

# Analysis of Voice File Forgery Detection Techniques: Focusing on Korean Academic Journals

## 음성 파일 위변조 검출 기법 분석: 국내 학술지를 중심으로

Yeongmin Son<sup>1</sup>, Jae Wan Park<sup>2</sup>

손영민<sup>1</sup>, 박재완<sup>2</sup>

<sup>1</sup> Master Student, Dept. of Media, Soongsil University, Korea, [son342@soongsil.ac.kr](mailto:son342@soongsil.ac.kr)

<sup>2</sup> Associate Professor, Global School of Media, Soongsil University, Korea, [jaewan.park@ssu.ac.kr](mailto:jaewan.park@ssu.ac.kr)

Corresponding author: Jae Wan Park

**Abstract:** Today, voice files are easily created through a variety of methods, such as phone calls, voice messages, and voice recordings, and are increasingly being submitted as evidence in court. However, due to the nature of digital files, they can be forged and altered, so there is a risk that they can be abused by individuals or organizations with malicious purposes. Accordingly, the importance of forgery detection techniques to ensure the integrity and reliability of voice files is increasing. The purpose of this paper is to examine limitations and suggest development directions through analysis of domestic research on voice file forgery detection techniques. Voice file forgery detection techniques are largely classified into frequency analysis, metadata and file structure analysis, and artificial intelligence utilization. These three detection techniques have limitations in detecting voice file forgery in the case of precise editing. Therefore, this study suggests the need to develop a new algorithm and build a data set to build a deep learning model for detecting forgery of voice files. Additionally, this study suggests the need for a deep learning-based authentication system to prove the integrity of forged voice files before they are submitted as evidence in court. This study is expected to contribute to the development of forgery detection techniques by analyzing the limitations of voice file forgery detection techniques and suggesting development directions.

**Keywords:** Voice File Forgery Detection, Audio Forensics, Voice File Editing, Deep Learning, Case Study

**요약:** 오늘날 음성 파일은 전화 통화, 음성 메시지, 음성 녹음 등 다양한 방법으로 손쉽게 생성되면서 법정에서 증거물로 제출되는 경우가 점차적으로 증가하고 있다. 그러나 디지털 파일의 특성상 위변조가 가능해 악의적 목적을 가진 개인이나 조직에 의해 악용될 수 있는 위험이 존재한다. 따라서 음성 파일의 무결성과 신뢰성을 보장하기 위한 위변조 검출 기법의 중요성이 증대되고 있다. 본 논문은 음성 파일 위변조 검출 기법에 대한 국내 연구의 분석을 통해 한계점을 조사하고 발전 방향을 제안하는 것을 목적으로 한다. 음성 파일 위변조 검출 기법은 크게 주파수 분석, 메타데이터 및 파일 구조 분석, 인공지능 활용으로 분류된다. 이러한 세 가지 검출 기법들은 정밀한 편집의 경우에 음성 파일 위변조를 검출하는 데 한계를 지닌다. 따

Received: July 11, 2023; 1<sup>st</sup> Review Result: August 16, 2023; 2<sup>nd</sup> Review Result: September 20, 2023  
Accepted: October 25, 2023

라서 본 연구는 음성 파일의 위변조 검출을 위한 딥러닝 모델 구축을 위한 새로운 알고리즘 개발 및 데이터 셋 구축의 필요성을 제안한다. 또한 본 연구는 위변조된 음성 파일이 법원에 증거물로 제출되기 전, 무결성을 입증하기 위한 딥러닝 기반의 인증 시스템의 필요성을 제안한다. 본 연구는 음성 파일 위변조 검출 기법에 대한 한계점 분석 및 발전 방향을 제시함으로써 위변조 검출 기법의 발전에 공헌할 것으로 기대된다.

**핵심어:** 음성 파일 위변조 검출, 오디오 포렌식, 음성 파일 편집, 딥러닝, 사례 연구

## 1. 서론

현대 디지털 통신 시스템에서 디지털 음성은 음성 통화, 음성 인식, 음성 메시지 등 다양한 분야에서 중요한 역할을 수행하고 있다. 하지만 디지털 음성은 디지털 데이터의 특성상 위변조가 가능하기 때문에 사회공학적 공격, 목소리 위조, 사기 등 다양하게 악용될 수 있는 위험이 존재한다. 이에 법정에서 음성 파일이 증거로서 효력을 발휘하기 위해 데이터의 무결성과 신뢰성을 보장하는 것이 중요하게 되었다[1].

