 알 수 있다. 이러한 이유는 비선형 위상특성을 가지는 두 필터를 하나의 필터로 대체함으로써 얻어지는 이득이다. 또한 지각가중 필터의 구조를 가지는 제안된 필터를 사용함으로써 일

반적인 상호부호화기에 적용하기가 쉽고, 후처리 필터의 연산량을 감소시킬 수 있다. 표 6은 제안된 필터를 사용한 경우의 객관적 음질을 일반적인 방법과 비교하여 보여 준다. 표에서 알 수 있듯이 제안된 상호부호화기를 위한 새로운 지각가중 필터를 사용함으로써 객관적 음질의 향상을 가져 올 수 있었다.

#### IV. 성능평가

제안된 상호부호화 알고리즘의 성능평가를 위하여, 지연시간과 연산량과 NTT-Korean data 96 음성 샘플<sup>[12]</sup>를 사용하여 객관적 음질 평가와 주관적 음질 평가를 하였다. 연산량 평가를 위해서는 실시간 구현을 위한 고정 소수점 실험(fixed-point simulation)후에 WMOPS(weighted millions operation per second)<sup>[10]</sup>을 사용하였고, 객관적 음질 평가를 위하여 PESQ(perceptual evaluation of speech quality)<sup>[11]</sup>을 사용하였다. 주관적 음질평가를 위해서는 A/B 선호도 평가를 실시하였다. AMR과 G.723.1간의 상호부호화기에 제안된 알고리즘(LSP 변환, 개선된 고속 피치 검색, 상호부호화기를 위한 새로운 지각가중 필터)을 적용하였다. 또한, AMR에서 G.723.1로의 상호부호화기의 경우에는 G.723.1의 적용 코덱북 검색과정의 특징을 이용하여 음질의 저하없이 연산량을 감소시킬 수 있는 고속 적응 코덱북 검색 방법<sup>[3]</sup>을 적용하였다.

##### 1. 연산량 평가

연산량 평가를 위해서 WMOPS<sup>[10]</sup>를 사용하여 tan-dem방식과 제안된 상호부호화기를 비교하였다. 표 7과

표 6. 제안된 필터의 객관적 음질(PESQ)

Table 6. Objective Quality(PESQ) of the proposed filter.

AMR(Kbps)		4.75	5.15	5.9	6.7	7.4	7.95	10.2	12.2
A→G(5.3)	Original	3.113	3.141	3.202	3.229	3.257	3.278	3.309	3.311
	Proposed	3.126	3.154	3.212	3.248	3.275	3.308	3.323	3.328
A→G(6.3)	Original	3.184	3.218	3.286	3.332	3.367	3.389	3.431	3.439
	Proposed	3.197	3.234	3.297	3.348	3.388	3.413	3.441	3.456
G(5.3)→A	Original	3.017	3.059	3.126	3.196	3.247	3.268	3.341	3.376
	Proposed	3.022	3.063	3.128	3.197	3.248	3.269	3.340	3.371
G(6.3)→A	Original	3.061	3.109	3.182	3.253	3.316	3.340	3.428	3.476
	Proposed	3.066	3.112	3.181	3.258	3.322	3.342	3.434	3.476

표 7. 제안된 상호부호화 알고리즘(AMR→G.723.1)의 연산량(WMOPS)

Table 7. WMOPS of the proposed transcoding algorithm(AMR→G.723.1).

AMR(Kbps)		4.75	5.15	5.9	6.7	7.4	7.95	10.2	12.2
A→G(5.3)	Tandem	20.458	20.453	20.457	20.472	20.417	20.470	20.432	20.476
	Transcoding	8.560	8.555	8.559	8.574	8.519	8.572	8.534	8.578
	Reduction(%)	58.158	58.172	58.161	58.118	58.275	58.124	58.232	58.107
A→G(6.3)	Tandem	23.744	23.739	23.743	23.758	23.703	23.756	23.718	23.762
	Transcoding	12.239	12.234	12.238	12.253	12.198	12.251	12.213	12.257
	Reduction(%)	48.454	48.465	48.456	48.426	48.538	48.430	48.507	48.418

표 8. 제안된 상호부호화 알고리즘(G.723.1→AMR)의 연산량(WMOPS)

Table 8. WMOPS of the proposed transcoding algorithm(G.723.1→AMR).

