

일반논문 (Regular Paper)

방송공학회논문지 제23권 제1호, 2018년 1월 (JBE Vol. 23, No. 1, January 2018)

<https://doi.org/10.5909/JBE.2018.23.1.86>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

심층 신경망을 이용한 음성 신호의 부호화 이력 검출

조 효 진^{a)}, 장 원^{a)}, 신 성 현^{a)}, 박 호 중^{a)‡}

Coding History Detection of Speech Signal using Deep Neural Network

Hyo-Jin Cho^{a)}, Won Jang^{a)}, Seong-Hyeon Shin^{a)}, and Hochong Park^{a)‡}

요 약

본 논문에서는 디지털 음성 신호의 부호화 이력을 검출하는 방법을 제안한다. 음성 신호를 디지털 방식으로 전송 또는 저장할 때 데이터양을 줄이기 위해 부호화한다. 따라서 음성 신호 파형이 주어질 때, 해당 신호가 원본인지 부호화된 신호인지 판단하고, 만일 부호화 되었다면 부호화 횟수를 검출하는 부호화 이력 검출 과정이 필요하다. 본 논문에서는 12.2kbps 비트율의 AMR 부호화기에 대하여 원본, 단일 부호화, 이중 부호화 여부를 판단하는 부호화 이력 검출 방법을 제안한다. 제안한 방법은 입력 음성 신호에서 음성 고유의 특성 벡터를 추출하고, 해당 특성 벡터를 심층 신경망으로 모델링 하는 방법을 사용한다. 본 논문에서 제안하는 특성 벡터가 일반적인 스펙트로그램으로부터 추출한 특성 벡터보다 우수한 부호화 이력 검출 성능을 제공하는 것을 확인하였다.

Abstract

In this paper, we propose a method for coding history detection of digital speech signal. In digital speech communication and storage, the signal is encoded to reduce the number of bits. Therefore, when a speech signal waveform is given, we need to detect its coding history so that we can determine whether the signal is an original or an coded one, and if coded, determine the number of times of coding. In this paper, we propose a coding history detection method for 12.2kbps AMR codec in terms of original, single coding, and double coding. The proposed method extracts a speech-specific feature vector from the given speech, and models the feature vector using a deep neural network. We confirm that the proposed feature vector provides better performance in coding history detection than the feature vector computed from the general spectrogram.

Keyword : coding history, feature vector, speech parameter, DNN

a) 광운대학교 전자공학과(Dept. of Electronics Engineering, Kwangwoon University)

‡ Corresponding Author : 박호중(Hochong Park)

E-mail: hcpark@kw.ac.kr

Tel: +82-2-940-5104

ORCID: <http://orcid.org/0000-0003-1600-6610>

※ 본 연구는 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2016R1D1A1B03930923).

· Manuscript received September 8, 2017; Revised October 31, 2017; Accepted October 31, 2017.

1. 서론

음성 신호를 디지털 방식으로 전송 또는 저장할 때 전송 채널 용량 또는 저장 메모리 용량을 줄이기 위해 부호화한다. 따라서 디지털 음성 신호 파형이 주어질 때, 신호의 전송 또는 변형 과정 등을 분석하기 위해 해당 신호가 원본인지 부호화된 신호인지 판단하고, 만일 부호화 되었다면 부호화 횟수를 검출하는 부호화 이력 (coding history) 검출 과정이 필요하다. 예로, 디지털 음성 통신에서 전송 과정에서 불필요한 부호화 과정을 거치면 이중 부호화 (double coding)가 실행되므로, 최종 수신한 음성 신호에서의 이중 부호화 검출을 통하여 통신 경로를 추정할 수 있다. 또한, 원본으로 가정된 음성 신호의 일부 구간에서 부호화 이력이 검출되거나, 한 번에 녹음된 음성 신호에서 원본, 단일 부호화 (single coding), 이중 부호화 구간이 동시에 나타나면 해당 음성 신호는 변형 또는 편집되어 연결된 것을 의미하므로, 부호화 이력 검출을 통하여 신호의 변형 또는 편집 여부를 추정할 수 있다.

