2013년 12월 전자공학회 논문지 제 50 권 제 12 호
Journal of The Institute of Electronics Engineers of Korea Vol. 50, NO. 12, December 2013 http://dx.doi.org/10.5573/ieek.2013.50.12.232

논문 2013-50-12-28

주파수 변환의 길이에 따른 위상 기반 오디오 정보 은닉 기술의 음질 및 성능 분석

(Performance Analysis of Audio Data Hiding Method based on Phase Information with Various Window Length)

조 기 호*, 김 남 수**

(Kiho Cho and Nam Soo Kim©)

요 약

주파수 영역에서의 위상값을 이용하는 오디오 정보 은닉 기술에서 주파수 변환의 길이의 역할은 매우 중요하다. 본 논문에서는 MCLT를 사용하는 주파수 위상 기반 오디오 정보 은닉 기술의 주파수 변환의 길이에 따른 음질과 실내 반향 환경에서의 성능을 분석하였다. 주파수 변환의 길이가 길어질수록 정보가 삽입된 오디오 신호의 음질이 떨어지고 실내 반향 환경에서의 검출 성능은 좋아지는 결과를 보였다. 음질 변화의 가장 큰 원인으로 주파수 변환 윈도우 안에서 급격한 신호 크기의 변화가 생길 때 그 신호가 퍼지게 되는 프리에코 현상의 존재를 확인하였다. 또한 반향에 강인하려면 윈도우 길이가 길어야 한다는 무선 통신 이론이 정보 은닉 기술 및 음향 신호 전송에도 적용이 됨을 확인하였다.

Abstract

The role of the window length of time-frequency transformation is important for the audio data hiding methods utilizing phase information. In this paper, the experiments for our audio data hiding method were conducted in order to evaluate the audio quality and robustness against reverberant environment. The experimental results showed the tendency that the worse audio quality but better robustness were obtained when the lengthy window was applied. The important reason for quality degradation was pre-echo which flatters the percussive sound. The results also indicated that the wireless communication theory related to the length of time-frequency transform can be applied in the field of audio data hiding and acoustic data transmission.

Keywords : 오디오 정보 은닉, MCLT (Modulated Complex Lapped Transform), 주파수 위상, 음향 데이터 전송, PEAQ (Perceptual Evaluation of Audio Quality)

Ⅰ. 서 론

오디오 정보 은닉은 오디오 신호에 사람의 귀에 들리지 않는 정보를 삽입하고 배포되어 사용자에게 특정 정보를 제공하는 기술이다. 오디오 정보 은닉은 저작권 증명 등을 위한 오디오 워터마킹, 스테가노그라피, 비밀 통신 (covert communication) 등에 이용되어 왔다[1].

지금까지 다양한 오디오 정보 은닉 (또는 오디오 워터마킹) 기법들이 제안되어 왔으며, 크게 샘플 비트 변

* 학생회원, ** 정회원, 서울대학교 전기·정보공학부 및 뉴미디어통신공동연구소
(Department of Electrical and Computer Engineering and INMC, Seoul National University)
© Corresponding Author (E-mail: nkim@snu.ac.kr)
※ 본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2013-H0301-13-4005).
접수일자: 2013년10월29일, 수정완료일: 2013년11월27일

경, 주파수 크기 변화, 에코 삽입, 주파수 위상 변화로 구분될 수 있다. 또한 성능 개선을 위해 대역 확산을 이용한 데이터 중복 삽입과 심리음향 모델의 이용[2], 그리고 채널에서 발생하는 에러를 복원하기 위한 채널 부호화 등이 이용된다[3].

이 중에서 음향 OFDM (Orthogonal Frequency Division Multiplexing)[4] 등 주파수 위상을 이용한 오디오 정보 은닉 기술들은 배포 과정에서 생길 수 있는 공격에 대한 강인성과 상대적으로 높은 데이터 전송량을 보인다. 하지만 음향 OFDM 에서는 위상의 급격한 변화로 인하여 음질 저하가 존재하며, 특히 음성이나 클래식 음악 등 조용한 음악에서 심한 음질 저하를 보인다[5].

