

# 다자간 음성통화 품질 향상을 위한 오디오 믹서 알고리즘

## Audio Mixer Algorithm for Enhancing Speech Quality of Multi-party Audio Telephony

류상현, 김형국<sup>†</sup>

(Sang-Hyeon Ryu and Hyoung-Gook Kim<sup>†</sup>)

광운대학교 전자공학과

(접수일자: 2013년 6월 27일; 수정일자: 2013년 8월 14일; 채택일자: 2013년 8월 28일)

**초 록:** 두세 명 혹은 그 이상의 참가자간사이의 다자간통화 시 음량불균형, 음량포화, 잡음레벨상승으로 인해서 음질 저하가 발생한다. 이 문제를 해결하기 위해서 본 논문은 소프트웨어 기반의 다지점제어장치를 위한 향상된 오디오 믹싱 알고리즘을 제안한다. 제안된 방식은 음성구간검출과 게인콘트롤이 결합된 기술로서 음성신호 분류, 음량 추정, 게인값 적용, 모든 채널의 음성신호를 믹싱하는 알고리즘들로 구성되어 있다. 제안된 오디오 믹싱 알고리즘은 효율적인 연산과 고품질의 음성을 제공하며, 실질적인 다자간 음성 통화에 적합하다.

**핵심용어:** 다자간 음성통화, 음성구간검출, 게인콘트롤, 음량균일화

**ABSTRACT:** The speech quality of multi-party audio telephony between two, three or more participants is decreased by audio volume imbalance, audio volume saturation and noise level increase. To solve this issue, this paper proposes an advanced audio mixing algorithm for software-based multi-point control unit. Our approach is based on the combined voice activity detection and gain control technique that consists of a set of algorithms that classify audio signals, estimate audio volumes, adjust gain factors and mix audio signals of all channels. The proposed audio mixing algorithm is computationally efficient, delivers high-quality speech, and is suitable for use in any practical multi-party audio telephony.

**Keywords:** Multi-party audio telephony, Voice activity detection, Gain control, Audio volume equalization

**PACS numbers:** 43.60.Dh, 43.72.Ja, 43.72.Kb

### 1. 서 론

통합 네트워크 환경으로 발전되고 있는 인터넷을 중심으로 Voice over Internet Protocol (VoIP)은 다양한 영역과 융합되어 기존의 이동통신 시장을 대체할 수 있는 대안으로 급부상하고 있다. 최근에는 일대일 방식의 VoIP 서비스에서 여러 명이 한꺼번에 음성통화 서비스를 이용할 수 있는 ‘다자간 음성통화’ 시스템<sup>[1-3]</sup>이 등장하고 있다. 다자간 음성통화 시스템은 컴퓨터 혹은 스마트폰 및 네트워크를 이용하여 음성 데이터를 실시간으로 전송함으로써 컴퓨터, 스마트

TV 혹은 스마트폰을 이용해 여러 명의 사용자가 음성대화를 가능하게 하는 시스템이다. 이러한 다자간 음성통화가 실시간으로 이루어지기 위해서는 다자간 음성통화 시에 여러 참여자들로부터 전송받은 음성 정보를 실시간으로 믹싱을 수행하는 오디오 믹서<sup>[1]</sup>의 역할이 매우 중요하다. 하지만 각 참여자들의 사용 환경과 디바이스 차이로 인해서 믹서에 송신된 음성의 음량과 배경잡음이 다양하기 때문에, 믹싱 시 음량의 한계점에 도달하는 음량포화, 여러 잡음의 합성으로 인한 잡음레벨상승, 그리고 어떤 참여자의 음성은 크게 반대로 어떤 참여자의 음성은 작게 들리는 음량불균형으로 인한 통화품질 저하가 발생한다. 기존의 오디오 믹서<sup>[4-5]</sup>는 믹싱 시 발생하는 음량포화와 잡음레벨상승을 막기 위해서 입력된 음

<sup>†</sup>Corresponding author: Hyoung-Gook Kim (hkim@kw.ac.kr)  
Department of Wireless Communication Engineering, Kwang-woon University Room 405, Chambit Building 447-1, Wolgye-Dong, Nowon-Gu, Seoul 139-701, Republic of Korea  
(Tel: 82-2-940-5574, Fax: 82-2-913-5006)

성신호의 크기를 줄여주는 게인콘트롤을 사용했기 때문에, 부정확한 게인콘트롤로 인해서 음량불균형을 해결하지 못했다.