오디오 신호의 부호화 이력을 검출하는 연구가 널리 진행되었고, 오디오 부호화기는 스펙트럼 정보를 양자화하므로 스펙트럼 정보의 왜곡 특성을 분석하여 부호화 이력을 검출한다. 예로, 고대역 스펙트럼을 분석하여 MP3 부호화의 비트율과 부호화 횟수를 검출하는 방법^[1,2], MDCT (modified discrete cosine transform) 계수를 분석하여 AAC (advanced audio coding) 부호화 횟수를 검출하는 방법^[3,4] 등이 개발되었다. 그러나 음성 부호화와 오디오 부호화는 서로 다른 방법으로 진행되고 부호화에 의한 신호 변형 특성에 차이가 있으므로 오디오 부호화 이력 검출 방법을 음성 신호에 그대로 적용할 수 없다. 최근 음성 신호가 AMR (adaptive multi-rate) 부호화기로 부호화될 때, 부호화 이력을 검출하는 방법이 개발되었다^[5]. 그러나 이 방법은 신경망을 이용하여 자동으로 특성을 추출하므로 추출된 특성이 구체적 의미를 갖지 못하여 부호화 이력 검출에 대한 이론적 설명과 분석을 제공하지 못하고, 또한 원본 검출을 지원하지 않는 한계를 가진다. 본 논문에서는 음성 파라미터를 기반으로 특성을 추출하여 12.2kbps 비트율의 AMR에 대하여 원본, 단일 부호화, 이중 부호화를 검출하는 새로운 방법을 제안한다.

본 논문에서 제안하는 방법은 입력 신호에서 음성 분석에 특화된 특성 벡터를 추출하고, 이를 분석하여 부호화 이력을 검출한다. 먼저, 20ms 길이의 프레임 단위로 다양한 음성 파라미터를 추출하고, 4초 길이의 텍스처 (texture) 프레임 동안 각 음성 파라미터의 변화 특성을 나타내는 통계 값을 구하여 텍스처 프레임 단위로 특성 벡터를 정의한다. 다음, 특성 벡터를 심층 신경망으로 모델링 하여 부호화 이력을 최종 결정한다. 제안하는 방법은 음성 파라미터 기반으로 동작하므로 부호화에 의한 음성 왜곡을 구체적으로 모델링 할 수 있는 장점을 제공한다. 제안하는 음성 파라미터 기반의 특성 벡터가 오디오 부호화 이력 검출에 사용되는 일반적인 스펙트럼 기반의 특성 벡터에 비하여 우수한 부호화 이력 검출 성능을 제공하는 것을 확인하였다.

II. 제안하는 부호화 이력 검출 방법

1. 개요

CELP (code excited linear prediction) 구조의 음성 부호화기는 음성 모델을 기반으로 음성 파라미터를 추출하고 양자화하는 구조를 사용한다. 따라서 음성 신호가 부호화되면 음성 특성을 모델링 하는 음성 파라미터에 왜곡이 나타나고, 만일 이중 부호화가 실시되면 해당 파라미터의 왜곡은 더 증가하게 된다. 따라서 음성 파라미터의 왜곡 특성을 분석하면 음성 부호화 이력 검출을 위한 중요한 정보를 추출할 수 있다. 또한, CELP 음성 부호화기는 LPC (linear predictive coding) 잔차 (residual) 신호에 대한 파형 양자화를 실시하므로, LPC 잔차 신호의 성질을 분석하면 부호화 이력에 대한 정보를 추가적으로 얻을 수 있다. 이와 같은 배경에 따라, 제안 방법은 입력 신호와 입력 신호의 LPC 잔차 신호에 대하여 각각 20ms 프레임 단위로 다양한 음성 파라미터를 추출하여 특성 벡터를 구한다.

음성 신호의 특성은 시간에 따라 변하고 그에 따라 부호화에 의한 왜곡 형태도 시간에 따라 변한다. 따라서 한 프레임에 대한 음성 파라미터만으로 부호화 이력을 검출하면 정확도가 크게 저하된다. 제안 방법은 시간에 따른 특성 변화를 종합적으로 분석하여 성능을 향상시키기 위하여 4초

길이의 텍스처 프레임을 정의하고, 텍스처 프레임 내에서 음성 파라미터의 변화를 통계적으로 분석하여 최종 특성 벡터를 정의한다.