이러한 문제를 해결하기 위해서는 주파수 윈도우 간의 중첩이 존재해야 하며, MCLT (Modulated Complex Lapped Transform)[6]를 사용하는 주파수 위상 기반 오디오 정보 은닉 기법으로 음향 OFDM이 가지고 있는 문제를 해결하였다[7]. MCLT는 인접한 MCLT 변환 단위(프레임)끼리는 서로 절반씩 중첩되기 때문에 음향 OFDM에서 사용하는 퓨리에 변환에서 생기는 블록 효과로 인한 음질 감소 등이 존재하지 않는다.

이러한 주파수 위상 기반 오디오 은닉 기법은 음향 데이터 전송에 이용할 수 있다[7]. 음향 데이터 전송은 정보가 삽입된 음원을 스피커로 재생하여 전송하고, 마이크로 녹음하여 추출하는 근거리 무선 전송 기법이다. 무선 통신 이론에 따르면, 스피커와 마이크 사이의 공간에서 생기는 채널에 대한 강인성은 주파수 변환의 길이가 길어질수록 좋아진다[8].

하지만 주파수 변환 윈도우 길이에 따른 위상 변화에 대한 음질에 대한 분석은 자세히 이루어 진 바 없다. 음성의 명료도에 대한 위상 스펙트럼의 중요성을 파악하는 실험 결과가 있지만[9], 오디오 정보 은닉 기법에 그대로 적용하기 어렵다. 또한 무선 통신 이론의 주파수 변환 길이에 따른 강인성에 대한 관계도 음향 데이터 전송 등의 오디오 정보 은닉 응용에서 경향성이 증명된 바 없다. 따라서 본 논문에서는, 기존에 제안된 주파수 위상 기반 오디오 은닉 기법[7]을 확장한 오디오 은닉 기법을 제안하고, 주파수 변환 윈도우 길이에 따른 음질 및 검출 성능의 변화를 분석한다.

Ⅱ. 주파수 위상 기반 오디오 정보 은닉 기법

1. MCLT

MCLT는 길이 2M의 실수 시간축 신호를 길이 M의 복소수 주파수 계수로 바꾸어주는 변환이다. MCLT는 윈도우 길이의 절반씩 중첩해 나가는 것이 특징이다. 시간 축 신호를 $x(n)$라고 하면 k번째 주파수에서의 MCLT 계수 $X(k)$는 다음과 같이 정해진다.

$$X(k) = \sqrt{\frac{2}{M}} \sum_{n=0}^{2M-1} x(n)h(n) \exp[-j(n+\frac{M+1}{2})(k+\frac{1}{2})\frac{\pi}{M}] \quad (1)$$

여기에서 $h(n)$은 사인 윈도우로 다음과 같다.

$$h(n) = -\sin[(n+\frac{1}{2})\frac{\pi}{2M}] \quad (2)$$

2. 정보의 삽입

정보를 삽입하는 과정은 그림 1과 같이 오디오 신호의 MCLT 계수를 구하여 데이터 비트 값에 따라 값을 변경한다. 변경된 MCLT 프레임은 다시 시간 축 신호로 되돌려 인접한 프레임과 절반씩 중첩하여 더한다.

주파수 위상 기반 오디오 정보 은닉 기법에서는 주파수 변환 계수의 위상 값에 따라 어떤 비트값을 삽입하는지 결정이 된다. 데이터 삽입은 오디오 신호의 MCLT 계수의 위상을 이상적 채널에서 검출 시 비트 값이 0 또는 π가 되도록 한다. 즉, 데이터가 삽입된 신

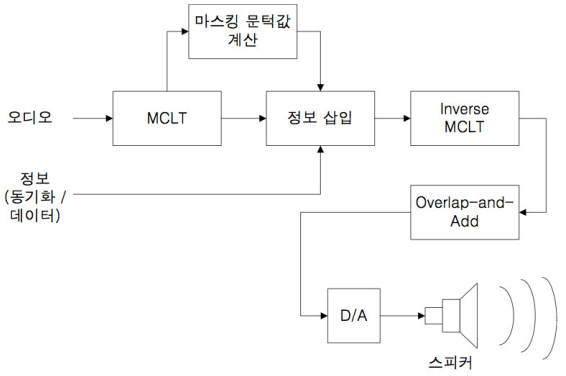

그림 1. 주파수 위상 기반 오디오 정보 은닉 기법의 정보 삽입 과정
Fig. 1. Data embedding procedure of proposed audio data hiding method.