본 논문에서는 음량포화 및 잡음레벨 상승과 음량 불균형을 해소하기 위해서, 음성구간검출을 적용함으로써 각 채널로부터 입력되는 연속된 음성 신호를 기준음량에 맞추는 게인콘트롤을 사용해서 음량포화를 방지하고 음량을 균일화시키는 방법을 이용해서 문제를 해결하고자 한다.

본 논문의 구성은 다음과 같다. 전체 시스템 구조에 대해서 II장에서 설명하고, III장에서는 음성구간 검출 알고리즘의 세부구조, IV장에서는 게인콘트롤에 대한 설명, V장에서는 믹싱콘트롤에 대해서 설명하며, VI장에서는 실험결과에 대해서 설명한다. 마지막으로 VII장에서 결론을 서술한다.

## II. 오디오 믹서 알고리즘의 구조

본 논문에서 제안한 다자간 음성통화 품질 향상을 위한 오디오 믹서 알고리즘의 구조는 Fig. 1과 같이 음성구간검출기, 게인콘트롤부, 믹싱콘트롤부로 구성된다.

기존의 오디오 믹서는 음성 및 비음성구간 구분 없이 임의의 음량을 기준으로 설정하여 모든 음량을 균일화하는 방식을 사용하기 때문에 잡음레벨까지도 상승하여 클리핑 현상이 발생하고, 급격한 게인값 변동을 수반하므로 음량균일화 유지 및 음량포화 방지가 어렵기 때문에, 본 논문에서는 잡음에 강한 음성구간검출기를 통해서 연속되는 음성구간단

위로 게인값을 구하여 적용함으로써 안정적인 음량을 유지할 수 있다.

또한, 음량균일화를 위한 게인콘트롤에서는 각 채널의 입력신호들에 대해 가장 큰 음량의 세션을 기준으로 음량을 균일화하여 게인값을 적용함으로써 실질적인 음량균일화가 이루어지지 않는 기존의 오디오 믹서에서 발생하는 문제점을 해결하여, 음량균일화를 이루고 음량포화를 방지한다.

각 채널의 출력에 적절한 음성 패킷의 합성을 위한 믹싱콘트롤부에서는 N개의 채널에서 음량이 균일화된 음성 패킷을 전달받아, 각 채널의 출력에 적절한 음성패킷들을 믹싱 후 인코딩하여 각 참여자에게 전송한다.

$$y_i^j(n) = \sum_{k=1, k \neq j}^N x_i^k(n) \cdot Gain_i^k. \quad (1)$$

여기서  $i$ 는 패킷 번호,  $n$ 은 패킷 내의 샘플 번호,  $j$ 는 출력채널의 번호,  $k$ 는 믹싱할 채널의 번호,  $N$ 은 오디오믹서의 총 채널의 수이고,  $y_i^j(n)$ 는 믹싱된 출력 신호,  $x_i^k(n)$ 은 각 채널의 입력 신호,  $Gain_i^k$ 는 각 채널에서 계산된 음성 게인이다.

각 사용자는 오디오 믹서에 자신이 전송한 음성 패킷을 다시 자신이 청취할 필요가 없기 때문에, 자신의 채널을 제외한 나머지 채널들의 음성패킷들이 합쳐진 음성 스트림을 오디오 믹서로부터 전달받는다.