제안하는 부호화 이력 검출 방법은 심층 신경망을 사용하여 특성 벡터를 모델링 한다. 제공되는 훈련 데이터에서 텍스처 프레임 단위로 특성 벡터를 구하고, 최적의 검출 동작을 수행하도록 심층 신경망을 훈련하여 훈련된 신경망을 구한다. 다음, 실제 입력 음성 신호가 주어지면 텍스처 프레임 단위로 특성 벡터를 추출하고 훈련된 신경망에 따라 특성 벡터를 분석하여 원본, 단일 부호화, 이중 부호화 중 하나로 분류한다. 즉, 제안하는 부호화 이력 검출은 텍스처 프레임 단위로 동작한다.

2. 프레임 단위 음성 파라미터 추출

20ms 프레임 단위로 음성 파라미터를 추출할 때, 부호화에 따라 큰 왜곡이 발생하는 음성 파라미터를 선택해야 부호화 이력 검출을 위한 핵심 정보를 얻을 수 있다. 이를 위해 부호화에 의한 음성 파라미터 특징 변화를 분석하였다. 그림 1은 원본, AMR 단일 부호화, AMR 이중 부호화한 신호의 스펙트럼 포락선의 예를 보여주며, 부호화에 의하여 스펙트럼 포락선에 차이가 나타나는 것을 알 수 있다. 따라서 음성의 스펙트럼 포락선 정보를 나타내는 LPC 계수 10개와 MFCC (Mel-frequency cepstral coefficient) 12개를 구하여 스펙트럼 포락선의 특징을 분석한다. 그림 2는 음성 신호의 부호화 횟수에 따른 LPC 계수와 MFCC 값의

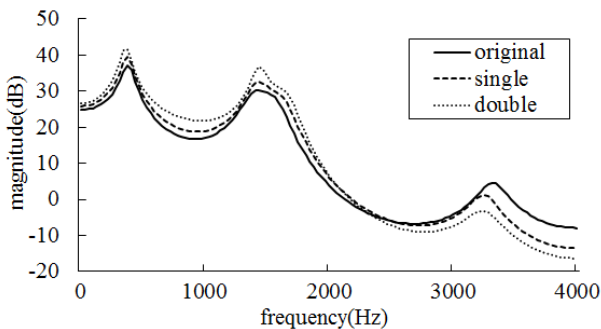


그림 1. 원본, AMR 단일 부호화, AMR 이중 부호화한 신호의 스펙트럼 포락선의 예

Fig. 1. Example of spectral envelopes of original, AMR single coded, and AMR double coded signals

평균 분포를 보여준다. LPC 계수와 MFCC가 부호화 횟수에 따라 서로 다른 분포 특성을 가지고 부호화 횟수에 대한 변별력을 가지고 있는 것을 알 수 있다.

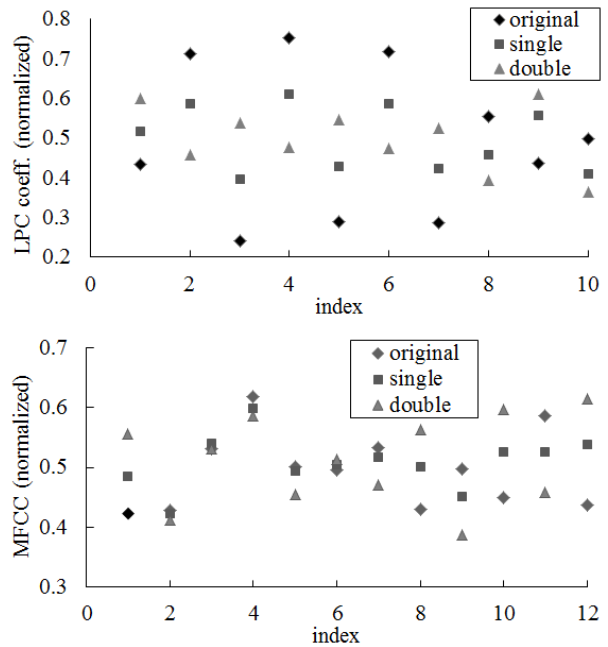


그림 2. 원본, AMR 단일 부호화, AMR 이중 부호화한 신호의 LPC 계수와 MFCC의 평균 분포 (위) LPC 계수, (아래) MFCC

Fig. 2. The average distribution of LPC coefficients and MFCC of original, AMR single coded, and AMR double coded signals (Top) LPC coefficients, (Bottom) MFCC