호를 다시 MCLT 변환을 하여 얻은 MCLT 계수 $\widehat{Y}_i(k)$가 다음과 같아야 한다.

$$\widehat{Y}_i(k) = |X_i(k)|b_i(k) \tag{3}$$

여기에서 i는 프레임 인덱스를 나타내고, $|X_i(k)|$는 원본 오디오 신호의 MCLT의 크기, 그리고 $b_i(k)$는 비트값을 나타내며 -1 또는 1의 값을 가진다.

데이터 삽입된 결과가 데이터 검출 과정에서 식 (3)과 같이 되려면 데이터가 삽입된 MCLT 계수 $\widehat{X}_i(k)$는 다음과 같이 결정된다.

$$\begin{aligned}\widehat{X}_i(k) &= 2|X_i(k)|b_i(k) - j[(\boldsymbol{A_{-1}})_k \boldsymbol{X_{i-1}} \\ &+ \frac{1}{2}X_i(k-1) - \frac{1}{2}X_i(k+1) + (\boldsymbol{A_1})_k \boldsymbol{X_{i+1}}]\end{aligned} \tag{4}$$

여기에서 $\boldsymbol{X_{i-1}}, \boldsymbol{X_{i+1}}$는 원본 오디오 신호의 $i-1$, $i+1$번째 프레임에서의 MCLT 계수 전체 벡터를 의미하고, $(\boldsymbol{A_m})_k$는 이전 프레임 $(m=-1)$과 다음 프레임 $(m=1)$의 중첩으로 인한 성분을 의미한다. $(\boldsymbol{A_m})_k$의 l번째 성분 $(\boldsymbol{A_m})_{kl}$은 다음과 같다.

$$(\boldsymbol{A_m})_{kl} = \begin{cases} (-m)\dfrac{(-1)^{l+d}}{\pi(2d-1)(2d+1)} & \text{if } |l-k| = 2d \\ \dfrac{(-1)^l}{4} & \text{elseif } |l-k| = 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

여기에서 d는 양수이다. 식 (4)를 보면, 한 위치에서 데이터를 삽입하기 위해서는 이전 프레임과 다음 프레임, 그리고 바로 아래 주파수와 위 주파수가 모두 필요함을 알 수 있다. 따라서 시간, 주파수 방향으로 한 인덱스씩 띄어서 데이터를 삽입한다. 즉, 시간 방향으로 $\ldots, i-2, i, i+2, \ldots$번째 프레임에, 주파수 방향으로 $\ldots, k-2, k, k+2, \ldots$번째 주파수에 데이터를 삽입한다.

데이터 전송은 패킷 단위로 하며, 본 논문에서는 2개의 데이터 프레임 사이에 동기화 프레임을 삽입한다. 동기화 프레임은 데이터 프레임의 위치와 채널 영향을 추정할 때의 파일럿 신호 역할을 한다.

3. 정보의 검출

정보를 검출하는 과정은 그림 2와 같이 동기화, MCLT 계수 추출, 수신된 비트를 판단하는 디코딩 과정으로 나뉘어진다. 데이터 프레임의 위치를 알기 위한

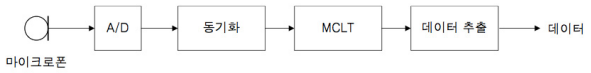

그림 2. 주파수 위상 기반 오디오 정보 은닉 기법의 정보 검출 과정

Fig. 2. Data extraction procedure of proposed audio data hiding method.

동기화 과정은 주파수 영역에서 상관 (correlation)을 구하여 그 값이 최대가 되는 지점을 구하여 그 지점에 해당하는 위치에서의 MCLT 계수를 얻는다. 동기화 과정이 끝난 뒤 얻어지는 파일럿 프레임의 MCLT 계수를 이용하여 채널에서 생기는 왜곡을 상쇄하기 위한 위너 (Wiener) 필터 기반의 채널 추정 및 보정을 수행 후, 데이터 프레임의 MCLT 계수의 위상 값에 따라 데이터 비트를 결정한다.