## III. 음성구간검출

기존의 오디오 믹서들은 입력되는 음성신호의 모든 구간에서 음량을 추정하기 때문에, 음성신호가 존재하지 않는 음량추정이 불필요한 구간에서의 부정확한 음량추정으로 인해 음량불균형이 해소되지 않았다.<sup>[2]</sup> 본 논문에서는 음성구간검출을 통해서 찾아진 사용자가 실제로 발생하는 구간인 연속되는 음성구간에서 기준음량으로 균일화시키기 위한 게인값을 구하기 때문에 잡음레벨의 상승으로 인한 클리핑현상을 방지한다. 또한, 연속된 음성구간단위로 게인값을 적용함으로써 급격한 게인값 변동을 방지하여 안정적인 음량을 유지할 수 있다. 하지만 다자

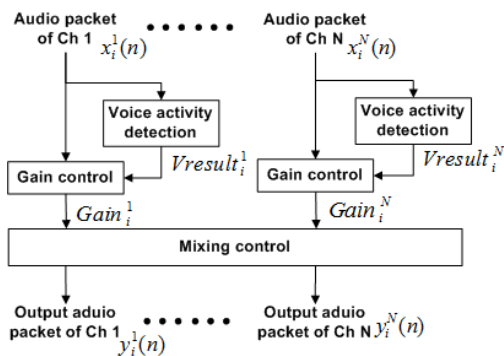


Fig. 1. Structure of audio volume equalization system.

간 음성통화 및 원격회의 시, 각 채널의 사용자들은 다양한 잡음환경에 노출되어있다. 기존의 음성구간 검출 방식들은 고정된 문턱값을 사용하기 때문에 예측할 수 없는 다양한 잡음환경과 잡음레벨의 변동에 적응하지 못하여 음성구간검출 정확도가 저하되는 문제를 보인다.<sup>[3]</sup> 이러한 잡음의 영향을 줄이기 위해서 잡음제거를 동반한 음성구간검출이 사용되거나, 가우시안 분포를 이용한 음성구간검출<sup>[6]</sup>이 사용되지만, 연산량의 증가로 인해 실시간 통화가 어려워지는 단점이 존재한다. 본 논문에서는 다양한 잡음 환경 및 잡음레벨이 변화하는 현재의 상황에 빠르게 적응하여 문턱값을 갱신시킴으로써 더 효과적으로 음성구간과 비음성구간을 검출할 수 있는 음성구간 검출방법을 제안한다.

Fig. 2는 본 논문에서 제안하는 음성구간검출기의 구조를 나타낸다.

입력된 음성 패킷은 저주파 대역에 집중되어있는 잡음에 대한 영향을 줄이기 위해 고주파통과필터를 통과시키고, 출력된 음성 신호를 이용해서 에너지  $E_i$  와 Auto-correlation(AC)  $AC_i$ 을 계산한다. 현재 입력된 음성 패킷에서 계산된 에너지와 AC가 각각의 설정된 문턱값  $\Psi_i^E, \Theta_i^A$ 보다 작은 경우에는 비음성구간으로 판별한다. 나머지 경우에 대해서는 모두 음성구간으로 판별한다.

$$Vresult_i = \begin{cases} \text{비음성구간,} & \text{if } E_i < \Psi_i^E \text{ AND } AC_i < \Theta_i^A \\ \text{음성구간,} & \text{otherwise} \end{cases} \quad (2)$$

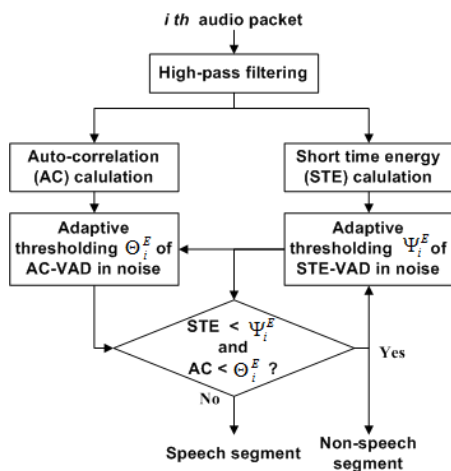


Fig. 2. Block diagram of voice activity detection.

비음성구간으로 판별된 경우, 현재의 잡음상황에 대처하기 위해서 적응적으로 에너지와 AC의 문턱값  $\Psi_i^E, \Theta_i^A$ 을 갱신한다.