음성 부호화에 의하여 특히 고대역 정보에 많은 왜곡이 발생하므로, 해당 왜곡을 간단하게 분석하기 위하여 프레

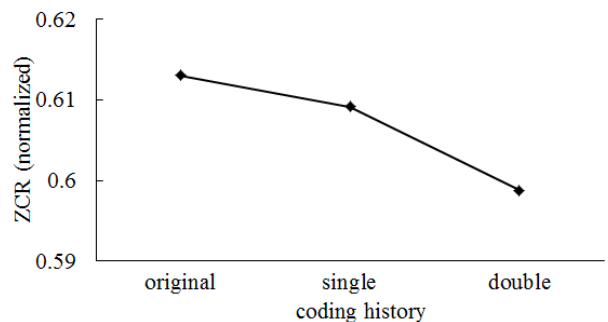


그림 3. 원본, AMR 단일 부호화, AMR 이중 부호화한 신호의 평균 영 교차율

Fig. 3. The average zero crossing rate of original, AMR single coded, and AMR double coded signals

임 단위로 영 교차율 (zero crossing rate, ZCR) 1개를 구한다. 그림 3은 부호화 횟수에 따른 평균 ZCR을 보여준다. 부호화 횟수에 따라 ZCR에 큰 차이를 가지고, 따라서 ZCR은 부호화 이력 검출에 큰 역할을 할 것이다. 이상과 같은 분석 결과에 따라 입력 신호로부터 20ms 프레임 단위로 LPC 계수 10개, MFCC 12개, ZCR 1개 등 23개 음성 파라미터를 구한다.

음성 신호의 LPC 잔차 신호는 이론적으로 평탄한 스펙트럼 포락선을 가진다. 그러나 LPC 잔차 신호의 양자화 과정에서 많은 왜곡이 발생하므로 실제 복원된 신호의 LPC 잔차 신호의 스펙트럼 포락선은 평탄하지 않게 된다. 따라서 LPC 잔차 신호의 스펙트럼 포락선의 왜곡 특성을 분석하면 부호화 이력에 대한 추가 정보를 추출할 수 있다. 이를 위해 프레임 단위로 입력 신호의 LPC 잔차 신호를 구하고, LPC 잔차 신호의 LPC 계수 10개, MFCC 12개, ZCR 1개 등 23개 음성 파라미터를 추가로 구한다. 이와 같은 과정으로 20ms 프레임 단위로 입력 신호와 입력 신호의 LPC 잔차 신호로부터 총 46개 음성 파라미터를 구한다.

3. 텍스처 프레임 단위 특성 벡터 추출

그림 4는 20ms 프레임 단위로 구한 46개 음성 파라미터를 기반으로 4초 텍스처 프레임 단위의 최종 특성 벡터를 구하는 과정을 보여준다. 프레임 단위로 구한 46개 음성 파라미터를 46차 벡터로 표시하고, 텍스처 프레임에 포함된 모든 프레임의 음성 파라미터를 그림 4와 같이 46×200

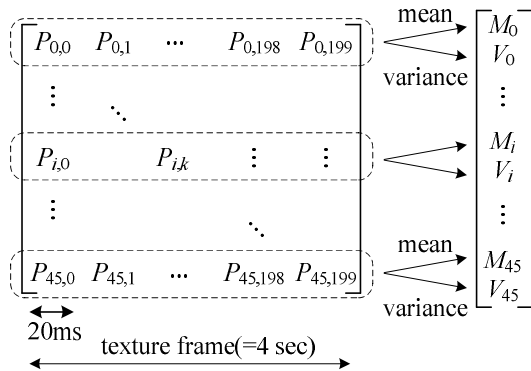


그림 4. 음성 파라미터로부터 특성 벡터를 구하는 구조
 Fig. 4. Structure of determining feature vector from speech parameters