Ⅲ. 주파수 변환 길이에 따른 성능 변화

1. 위상 변화 시스템

위상 변화 시스템은 입력 신호를 주파수 영역으로 변환한 후, 변환된 주파수 계수들의 크기는 그대로 두고 위상만 변화시킨 후, 다시 시간 영역으로 역변환하는 시스템을 의미한다. 주파수 영역으로 변환하는 행렬을 F라고 하고, 각 주파수 인덱스 k의 위상 변화량을 $\phi(k)$라고 할 때, 입력 신호 x와 출력 신호 y의 관계는 다음과 같다.

$$y = \mathrm{F}^{-1}\Phi\mathrm{F}x \tag{6}$$

여기에서 Φ는 위상 변화를 나타내는 대각 (diagonal) 행렬이며, (k,k)번째 항은 $e^{j\phi(k)}$이다.

2. 위상 변화에 대한 인간의 청각 지각

이러한 위상 변화 시스템에서 생기는 신호 변화의 청각적 특성에 대해서 알려진 바로는, 인간은 오디오 신호의 위상 변화에 대체적으로 둔감하다[10]. 특히, 일반적인 음악이나 음성에 대해서 보통의 경우 위상 변화에 따른 음질 차이를 크게 느끼지 못하는 경향성이 있다[10].

하지만 위상 스펙트럼이 차지하는 중요성은 위상을 구하기 위한 주파수 변환의 길이가 길어질수록 커지는 경향을 보인다[9]. 인간이 음성을 인지하는 데에 있어서

의 위상 스펙트럼의 역할을 측정하기 위한 실험에서 퓨리에 변환 윈도우의 길이가 128 msec 이상이 되면 크기 (magnitude) 스펙트럼보다 오히려 더 중요하다는 결과가 있다[9].

이러한 기존 실험 결과들을 바탕으로 다음 장에서는 주파수 위상 기반 오디오 정보 은닉 기법의 주파수 변환 길이에 따른 음질 변화를 분석한다.

3. 주파수 변환 길이와 채널에 대한 강인성의 관계

음향 신호 전송에서의 채널에 대한 영향은 시간 축에서 필터와 컨볼루션 연산을 한 것이라고 볼 수 있다. 채널을 나타내는 필터의 길이가 길어질수록 이전 프레임들의 영향이 더 커지게 된다. 무선 통신에서는 채널 필터의 길이와 관련된 파라미터로 최대 지연 확산을 이용하는데, 주파수 변환의 윈도우 길이는 최대 지연 확산보다 훨씬 길도록 정해야 한다[8]. 즉, 주파수 변환의 길이가 길어질수록 필터 형태의 채널에 더 강인하다.

Ⅳ. 실 험

1. 실험 방법

음질 변화의 측정은 객관적 음질 측정 방식인 PEAQ (Perceptual Evaluation of Audio Quality)를 사용하였다[11]. PEAQ 점수는 원본과 실험 대상을 비교하여 0점부터 -4점까지 정의가 된다. 0점은 원본과 실험 대상이 완전히 같다는 의미이고, -1점은 차이가 느껴지지만 음질 저하가 귀에 거슬리지 않는 수준이다. -2점과 -3점은 음질 저하가 귀에 거슬리는 수준이고, -4점은 심각한 음질 저하를 의미한다. 검출 성능 측정은 비트 오류율 (Bit error rate, BER)을 사용하였다. 전체 수신된 비트 중 오류가 있는 비트의 비율을 의미한다.

음원은 30초 길이의 44.1 kHz 샘플링 주파수를 가진 16개의 서로 다른 음악을 사용하였다. 데이터 삽입은 3.5 kHz부터 6.5 kHz 대역을 이용하였다. 주파수 변환은 MCLT를 사용하였으며, MCLT 길이는 512 부터 8192 까지를 실험 범위로 하였다. (이 때, 시간 축 윈도우 길이는 각각 1024 샘플부터 16384 샘플이다.)

2. 주파수 변환 길이에 따른 음질 변화

주파수 위상 기반 오디오 정보 은닉 기법의 주파수 변환 길이에 따른 PEAQ 점수는 그림 3과 같다.