음성 파일 위변조는 음성 파일의 내용을 조작하거나 새로운 음성 파일을 만들어 내는 것을 의미한다. 이러한 위변조 기법은 기술의 발전과 함께 점점 더 정교해지고 있으며, 최근 인공지능 기술을 이용해 위변조를 수행하는 사례가 증가하고 있다[2]. 따라서 음성 파일 위변조에 대응하기 위해서는 위변조 기법을 이해하고, 이를 탐지하기 위한 기술을 개발하는 것이 요구된다. 본 연구는 이를 위한 사례 연구로서 국내에 출판된 저널 중 음성 파일 위변조 검출에 대한 다양한 기법들의 분석을 통해 음성 파일 위변조 검출에 대한 이해를 높이고, 한계점 분석 및 발전 방향을 제시하는 것을 목적으로 한다.

디지털 포렌식은 녹음된 장치를 수집하고 녹음 장치로부터 음성 파일을 수거하여 녹음 장치 및 음성 파일을 분석하는 작업이 필요하다. 하지만 일반적으로 경미 사건의 경우 녹음 파일 및 녹취록을 증거물로 제출한다. 또한 개인정보보호를 이유로 경찰이 수사 초기에 녹음된 장치를 압수하는 것은 쉽지 않다. 이에 녹음 장치가 새롭게 포맷되어 초기화되었을 경우 디지털 포렌식은 제대로 작동되기 어렵다. 따라서 본 연구는 녹음 장치를 분석에 수행하는 논문은 제외하고, 음성 파일 통해 오디오 포렌식을 수행한 논문으로 연구의 범위를 정한다.

위변조 탐지 기법은 주로 머신러닝(machine learning), 신호처리, 메타데이터 및 파일 구조 분석 기법을 활용하여 발전하고 있다[3]. 머신러닝은 대량의 훈련 데이터를 기반으로 음성 파일의 패턴과 특징을 학습하여 위변조 여부를 판별한다. 신호처리 기법은 음성 신호의 주파수 대역에서 발생하는 특성 등을 활용하여 원본과 위변조 파일을 구분한다. 메타데이터 분석은 원본 파일과 위변조 파일의 메타데이터 및 구조를 대조하여 위변조를 판별하는 기법이다. 따라서 본 연구에서는 크게 3가지 유형에 따른 위변조 검출 기법에 대한 깊이 있는 사례 조사 및 분석을 통해 현재 위변조 검출 기법의 특징과 한계점을 조사하였다. 또한 본 논문은 기존 연구의 한계점을 기반으로 음성 파일 위변조 탐지 기술의 발전 방향을 제시한다. 음성 위변조 탐지 기술의 발전은 디지털 통신 시스템의 보안을 향상시키고, 법정에서 증거물로 제출된 음성 파일에 대한 법관의 판단 기준에 중요한 역할을 수행할 것으로 기대된다.

## 2. 음성 파일 위변조 검출의 현황

스마트폰과 같은 미디어 기기의 확산과 기술의 진보로 인해 음성을 이용하는 디지털 서비스가 증가하고 있다. 이를 이용해 목소리 인증, Speech-to-Text(STT), 화자 인식, 딥 보이스(Deep Voice) 등의 기술이 발전하고 있다. 또한 법정 증거물로서 음성 파일을 제출하는 사례가 증가하고 있다. 음성 파일이 신뢰성을 가지기 위해서는 음성 파일이 원본임을 보장해야 한다. 재판부는 음성 파일에 대해 “녹음파일 사본에 편집의 흔적을 발견할 수 없고, 이 사건 녹음파일 사본의 파일 정보와 녹음 주파수 대역이 위 디지털 녹음기로 생성한 파일의 그것들과 같다”라는 대법원 판례의 근거를 인용하여 증거로 채택한다[3]. 이는 음성 파일의 메타데이터 및 구조는 위변조가 거의 불가능하다는 전제하에서 기인된 것으로 볼 수 있다[3]. 그러나 음성파일의 메타데이터 및 구조의 위변조는 가능하며, 현재 음성 편집 기술을 이용해 주파수 대역의 손상 없이 음성 파일의 위변조가 가능하다.