AMR(Kbps)		4.75	5.15	5.9	6.7	7.4	7.95	10.2	12.2
G(5.3)→A	Tandem	10.768	9.778	11.096	13.411	12.644	13.166	12.830	13.120
	Transcoding	8.512	7.522	8.767	11.084	10.317	10.839	10.295	10.057
	Reduction(%)	20.951	23.072	20.975	17.351	18.404	17.674	19.758	23.346
G(6.3)→A	Tandem	10.772	8.778	10.094	12.411	11.644	12.166	11.830	12.120
	Transcoding	8.516	6.522	7.767	10.084	9.317	9.839	9.295	9.057
	Reduction(%)	20.943	25.701	23.053	18.749	19.85	19.127	21.429	25.272

표 8은 제안된 상호부호화 알고리즘의 연산량을 tandem 방식과 비교하여 보여준다. 표에서 보여주듯이 제안된 상호부호화 알고리즘을 사용함으로써 tandem 방식에 비해, AMR에서 G.723.1로의 상호부호화기의 경우 약 48.4%~58.2%의 연산량이 감소하였고, 반대 방향의 경우에는 약 18%~23%의 연산량이 감소하였다.

### 2. 객관적 음질 평가

객관적 음질 평가를 위해서는 NTT-Korean data 96 샘플<sup>[12]</sup>을 사용하여, ITU-T 표준안인 PESQ<sup>[11]</sup>을 사용하였다. 그림 6과 7은 각각 AMR에서 G.723.1 방향과 G.723.1에서 AMR로의 상호부호화기의 객관적 음질을 tandem 방식과 비교하여 보여준다. 그림에서는 편의상 G.723.1의 5.3 kbps와 6.3 kbps의 객관적 음질의 평균을 보여준다. 그림 6에서 보여주듯이 AMR에서 G.723.1 방향의 상호부호화기를 사용함으로써 전체 전송률에 대하여 tandem 방식에 비하여 음질이 향상됨을 알 수 있다.

그림 7에서는 반대 방향의 상호부호화기의 객관적 음질을 보여준다. 이 경우의 상호부호화기는 낮은 전송률에서는 tandem 방식에 비하여 다소 음질이 저하되고, 높은 전송률에 대하여는 음질이 향상되었다. 하지만, 전체 전송률의 평균에 대한 객관적 음질은 향상되었다.

### 3. 주관적 음질 평가

주관적 음질 평가를 위해서는 A/B 선호도 평가를 하였다. 선호도 평가에는 24 명이 참여하였고 남성/여성 화자 각각 4명이 발음한 총 96개의 샘플을 사용했다.

표 9와 10은 A/B 선호도 평가의 결과를 보여준다. 주관적 음질 평가는 AMR의 5.15, 7.95 와 12.2kbps를 사용하였고, 표에서 보여주듯이 선호도 평가에서 G.723.1에서 AMR로의 상호부호화기의 경우와 반대 방향의 경우 모두 tandem 방식에 비하여 높은 선호도를 나타냄을 알 수 있다. G.723.1에서 AMR로의 상호부호화기의 낮은 전송률의 경우 객관적 음질 평가에서는



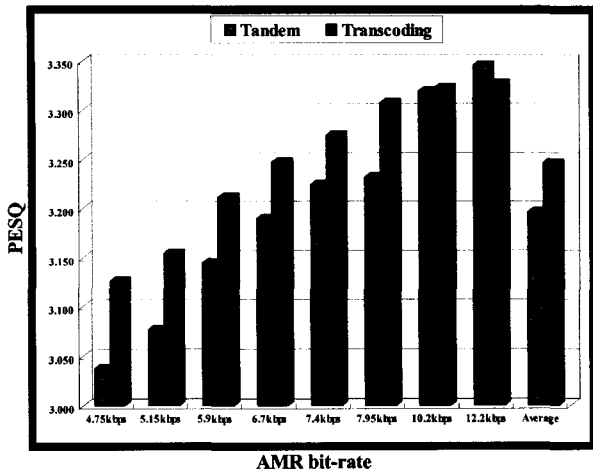


그림 6. 제안된 상호부호화기의 객관적 음질 (AMR→G.723.1)

Fig. 6. Objective quality of the proposed transcoder (AMR→G.723.1).