행렬로 표시한다. 여기서, 각 요소를 $P_{i,k}$ 로 표시하고, $0 \leq i \leq 45$ 는 각 음성 파라미터를 나타내는 인덱스이고, $0 \leq k \leq 199$ 은 텍스처 프레임에 포함된 프레임 인덱스이다. 다음, 행렬의 각 행 (row)에 대한 평균 M_i 와 분산 V_i 를 각각 구하여 이를 92차 벡터 Y_i 로 표시하고, t 는 텍스처 프레임 인덱스이다. 이와 같이 구한 Y_i 는 4초 텍스처 프레임 내에서 음성 파라미터의 단구간 (short term) 특성을 나타내는 정보이다.

음성 신호의 왜곡을 더 상세하게 분석하기 위하여 텍스처 프레임보다 더 긴 구간에서의 특성 분석이 필요하다. 이를 위해 앞에서 구한 Y_i 에 대하여 이웃 텍스처 프레임 사이의 변화량 $Y_i - Y_{i-1}$ 을 구한다. 이렇게 구한 $Y_i - Y_{i-1}$ 은 신호의 장구간 (long term) 특성 정보를 제공한다.

제안하는 부호화 이력 검출 방법은 위의 과정을 통하여 각 텍스처 프레임별로 Y_i 를 구하여 $X_i = [Y_i^T (Y_i - Y_{i-1})^T]^T$ 로 정의되는 184차 특성 벡터 X_i 를 추출하고, 이를 심층 신경망으로 모델링 하여 부호화 이력 검출 동작을 수행한다. 제안 방법은 입력 신호와 입력 신호의 LPC 잔차 신호의 특성을 모두 분석하고, 시간 축에서 단구간 특성과 장구간 특성을 동시에 분석하여 부호화 이력 검출의 성능을 향상시킨다.

4. 심층 신경망

제안하는 부호화 이력 검출 방법은 심층 신경망을 사용하여 특성 벡터를 모델링 한다. 심층 신경망의 매개변수에는 신경망 구조를 결정하는 매개변수와 훈련 과정을 결정하는 매개변수가 있다. 심층 신경망에 사용할 최적의 매개변수를 구하는 이론적인 방법은 없으며, 반복적인 실험을 통하여 최적의 값들을 찾아야 한다. 본 논문에서는 다양한 조합의 매개변수 값에 대하여 실험을 진행하고 가장 우수한 성능을 제공하는 경우를 찾아 최적 매개변수 값들을 결정하였다.

신경망의 은닉층은 2개, 각 은닉층의 뉴런 개수는 100개와 50개로 결정하였고, 3가지 부호화 이력으로 분류하므로 출력층의 뉴런 수는 3개로 한다. 심층 신경망의 활성화 함수로 ReLU^[6] (rectified linear unit)를 사용하고, 출력층에는 softmax^[6] 함수를 적용하여 출력층 뉴런 값은 각 부호화 이

력에 대한 확률을 나타낸다. 비용 함수는 교차 엔트로피^[6] (cross-entropy) 함수를 사용한다. 훈련 과정에서 Xavier's initialization^[7]을 사용해 각 뉴런의 가중치를 초기화하고, batch normalization^[8]을 각 층에 적용하여 가중치 및 바이어스의 정규화를 수행하고, 이 동작을 통하여 학습 속도와 훈련 성능을 향상시켰다. 훈련 과정은 총 100회 epoch를 수행하였다.