그림 3에서 알 수 있듯이, 주파수 변환 윈도우 길이가 512로 짧은 경우에는 PEAQ 점수가 -0.233점으로 높은 음질을 보였지만, 윈도우 길이가 길어질수록 PEAQ 점수도 떨어지고 음원마다 편차도 커지는 결과를 얻었다. 즉, 윈도우 길이가 길어질수록 위상 변화에 의한 음질 저하가 커진다는 의미이고, 이것은 음성 인지 실험 결과[9]와 일치하는 경향성을 보였다. 특히 MCLT 길이가 8192인 경우 (시간 축 윈도우 길이로는 16384 샘플, 시간으로는 371 msec) PEAQ 점수가 평균 -0.883점으로 음질 저하가 듣기에 신경 쓰일 수 있는 수준이다.

이러한 음질 저하의 가장 큰 원인으로는 그림 4와

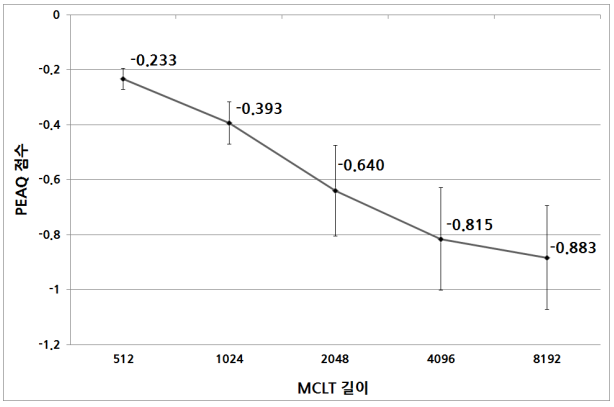

그림 3. MCLT 길이에 따른 제안된 오디오 정보 은닉 기법의 객관적 음질 평가 (PEAQ) 점수 (점수 위 막대는 95% 신뢰구간을 의미)

Fig. 3. PEAQ score of data hiding based on phase of MCLT coefficients with various window length. Vertical lines on the top of bars denote the 95% confidence intervals.

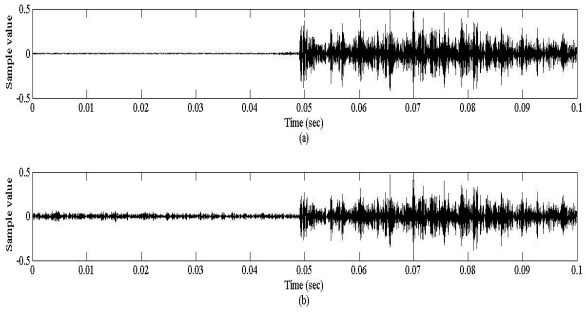

그림 4. MCLT 윈도우 길이가 0.371초인 경우의 프리에코 발생 예시: (a) 원본 신호, (b) 데이터가 삽입된 신호

Fig. 4. Example of pre-echo where the length of MCLT window is 0.371 second: (a) original signal, (b) data-embedded signal.

같은 프리에코 (pre-echo) 현상을 꼽을 수 있다. 주파수 축에서의 위상은 시간 축에서 지연 (delay)과 밀접한 관계가 있다. 위상이 변화하게 되면 자연스럽게 기존 시간 축 신호의 주파수에 따른 지연이 변화하게 된다. 그런데 만약에 하나의 윈도우 안에 드럼이나 음악의 시작 (onset)과 같이 시간 축에서 급격한 크기 변화가 일어나는 충돌성 (percussive) 신호가 있는 경우, 소리가 그림 4의 (b)와 같이 한 윈도우 시간 전체로 흩뿌려지게 되며, 이러한 현상이 프리에코이다. 주파수 변환 윈도우의 길이가 길어질수록 충돌성 신호의 개수와 흩뿌려지는 시간이 길어지게 되어 음질이 나빠지게 되는 것이다.