문턱값을 갱신하는 과정은 Fig. 3과 같다. 잡음상황의 변화를 추정하기 위해서 Recursive-averaging 방법을 통해 비음성구간의 평균에너지  $E_i^{sil}$ 와 평균AC  $AC_i^{sil}$ 을 계산한다. 계산된 평균에너지와 평균AC는 이전 패킷의 평균에너지와 평균AC의 비교를 통해서 잡음상황의 변화를 추정한다.  $i$ 는 패킷 번호,  $\beta_E$  ( $0 < \beta_E < 1$ )는 에너지의 스무딩 상수,  $\beta_A$  ( $0 < \beta_A < 1$ )는 AC의 스무딩 상수이다.

에너지 상승구간의 카운팅 수  $cnt_{up}^E$ 이 상승구간의 카운팅 문턱값  $N_{up}$ 보다 크면, 잡음의 에너지가 상승중이라 판단하고, 이전의 평균에너지보다 상승된 현재의 평균에너지  $E_i^{sil}$ 에 가중치  $\alpha_E$ 를 적용하여 에너지문턱값  $\Psi_i^E$ 를 갱신한다. 그리고 하강구간의 카운팅 수  $cnt_{down}^E$ 가 하강구간의 카운팅 문턱값  $N_{down}$ 보다 크면, 잡음의 에너지가 감소되고 있다고 판단하여 이전의 평균에너지보다 감소된 현재의 평균에너지  $E_i^{sil}$ 에 가중치  $\alpha_E$ 를 적용하여 에너지의 문턱값  $\Psi_i^E$ 를 갱신한다. 여기서  $\alpha_E$  ( $5 < \alpha_E < 9$ )는 에너지 문턱값의 가중치이고,  $\alpha_A$  ( $9 < \alpha_A < 12$ )는 AC 문턱값의 가중치이며,  $N_{up}$  ( $N_{up} = 3$ )은 상승구간의 카운팅 문턱값,  $N_{down}$  ( $N_{down} = 7$ )은 하강구간의 카운팅 문턱값이다.

If ( $Vresult_i = \text{비음성구간}$ )

$$E_i^{sil} = \beta_E \cdot E_{i-1}^{sil} + (1 - \beta_E) \cdot E_i$$

If ( $E_i^{sil} > E_{i-1}^{sil}$ )

$$\Psi_i^E = \begin{cases} \beta_{up} \cdot \Psi_{i-1}^E + (1 - \beta_{up}) \cdot \alpha_E \cdot E_i^{sil}, & \text{if } cnt_{up}^E > N_{up} \\ \Psi_{i-1}^E, & \text{otherwise} \end{cases}$$

else

$$\Psi_i^E = \begin{cases} \beta_{down} \cdot \Psi_{i-1}^E + (1 - \beta_{down}) \cdot \alpha_E \cdot E_i^{sil}, & \text{if } cnt_{down}^E > N_{down} \\ \Psi_{i-1}^E, & \text{otherwise} \end{cases}$$

$$AC_i^{sil} = \beta_A \cdot AC_{i-1}^{sil} + (1 - \beta_A) \cdot AC_i$$

If ( $AC_i^{sil} > AC_{i-1}^{sil}$ )

$$\Theta_i^A = \begin{cases} \beta_{up} \cdot \Theta_{i-1}^A + (1 - \beta_{up}) \cdot \alpha_A \cdot AC_i^{sil}, & \text{if } cnt_{up}^A > N_{up} \\ \Theta_{i-1}^A, & \text{otherwise} \end{cases}$$

else

$$\Theta_i^A = \begin{cases} \alpha_A \cdot AC_i^{sil}, & \text{if } cnt_{down}^A > N_{down} \\ \Theta_{i-1}^A, & \text{otherwise} \end{cases}$$

Fig. 3. Adaptive thresholding approach in noisy environment.