III. 성능 분석

본 논문에서는 TIMIT 음성 DB를 사용하여 부호화 이력 검출 성능을 측정하였다. TIMIT DB 음성 신호를 8kHz로 다운 샘플링한 후 12.2bps AMR 부호화기를 이용하여 단일 부호화와 이중 부호화를 실시하여 원음, 단일 부호화, 이중 부호화된 음성 PCM 파형을 각각 구한다. 여기서, 각 신호의 길이는 약 4시간 13분이다. 이중 부호화를 구현할 때, 첫 번째 부호화와 두 번째 부호화의 AMR 프레임 오프셋 (offset)을 무작위로 설정하여 프레임 오프셋에 대한 조건을 제거하였다. 원본, 단일 부호화, 이중 부호화된 신호 각각에 대하여 4초 단위로 특성 벡터를 추출하고, 이때 음성 파라미터를 추출하는 프레임과 AMR 프레임 경계가 무작위 관계를 가지도록 설정하였다. 각 부호화 이력별 특성 벡터 중에서 무작위로 90%를 선택하여 신경망을 훈련시키고, 나머지 10%는 실험 (test) 데이터로 사용한다. 이와 같은 방법으로 서로 다른 데이터에 대한 훈련과 실험을 10번 진행하는 10-fold cross validation을 실시하여 최종 평균 성능을 측정하였다.

표 1은 제안하는 부호화 이력 검출 방법의 성능을 보여준다. 원본 신호가 입력되면 모두 원본 신호로 판정하여

100% 정확도를 가지고, 원본 신호와 부호화된 신호 사이의 상호 오분류는 없다. 따라서 단일 부호화와 이중 부호화 구분 없이 원음과 부호화 여부만 판단하면 정확도는 100%가 된다. 단일 부호화와 이중 부호화 사이에 상호 오분류는 존재하지만 그 빈도가 매우 낮으며, 모든 부호화 이력에 대한 전체 평균 정확도는 99.1%이다.

제안하는 방법의 성능을 기존 방법과 비교하기 위해 오디오 분석에 널리 사용되는 스펙트로그램 기반의 특성 벡터를 정의하고 성능을 측정하였다^[3,9]. 공정한 비교를 위하여 제안 방법과 동일하게 20ms 프레임 단위로 DFT를 적용하여 각 프레임의 스펙트럼을 구하고, 46개의 멜 스케일 대역 필터와 로그 연산을 적용하여 대역별 필터 출력 46개를 구한다. 다음, 그림 4와 동일하게 200개 프레임을 연결하여 4초 텍스처 프레임을 정의하고, 각 대역에 대하여 평균과 분산을 각각 구하여 92차 벡터를 얻는다. 마지막으로, 제안 방법과 동일하게 이웃 텍스처 프레임 사이의 차이를 구하여, 최종적으로 184차 특성 벡터를 정의한다. 따라서 제안하는 음성 파라미터 기반의 특성 벡터와 기존 스펙트로그램 기반의 특성 벡터를 구하는 프레임 및 텍스처 프레임 구조는 동일하고, 특성 벡터의 차수도 동일하다.

표 2는 스펙트로그램 기반의 특성 벡터를 심층 신경망으로 훈련하여 부호화 이력을 검출한 성능을 보여준다. 여기서, 심층 신경망은 제안한 방법과 동일한 구조와 훈련 방법을 사용하였고, 따라서 표 1과 표 2의 성능 차이는 오로지 특성 벡터의 차이에 의한 것이다. 원본 신호가 입력될 때, 제안한 특성 벡터를 사용하면 원본 신호로 100% 검출하였으나 스펙트로그램 기반의 특성 벡터를 사용하면 검출 정확도가 87.3%로 저하되고, 부호화된 신호와의 오분류가 발생한다. 또한, 단일 부호화와 이중 부호화 신호에 대한 정확도도 많이 저하되고, 전체 평균 정확도는 85.1%가 되어 제

표 1. 제안하는 부호화 이력 검출 방법의 성능 (%)
Table 1. Performance (%) of the proposed coding history detection method

| True \ Estimated | Original | Single | Double | Ave. |
|------------------|----------|--------|--------|------|
| | Original | 100.0 | 0.0 | |
| Single | 0.0 | 98.2 | 1.8 | 99.1 |
| Double | 0.0 | 0.9 | 99.1 | |

표 2. 스펙트로그램 기반의 특성 벡터를 사용할 때의 성능 (%)
Table 2. Performance (%) when using the spectrogram-based feature vector

| True \ Estimated | Original | Single | Double | Ave. |
|------------------|----------|--------|--------|------|
| | Original | 87.3 | 12.7 | |
| Single | 2.5 | 83.8 | 13.7 | 85.1 |
| Double | 0.0 | 15.8 | 84.2 | |

안 방법에 비하여 14.0% 포인트 감소한다. 만일, 원음과 부호화 여부만 판단하면 평균 정확도는 93.03%가 된다.