3. 주파수 변환 길이에 따른 검출 성능 변화

실내 반향이 존재하는 환경에서 제안된 오디오 정보 은닉 기술의 주파수 변환 길이에 따른 검출 성능을 측정하였다. 실내 반향에 해당하는 채널 필터는 이미지 방식 기반의 시뮬레이션으로 구하였다[12]. 실내의 크기와 스피커, 마이크의 위치는 그림 5에 표시하였으며 (좌측 하단이 원점), 방 벽면의 반사 계수 (reflection coefficient)는 -0.9로 반향이 큰 상황을 전제하였다. 이뿐 아니라 20 dB의 백색 잡음을 추가하였다.

그리고 그림 5의 환경에서 정의된 채널 필터는 그림 6과 같다. 처음 임펄스가 생기는 시간부터 마지막 샘플까지의 길이는 약 0.19초이다.

그림 6에서 정의된 채널 하에서, 주파수 변환 길이에

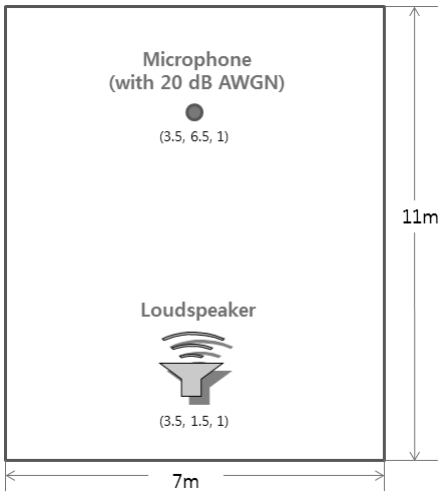

그림 5. 채널 실험 환경 (높이 3m, 반사계수 -0.9)
Fig. 5. Environment of simulated room.
(Height: 3 m and reflection coefficient: -0.9)

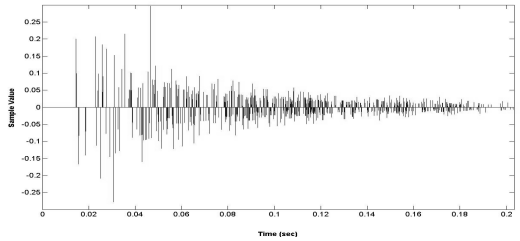

그림 6. 채널 실험 환경의 임펄스 응답
Fig. 6. Impulse response of simulated room.

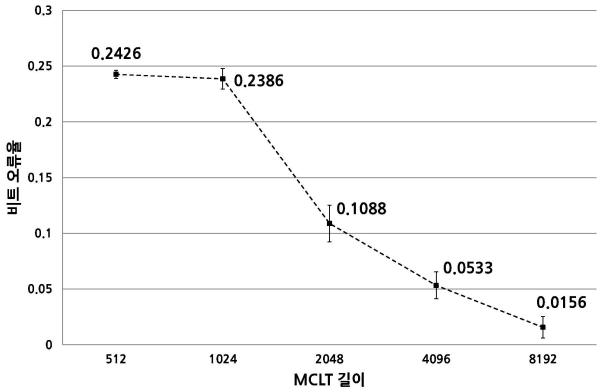

그림 7. MCLT 길이에 따른 제안된 오디오 정보 은닉 기법의 그림 6에서 정의된 채널에서의 비트 오류율 (점수 위 막대는 95% 신뢰구간을 의미)
Fig. 7. Bit error rate of data hiding based on phase of MCLT coefficients under the channel displayed in Fig. 6 with various window length. Vertical lines on the top of bars denote the 95% confidence intervals.

따른 제안된 오디오 정보 은닉 기술의 비트 오류율은 그림 7과 같다.

그림 7에서 알 수 있듯이, 주파수 변환의 길이가 길어질수록 비트 오류율이 낮아짐을 알 수 있다. MCLT 길이가 4096일 때 약 0.19초, 8196일 때 약 0.37초이므로, 채널 필터의 길이보다 윈도우 길이가 긴 경우에는 비트 오류율이 그렇지 않은 경우에 비해 매우 낮아진다. 이것은 주파수 변환 길이가 길면 채널의 시간 지연에 강인하다는 무선 통신 이론이 정보 은닉 기술과 음향 신호 전송 분야에도 적용이 가능함을 확인할 수 있는 결과이다.