표 9. 주관적 음질평가(G.723.1→AMR)

Table 9. Subjective quality of the proposed transcoder (G.723.1→AMR).

Direction\bit-rate (kbps)	Preference		
	Tandem	No preference	Transcoding
G(5.3)→AMR(5.15)	29.17%	23.96%	46.88%
G(5.3)→AMR(7.95)	25.00%	33.33%	41.67%
G(5.3)→AMR(12.2)	22.92%	37.50%	39.58%
<b>Average</b>	<b>25.69%</b>	<b>31.60%</b>	<b>42.71%</b>
G(6.3)→AMR(5.15)	35.42%	34.38%	30.21%
G(6.3)→AMR(7.95)	36.46%	27.08%	36.46%
G(6.3)→AMR(12.2)	39.58%	34.38%	26.04%
<b>Average</b>	<b>37.15%</b>	<b>31.94%</b>	<b>30.90%</b>

tandem 방식에 비하여 다소 낮은 객관적 음질을 보였지만, 주관적 음질 평가에서 더욱 선호함을 알 수 있다. 이는 상호부호화 알고리즘을 사용함으로써 음성의 명료도가 높아지는 특징을 가지게 되고, 청취자는 이러한 음색을 더욱 선호하기 때문이다.

### V. 결 론

본 논문에서는 무선 통신에 사용되는 AMR과 VoIP에 사용되는 G.723.1 음성부호화기 간의 상호부호화 알고리즘을 제안하였다. 제안된 알고리즘은 크게 LSP 변환, 개선된 고속 피치 검색, 상호부호화기를 위한 새로운 지각가중 필터를 사용하여 송신단의 비트열을 수신단의 비트열로 변환해 준다. 제안된 알고리즘의 성능평

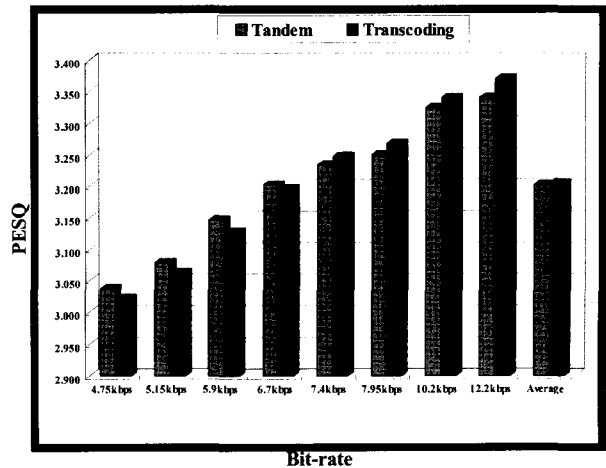


그림 7. 제안된 상호부호화기의 객관적 음질 (G.723.1→AMR)

Fig. 7. Objective quality of the proposed transcoder (G.723.1→AMR).

표 10. 주관적 음질평가(AMR→G.723.1)

Table 10. Subjective quality of the proposed transcoder (AMR→G.723.1).

Direction\bit-rate (kbps)	Preference		
	Tandem	No preference	Transcoding
AMR(5.15)→G(5.3)	19.79%	25.00%	55.21%
AMR(7.95)→G(5.3)	27.08%	36.46%	36.46%
AMR(12.2)→G(5.3)	29.17%	39.58%	31.25%
<b>Average</b>	<b>25.35%</b>	<b>33.68%</b>	<b>40.97%</b>
AMR(5.15)→G(6.3)	21.88%	36.46%	41.67%
AMR(7.95)→G(6.3)	25.00%	45.83%	29.17%
AMR(12.2)→G(6.3)	25.00%	45.83%	29.17%
<b>Average</b>	<b>24.65%</b>	<b>40.63%</b>	<b>34.72%</b>

가를 위하여 연산량, 객관적 음질과 주관적 음질 평가를 하였다. 제안된 상호부호화 알고리즘을 사용함으로써, tandem 방식에 비하여, 전송률과 방향에 따라 약 20%-58%의 연산량을 감소시킬 수 있고, 객관적 음질의 향상을 가져 올 수 있다. 주관적 음질 평가를 통하여, 제안된 상호부호화 알고리즘은 더욱 높은 선호도를 보임을 알 수 있었다. 이러한 결과들은 제안된 상호부호화 알고리즘이 기존 tandem방식에 비해 높은 성능을 보임을 입증한다. 또한, 제안된 LSP변환 방법, 개선된 고속 피치 검색 방법과 상호부호화기를 위한 새로운 지각가중 필터는 다른 상호부호화기의 송신단과 수신단의 음성부호화기의 특징을 이용하여, 쉽게 확장하여 적용할 수 있고, 많은 연산량을 음질 저하없이 감소시킬 수 있는 장점을 가지고 있다.