위의 비교 실험을 통해 제안하는 특성 벡터가 일반적인 스펙트로그램 기반의 특성 벡터에 비하여 우수한 부호화 이력 검출 성능을 제공하는 것을 확인하였다. 따라서 기존 오디오 신호에 적용된 기술을 음성 신호에 그대로 적용할 수 없고, 제안 방법과 같이 음성을 위하여 음성에 특화된 새로운 특성 벡터가 필요하다는 것을 알 수 있다.

제안하는 특성 벡터는 여러 종류의 음성 파라미터를 기반으로 계산된다. 따라서 각 음성 파라미터가 부호화 이력 검출에 의미 있는 역할을 하는지 확인할 필요가 있고, 이를 통해 제안한 특성 벡터에 불필요한 잉여 정보가 없는 것을 검증할 수 있다. 각 음성 파라미터의 역할을 개별적으로 확인하기 위해 전체 특성 벡터에서 해당 파라미터로부터 계산된 성분을 제외한 변형된 특성 벡터를 새로 구하고, 이를 이용하여 심층 신경망을 다시 훈련한 후 부호화 이력 검출 성능을 측정하는 실험을 진행하였다. 이 실험에서 성능이 저하되는 양이 해당 파라미터의 역할에 해당한다.

표 3은 제안하는 184차 특성 벡터에서 각 종류의 음성 파라미터와 관련된 성분을 제외한 특성 벡터로 성능을 측정 한 결과이다. 제안한 184차 특성 벡터에서 MFCC 관련된

값은 총 96개이고, 이를 제외한 88차 특성 벡터만을 사용할 경우 평균 정확도는 97.0%로 저하되어 MFCC 정보가 큰 역할을 하는 것을 알 수 있다. ZCR 관련된 성분은 총 8개이며, 이를 제거하면 성능이 97.9%로 저하된다. ZCR 관련 성분 수는 전체 184개에 대하여 $8/184 = 4.3\%$ 에 불과하지만, 검출 성능을 1.2% 포인트 향상시키는 중요한 역할을 하는 것을 알 수 있다. 텍스처 프레임 사이의 차이 성분 $Y_t - Y_{t-1}$ 을 제거하면 92차 특성 벡터가 되며, 이 경우 성능은 97.4%로 저하되어 장구간 정보가 중요한 역할을 하는 것을 확인할 수 있다. 결론적으로, 제안한 184차 특성 벡터에 포함된 모든 성분이 부호화 이력 검출에 필요한 유효한 정보라는 것을 확인할 수 있다.

IV. 결 론

본 논문에서는 디지털 음성 신호 파형이 주어질 때, 해당 신호가 원본 신호, AMR 단일 부호화된 신호, AMR 이중 부호화된 신호인지 검출하는 방법을 제안하였다. 음성 부호화기 동작을 기반으로 부호화에 의하여 왜곡되는 정보를 분석하였고, 부호화에 의한 왜곡 특성을 가장 잘 표현하는 LPC, MFCC, ZCR 등의 음성 파라미터를 추출한다. 또한, 시간 축에서 해당 파라미터의 특성 변화를 단구간과 장구간에서 동시에 분석하여 최종 184차 특성 벡터를 구하고, 이를 심층 신경망으로 모델링 하여 부호화 이력을 검출한다. 제안한 특성 벡터를 사용하면 기존 스펙트로그램 기반의 특성 벡터를 사용할 때에 비하여 부호화 이력 검출 성능이 향상되는 것을 확인하였다.