AC 상승구간의 카운팅 수  $cnt_{up}^A$  이 상승구간의 카운팅 문턱값  $N_{up}$  이상일 때, 잡음의 AC가 상승하는 추세라 판단하고 이전의 평균AC보다 상승된 현재의 평균 AC  $AC_i^{sil}$ 에 가중치  $\alpha_A$ 를 적용해서 AC의 문턱값  $\theta_i^A$ 를 갱신한다. 그리고 AC 하강구간의 카운팅 수  $cnt_{down}^A$ 이 하강구간의 카운팅 문턱값  $N_{down}$  이상일 때, 잡음의 AC가 하강하는 추세라 판단하고, 이전의 평균AC보다 감소된 현재의 평균AC에 가중치  $\alpha_A$ 를 적용해서 AC의 문턱값  $\theta_i^A$ 를 갱신한다.

갱신된 각각의 문턱값은 다음에 입력되는 음성 패킷의 음성구간검출에 사용하기 위해 사용되며, 판별된 음성구간검출 결과는 게인콘트롤부로 전달된다.

#### IV. 게인 콘트롤

기존의 게인콘트롤은 각 채널별로 정확한 음량추정을 통한 게인을 계산하지 않고, 모든 채널의 음성 패킷을 믹싱 후 추정된 게인을 적용했기 때문에 음량불균형이 해소되지 않는 문제점이 있었다.<sup>[4-5]</sup>

본 논문에서는 입력되는 각 채널의 음성신호들을 기준음량에 맞추는 게인콘트롤을 사용한다. 각 채널의 모든 음성신호들을 기준음량에 맞추어 게인을 조정하게 되면 빈번한 게인값 변동과 무리한 음량조절로 인해서 음성신호 간의 불연속적인 잡음이 발생하거나, 음성신호파형에 왜곡이 발생하게 되는데, 이러한 문제를 해결하기 위해서 본 논문에서 Fig. 4의 구조를 갖는 게인콘트롤 방식을 제안한다.

먼저, 입력된 음성패킷의 크기를 나타내는 음성패킷 내 가장 큰샘플값인  $Peak_i$ 를 다음 식(3)과 같이 각 샘플값의 대소비교를 통해서 찾는다.

$$Peak_i = \text{Max} [x_i(0), \dots, x_i(n)], n = 0, \dots, M \quad (3)$$

여기서  $i$ 는 패킷 번호,  $n$ 은 패킷 내의 샘플번호,  $M$ 은 패킷 내의 샘플 수이다.

음성구간검출결과를 기반으로 연속된 음성구간의 음량크기를 나타내는 음성신호의 정점  $MaxPeak_i$ 를 식(4)를 통해 검출한다.  $MaxPeak_i$ 는 연속된 음성구간 내에서 각 음성패킷의 크기를 나타내는  $Peak_i$  중에서 최대값을 의미한다.

$$MaxPeak_i = \begin{cases} Peak_i & , \text{ if } Peak_i > MaxPeak_{i-1} \\ MaxPeak_{i-1} & , \text{ otherwise} \end{cases} \quad (4)$$

연속된 음성구간의 음량크기를 나타내는 음성신호의 정점  $MaxPeak_i$ 와 기준음량  $Basic Volume$ 의 비율로 기본게인값을 계산한다.

$$Gain_i^{Basic} = \frac{Basic Volume}{MaxPeak_i} \quad (5)$$

여기서, 기준음량  $Basic Volume = 15000$ 으로, 이는 음성신호의 표현범위 절반에 해당하는 값이다. 음량포화를 방지하고 안정적으로 음량을 균일화하기 위해 중간대역의 음량을 기준으로 게인값을 계산한다. 찾아진  $MaxPeak_i$ 를 기준음량에 맞추어 음량을 균일화시킴으로써 음량의 포화를 방지하기 위함이다. 그리고 연속된 음성구간을 한 문장으로 설정하여 게인값을 구하고, 다음 연속된 음성구간에 동일한 게인값을 적용함으로써 빈번한 게인값 변동으로 인한 음성신호간의 불연속점과 음성신호의 왜곡을 방지한다.