V. 결 론

본 논문에서는, 주파수 위상 기반 오디오 정보 은닉 기술에서 주파수 변환의 길이에 따른 음질 및 검출 성

능의 변화를 분석하였다. 실험 결과, 주파수 변환의 길이가 길어지면 음질이 떨어지는 대신, 필터 형태의 채널 임펄스 응답에 대해서는 더 나은 성능을 보이는 trade-off 관계를 가졌다. 따라서 목표로 하는 응용에 따라서 알맞은 주파수 변환 길이를 결정하는 것이 관련 시스템 설계에 매우 중요한 역할을 한다고 볼 수 있다. 더 나아가, 이러한 실험 결과에 따라 오디오 또는 음성을 합성 또는 재현하는 경우에 주파수 변환 윈도우 길이를 결정하는 문제에 있어서 본 논문의 실험 결과가 참고가 될 수 있을 것이다.

REFERENCES

[1] N. Cvejic and T. Seppänen, *Digital Audio Watermarking Tehcniques and Technologies*, IGI Global, 2008.
[2] J. S. Noh and K. H. Rhee, "High quality audio watermarking using spread spectrum and psychoacsoutic model,," *The Institute of Electronics of Engineers of Korea - Signal Processing*, vol. 43, no. 5, pp. 48-56, Sep. 2006.
[3] Y. Suzuki, R. Nishimura, and H. Tao, "Audio watermarking enhanced by LDPC coding for air transmission", in Proc. of *IIH-MSP 2006*, pp. 23-26, Pasadena, CA, USA, Dec. 2006.
[4] H. Matsuoka, Y. Nakashima, and T. Yoshimura, "Acoustic OFDM System and Performance Analysis," *IEEE Trans. Fundamentals*, vol. E91-A, no. 7, pp. 1652-1658, Jul. 2008.
[5] K. Cho, H. S. Yun, J.-H. chang, and N. S. Kim, "An analysis on audio quality deterioration of acoustic OFDM," *Journal of the Acoustic Society of Korea*, vol. 28, no. 2, pp. 107-111, Feb. 2009.
[6] H. S. Malvar, "Fast algorithm for the modulated complex lapped transform," *IEEE Signal Processing Letter*, vol. 10, no. 1, pp. 8-10, Jan. 2003.
[7] K. Cho, H. S. Yun, and N. S. Kim, "Data hing technique based on modulated complex lapped transform for acoustic data transmission," in *Proc. IEEK Autumn Conference*, vol. 32, no. 2, pp. 235-236, Nov. 2009.
[8] H. Schulze and C. Lueders, *Theory and Applications of OFDM and CDMA: Wideband Wireless Communications*, Jone Wiley & Sons, 2005.
[9] K. K. Paliwal and L. D. Alsteris, "On the usefulness of STFT phase spectrum in human listening tests," *Speech Communication*, vol. 45, issue. 2, pp. 153-170, Feb. 2005.
[10] S. P. Lipshitz, M. Pocock, and J. Vanderkooy, "On the audibility of midrange phase distortion in audio systems," *Journal of Audio Engineering Society*, vol. 30, pp. 580-595, Sep. 1982.
[11] "Method for objective measurements of perceived audio quality," *ITU-R Recommendation BS. 1387-1*, Nov. 2011.
[12] S. G. McGovern, "Fast Image Method for Impulse Response Calculators of Box-shaped Rooms," *Applied Acoustics*, vol. 70, issue. 1, pp. 182-189, Jan. 2009.

저 자 소 개

조 기 호(학생회원)
2007년 서울대학교 전기공학부 학사 졸업.
2009년 서울대학교 전기컴퓨터 공학부 석사 졸업.
2009년~현재 서울대학교 전기정보공학부 박사과정.
<주관심분야 : 오디오정보은닉, 오디오신호처리, 입체음향>

김 남 수(정회원)
1988년 서울대학교 전기공학과 학사 졸업.
1990년 한국과학기술원 전기 및 전자공학과 석사 졸업.
1994년 한국과학기술원 전기 및 전자공학과 박사 졸업.
1993년~1998년 삼성종합기술원 전문연구원.
1998년~현재 서울대학교 전기정보공학부 교수.
<주관심분야 : 음성신호처리, 오디오신호처리, 입체음향, 기계학습>