참 고 문 헌 (References)

[1] B. D'Alessandro and Y. Q. Shi, "MP3 bit rate quality detection through frequency spectrum analysis," Proc. 11th ACM Workshop on Multimedia and Security, pp. 57-61, 2009.
 [2] T. Bianchi, A. De Rosa, M. Fontani, G. Rocciolo and A. Piva, "Detection and classification of double compressed MP3 audio tracks," Proc. 1st ACM Workshop on Information Hiding and Multimedia Security, pp. 159-164, 2013.
 [3] D. Luo, W. Luo, R. Yang and J. Huang, "Identifying compression history of wave audio and its applications," ACM Trans. on Multimedia

표 3. 일부 특성 벡터 성분을 제거한 후의 부호화 이력 검출 성능 (%)
 Table 3. Performance (%) of coding history detection after eliminating a subset of components in feature vector

| Eliminated feature components | Est. True | Original | Single | Double | Ave. |
|-------------------------------|-----------|----------|--------|--------|------|
| | | | | | |
| MFCC | Original | 100.0 | 0.0 | 0.0 | 97.0 |
| | Single | 0.1 | 93.8 | 6.1 | |
| | Double | 0.0 | 2.9 | 97.1 | |
| LPC coefficients | Original | 99.9 | 0.1 | 0.0 | 97.3 |
| | Single | 0.5 | 94.0 | 5.5 | |
| | Double | 0.0 | 2.0 | 98.0 | |
| LPC residual | Original | 99.4 | 0.6 | 0.0 | 97.3 |
| | Single | 0.8 | 94.8 | 4.4 | |
| | Double | 0.0 | 2.2 | 97.8 | |
| Texture frame difference | Original | 99.7 | 0.2 | 0.1 | 97.4 |
| | Single | 0.2 | 94.7 | 5.1 | |
| | Double | 0.1 | 2.1 | 97.8 | |
| ZCR | Original | 99.8 | 0.2 | 0.0 | 97.9 |
| | Single | 0.7 | 96.5 | 2.8 | |
| | Double | 0.0 | 2.7 | 97.3 | |

- Computing, Communications, and Applications, vol. 10, no. 3, pp. 30:1-30:19, 2014.
- [4] D. Seichter, L. Cuccovillo and P. Aichroth, "AAC encoding detection and bitrate estimation using a convolutional neural network," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 2069-2073, 2016.
- [5] D. Luo, R. Yang, B. Li and J. Huang, "Detection of Double Compressed AMR Audio Using Stacked Autoencoder," IEEE Trans. on Information Forensics and Security, vol. 12, no. 2, pp. 432-444, 2017.
- [6] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," Nature, 521.7553: 436-444, 2015.
- [7] K. L. Priddy and P. E. Keller, Artificial neural networks: an introduction, SPIE Press, 2005.
- [8] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," Int. Conf. on Machine Learning(ICML), pp. 448-456, 2015.
- [9] H.-W. Yun, S.-H. Shin, W.-J. Jang and H. Park, "On-line audio genre classification using spectrogram and deep neural network," J. of Broadcast Engineering, vol. 21, no. 6, pp. 977-985, Nov. 2016.

저 자 소 개



조 효 진

- 2017년 2월 : 광운대학교 전자공학과 공학사
- 2017년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- ORCID : <http://orcid.org/0000-0003-2296-2270>
- 주관심분야 : 오디오/음성 신호처리, 딥 러닝



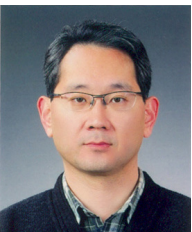
장 원

- 2017년 2월 : 광운대학교 전자공학과 공학사
- 2017년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- ORCID : <http://orcid.org/0000-0002-4711-780X>
- 주관심분야 : 오디오/음성 신호처리, 딥 러닝



신 성 현

- 2016년 2월 : 광운대학교 전자공학과 공학사
- 2016년 3월 ~ 현재 : 광운대학교 전자공학과 석박사통합과정
- ORCID : <http://orcid.org/0000-0002-2343-8983>
- 주관심분야 : 오디오/음성 신호처리, 딥 러닝



박 호 종

- 1986년 2월 : 서울대학교 전자공학과 공학사
- 1987년 12월 : Univ. of Wisconsin-Madison 공학석사
- 1993년 5월 : Univ. of Wisconsin-Madison 공학박사
- 1993년 9월 ~ 1997년 8월 : 삼성전자 선임연구원
- 1997년 9월 ~ 현재 : 광운대학교 전자공학과 교수
- ORCID : <http://orcid.org/0000-0003-1600-6610>
- 주관심분야 : 오디오/음성 신호처리, 3D 오디오, 음악정보처리