기준음량에 맞추는 게인값을 계산하더라도 음성신호에 적용 시, 음량포화의 상황이 발생하기 때문에, 이를 방지하기 위해서 음량포화방지를 위한 게인값을 식(6)과 같이 계산한다.

$$\begin{aligned} & \text{If} (Gain_i^{basic} \cdot Peak_i > Volume\_Limit) \\ & \quad Gain_i = \frac{Volume\_Limit}{Peak_i} \\ & \text{else} \\ & \quad Gain_i = Gain_i^{basic} \end{aligned} \quad (6)$$

여기서  $Volume\_Limit = 32767$ 로, 음성신호 표현범

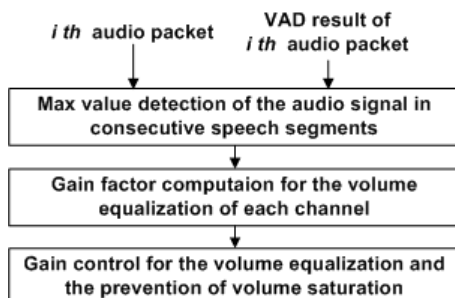


Fig. 4. Block diagram of audio gain control.

위의 최대값이고,  $Peak_i$ 는 현재 입력된 음성패킷의 크기이다.

현재 음성패킷에 기본게인값을 적용 시, 음량한계점보다 작으면 게인값을 적용해도 음량포화가 발생하지 않는다고 판단하여, 기본게인값을 현재 음성패킷에 적용하기 위한 게인값으로 결정한다. 반대로 현재 음성패킷에 게인값 적용 시 음량한계점을 초과하면 음량포화가 발생하는 상황으로 판단하여, 음량한계점과 현재 입력된 음성패킷의 크기의 비율로 음량포화방지를 위한 게인값을 계산한다.

최종적으로 게인컨트롤되어 음량이 균일화된 각 채널의 음성 패킷은 믹싱컨트롤부로 입력되어 각 채널에 해당하는 출력 음성 패킷으로 합성된다.

### VI. 실험결과

본 논문에서 제안하는 오디오 믹서 알고리즘의 성능을 측정하기 위해 다양한 잡음환경 및 다양한 음량 크기를 가지는 10가지 실험군을 구성하였다. 각 실험군은 16 kHz 샘플링레이트로 녹음되었으며, 성인 남성 2명과 여성 1명이 한국어로 대화하는 10가지 상황으로 구성하였다.

제안한 방식의 성능 비교평가를 위해, 출력신호들에 대한 스무스한 게인을 적용하는 믹싱 방식인 Enhanced Align-to-Greatest Weighted 방식<sup>[4]</sup>(M1), 각 채널을 개별적으로 게인컨트롤 한 후 믹싱하는 방식인 Automatic Gain Controller and Adjustment Rate 방식<sup>[5]</sup>(M2), 그리고 입력된 음성신호에 대해서 잡음제거와 음성구간검출<sup>[7]</sup>을 수행 후 믹싱을 하는 방식인 Enhanced Stream Regulation and Mixing 방식<sup>[4]</sup>(M3)을 적용하여 PESQ, Power LOSS, 출력잡음레벨, 처리속도 등의 성능을 측정하였다.

PESQ는 객관적인 음질평가 지표로서, 1~5의 점수로 표현되며 수치가 높을수록 좋은 음질인 경우이다. 게인컨트롤 과정이 없는 일반믹싱 TM(True Mixing)을 기준으로 각 믹서별 출력결과의 PESQ 측정을 통해서 믹싱방식에 따른 음질 훼손정도를 실험하였다. Fig. 5를 보면, 10개 실험군에 대한 각 믹서별 PESQ 평균을 기재하였다. 제안된 방식이 기존의 믹서방식보다 높은 PESQ를 보이고 있어, 음성신호파형의 왜곡이 적고

음질이 좋다는 것을 알 수 있다.