참 고 문 헌

[1] ITU-T Rec. G.723.1 "Dual-rate speech coder for multi-media communications transmitting at 5.3 and 6.3 kbit/s," 1996

[2] 3GPP TS 26.090 V5.0.0, AMR speech codec; Transcoding functions, Jun 2002

[3] S.W.Yoon, S.K. Jung, Y.C. Park, and D.H. Youn, "An efficient transcoding algorithm for G.723.1 and G.729A speech coders," in Proc. Eurospeech 2001, pp. 2499~2502, Sep 2001.

[4] K.T. Kim, S.K. Jung, Y.C. Park, Y.S. Choi, D.H. Youn, "An efficient transcoding algorithm for G.723.1 and EVRC speech coders," in Proc. IEEE VTS 54th Vehicular Technology Conference (VTC 2001), vol.3, pp.1561-1564, Oct.7-10, 2001

[5] J.K. Choi, C.H. Lee, H.G. kang, Y.C. Park and D.H. Youn, "Improvement issues on transcoding algorithm for the flexible usage to the various pairs of speech codec," in ICASSP 2004, to be published, 2004.

[6] H.G. Kang, H.K. Kim and Richard V. Cox, "Improving the Transcoding Capability of Speech codecs," in IEEE Transaction on Multimedia, VOL. 5, NO. 1, Mar., 2003.

[7] A.M.Kondoz, Digital speech coding for low rate communication system, John Wiley & Sons, 1994.

[8] W.B. Kelson, Speech coding and synthesis, Elsevier Science B.V., 1995

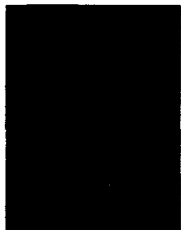
[9] Simon Haykin, ADAPTIVE FILTER THEORY 4th Edition, Pentice-Hall, Inc., 2002

[10] ITU-T Draft Rec P.191 "Software tools for speech and audio coding standardization," Nov 2000.

[11] ITU-T Draft Rec P.862 "Perceptual evaluation of speech quality (PESQ), an objective method of end-to-end speech quality assessment of narrowband telephone networks and speech codecs," May 2000.

[12] <http://www.ntt-at.com>

저 자 소 개



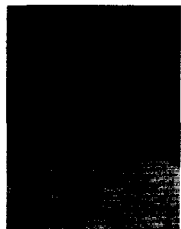
최진규(정회원)  
 2002년 중앙대학교 전자공학과  
 학사졸업.  
 2004년 연세대학교 전기전자  
 공학과 석사졸업.  
 2004년 LG전자 근무

<주관심분야: 신호처리, 음성부호화, 상호부호화>



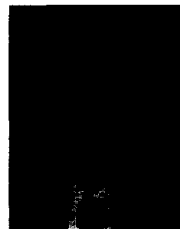
강홍구(정회원)  
 1989년 연세대학교 전자공학과  
 학사졸업.  
 1991년 연세대학교 전기전자  
 공학과 석사졸업.  
 1995년 연세대학교 전기전자  
 공학과 박사졸업.

2002년~ 현재 연세대학교 전기전자공학과 조교수  
 <주관심분야: 적응신호처리, 음성부호화, 화자인식>



윤성완(정회원)  
 1999년 연세대학교 전자공학과  
 학사졸업.  
 2001년 연세대학교 전기전자  
 공학과 석사졸업.  
 2001년~현재 연세대학교 전기  
 전자공학과 박사과정

<주관심분야: 신호처리, 음성부호화, 화자인식>



윤대희(정회원)  
 1977년 연세대학교 전자공학과  
 학사졸업.  
 1979년 Kansas State University  
 석사졸업.  
 1982년 Kansas State University  
 박사졸업.

1985년~현재 연세대학교 전기전자공학과 정교수  
 <주관심분야: 적응신호처리, 음성신호처리, 오디오  
 신호처리>