다음으로, Power Loss는 입력된 음성신호가 믹서를 거쳐서 적절한 음량조절이 이루어졌는지를 나타낸다. 믹서에서 불필요한 음량의 감소나 증가 없이 음량 균일화가 수행되는 것을 확인하기 위해 PowerLoss를 측정하였다. 초기 입력된 음성신호들의 음량을 기준으로 하기 위해서 TM의 음성신호를 기준으로 각 Power Loss를 측정하였다. Power Loss는 수치가 0에 가까울수록 불필요한 음량의 감소 및 증가가 없다는 의미이다. Fig. 5의 Power Loss를 측정한 결과를 보면, 제안된 방식은 출력 음성신호가 기존의 방식들에 비해 적은 Power Loss를 보였다. 게인컨트롤을 하더라도 원래의 음량에서 불필요한 음량감소 및 증가가 없다는 것을 의미한다.

Output noise level에 대한 실험은 -19 dB의 백색잡음과 -38 dB의 핑크잡음이 입력될 때, 각 믹서별로 출력결과의 잡음레벨을 측정한 실험이다. Fig. 6의 Output noise level에 대한 실험결과를 보면, M3는 잡음제거가 선행되기 때문에 출력잡음레벨이 낮다. M1과 M2는 잡음이 입력되었을 때, 잡음에서도 게인컨트롤을 해서 잡음레벨에 변동이 있다. 하지만 제안된 방식은 잡음만 존재하는 구간에서는 게인컨트롤을 하지 않기 때문에, 각 잡음을 합성한 잡음레벨인 -18.8 dB가 측정되

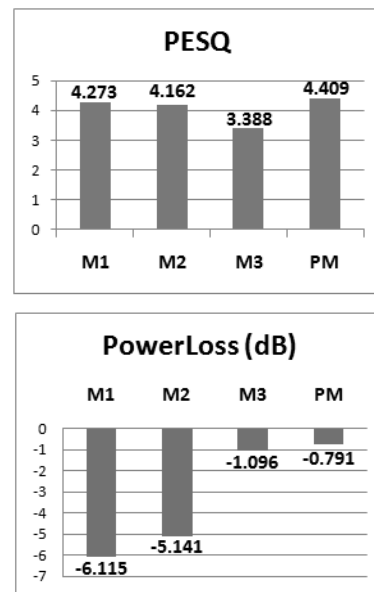


Fig. 5. PESQ and PowerLoss result comparisons of audio mixer algorithms.

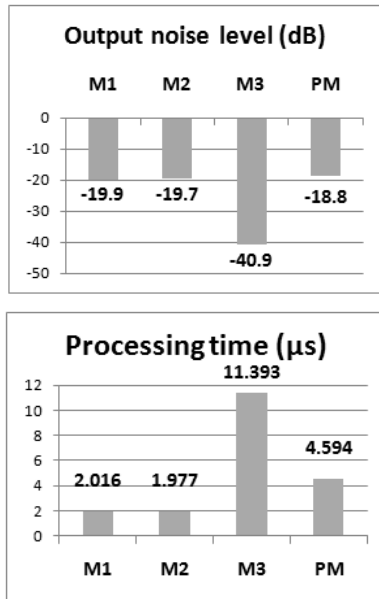


Fig. 6. Output noise level and processing time result comparisons of audio mixer algorithms.

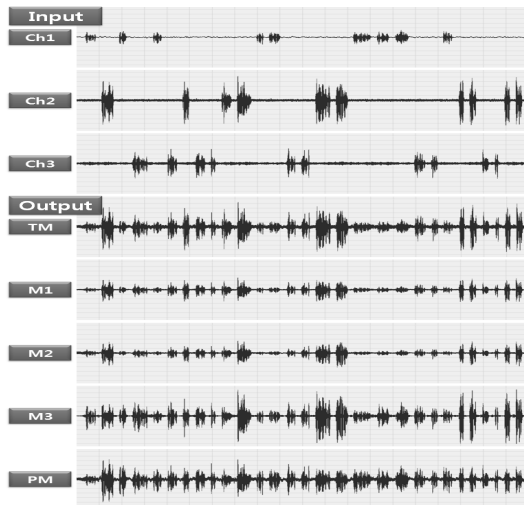


Fig. 7. Comparison of the mixing output signals to the input signals.

었다. 이는 입력된 두 잡음을 게인콘트롤 없이 믹싱한 결과와 같으며, 불필요한 게인콘트롤을 방지하여 연산량을 감소시켰다는 것을 알 수 있다.

마지막으로, 처리속도의 실험결과는 각 믹서별로 20 ms의 1패킷 단위 처리속도를 나타낸다. Fig. 6의 Processing time에 대한 실험 결과를 보면, 제안된 방식은 간단한 연산만을 수행하는 M1과 M2보다는 처리시간이 길지만, 잡음제거가 선행되어 매우 긴 처리시

간을 가지는 M3에 비해 불필요한 연산을 방지하여 처리시간을 줄였다.

그리고 Fig. 7은 3명의 화자로부터 음성신호가 입력되는 상황에서 각 방식의 출력신호를 비교하였으며, 비교를 위해 TM의 결과를 추가하였다. Fig. 7에서와 같이 각 채널로부터 입력되는 음성신호의 크기에 차이가 있는 상황에서는 기존의 믹서방식들은 각 문장들이 균일화되지 않은 채로 출력된 것을 확인할 수 있다. 반면, 제안된 방식은 각 채널의 음량추정을 위한 초기 문장을 제외한 이후의 음성신호에 게인을 적용하여 음량이 균일화된 것을 확인할 수 있다. 이는 게인을 조정하더라도 음성신호의 Power Loss가 적으며 음질 저하를 방지하여 높은 PESQ를 유지하는 오디오 믹서라 할 수 있다.

## VII. 결 론

본 논문에서는 다자간 음성통화 품질 향상을 위한 오디오 믹서 알고리즘을 제안하였다. 제안된 방식은 잡음환경에서도 음성구간 검출의 정확도를 향상시켜 오디오 믹싱 시 음량포화를 방지하고 음량균일화를 효과적으로 수행하였다. PESQ, Power Loss, 출력잡음레벨, 처리속도 등의 성능측정에 있어서 제안된 오디오 믹서 알고리즘은 기존방식과 비교하여 우수한 성능을 제시함을 알 수 있었다. 본 연구를 통해 개발된 오디오 믹서 알고리즘은 향후 실재감 혹은 현장감을 향상시키는 다자간 몰입형 화상통화에 적용할 수 있으리라고 생각된다.

## 감사의 글

본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음(NIPA-2013-H0301-13-4005).

## References

1. D. Song, Y. Mo, and F. Wang, "Architecture of multiparty conferencing using SIP," in Proc. IEEE Int'l. Conf. on WiCOM, 2, 1361-1364 (2005).

2. F. Xing G, U. Wei-Kang, and Y. Xiu-qing, "Research on fast real time adaptive audio mixing in multimedia conference," J. of Zhejiang Univ. Sci. **6**, 507-512 (2005).
3. S. V. Gerven and F. Xie, "A comparative study of speech detection methods," in Proc. of EUROSPEECH, **3**, 1095-098 (1997).
4. S. P. Chandra, K. M. Senthil, and M. P. P. Bala, "Audio mixer for multi-party conferencing in VoIP," in Proc. IEEE Int'l. Conf. on IMSAA, 1-6 (2009).
5. V. M. Baskaran, and K. Wong "Audio mixer with automatic gain controller for software based multipoint control unit," APCCAS, 164-167 (2010).
6. Y. S. Um, J. H. Chang, and D. K. Kim, "Signal subspace-based vioc activity detection using generalized gaussian distortion" (in Korean), J. Acoust. Soc. Kr. **32**, 131-137 (2013).
7. 3GPP TS 26.194, *AMR Wideband Speech Codec : Voice Activity Detection (VAD)*, 2004.

## 저자 약력

### ▶ 류 상 현(Sang-Hyeon Ryu)



2012년 2월: 광운대학교 전파공학과 공학사  
 2012년 3월 ~ 현재: 광운대학교 전파공학과 석사과정

### ▶ 김 형 국(Hyoung-Gook Kim)



2002년 7월: 독일 SIEMENS/Cortologic AG 책임연구원  
 2005년 3월: 독일 베를린 공과대학교 Assistant Professor  
 2007년 2월: 삼성중합기술원 수석연구원  
 2007년 3월 ~ 현재: 광운대학교 전파공학과 부교수