기존 음성 편집 기술인 ‘insertion’, ‘deletion’, ‘splicing’은 편집 영역의 전후의 시간 프레임이 변화하여 생성되는 다양한 특징에 의해 위변조 여부가 발견될 가능성이 있다. 또한 동일한 파일 내에서 음성 세그먼트(segment)를 복사한 후 원하는 위치로 붙여넣기하는 ‘copy-move’은 동일한 환경 내의 세그먼트를 복사한다는 점에서 삽입과 구분될 수 있으며, 이는 검출에 어려움이 있을 수 있다[4]. 더욱이 ‘mix paste’ 기법은 원본과 동일한 시간 프레임을 유지하면서 동일 파일 내 혹은 다른 파일 내의 음성 세그먼트를 합성하는 기법으로, 이를 이용하면 정교한 편집 및 목소리를 겹치게 할 수 있다. 음성 파일의 정교한 편집인 경우에는 음성 파일만으로 위변조를 검출하는 것은 거의 불가능할 수 있다.

결론적으로 음성 파일의 메타데이터의 조작은 가능하며, 또한 편집한 파일의 구조를 원본과 동일하게 만드는 것도 가능하다. 현재 딥 보이스 기술을 이용해 생성한 음성을 원본 음성에 합성하여 법정에 증거로 제출할 가능성도 간과할 수 없는 상황에 처해있다[5]. 따라서 음성 파일이 법정에서 효력을 발휘하기 위해서는 음성 파일의 무결성과 신뢰성을 판단할 근거가 필요하며[6], 이에 음성 파일 위변조 검출을 위한 효과적인 기술을 개발하는 것이 요구되고 있다.

## 3. 음성 파일 위변조 검출 기법 분석

국내 저널에 발표된 음성 파일 위변조 검출 기법은 크게 (1) 주파수 분석, (2) 메타데이터 및 파일 구조 분석, (3) 인공지능 활용으로 나누어질 수 있다.

### 3.1 주파수 분석

주파수 분석 기법은 음성 신호를 주파수 신호로 변환하는 알고리즘인 Fast Fourier Transform(FFT)를 이용하는 기법으로, 이를 위해 스펙트럼(spectrum)과 스펙트로그램(spectrogram)이 주로 활용된다. 일반적으로 주파수 분석은 차단 주파수(cut-off frequency), Zero Padding에 의한 지연시간, 주변 잡음을 활용한 불연속 구간 분석 등을 수행하게 된다[1][7]. 녹음 장치가 지닌 녹음 파일의 특성을 이해하고 정교한 편집을 수행할 경우 검출이 불가능할 수 있다.

주파수 영역에서는 인간이 들을 수 있는 음성 외에도 많은 정보가 포함되어 있다.

대표적으로 Electrical Network Frequency(ENF) 신호가 있다. ENF 신호는 60Hz 안팎의 교류 신호로, 발전 및 송전 과정에서 발생한다. 이 신호의 불연속점을 찾아 음성이 인위적으로 splicing 및 insertion, 중간 영역의 deletion된 구간이 있는지 찾아내는 것이 가능하다[8]. 이러한 주파수 분석 기법은 서로 다른 녹음 기기나 환경에서 녹음된 두 가지 이상의 파일을 이용해 편집한 위변조 음성을 검출하는데 좋은 성능을 보인다. 하지만 전력망 주파수 신호의 데이터베이스화가 필요하며[7], 동일한 장소와 시간에서 생성된 녹음 파일은 위변조 검출이 어렵다. 즉, 동일한 음성 파일 내에서 ‘mix paste’ 기법을 사용하여 편집에 따른 주파수 특징이 뚜렷하게 나타나지 않은 경우에는 위변조 검출이 불가능하다. 또한 최신 녹음 장치는 잡음을 제거하는 기능이 기본적으로 탑재되어 있어 주파수 특징을 얻기가 어렵다[1]. 더욱이 주파수 특징을 얻는다 해도 음성 편집 기술의 발달에 따라 주파수 영역에서의 편집이 가능하기 때문에 정교하게 편집된 음성 파일에는 대응이 어렵다. 더욱이 오디오 스펙트럼을 분석하는 알고리즘을 통해 비음성 구간의 삭제 검출을 검출하는 연구가 진행되었다[9]. 하지만 ‘mix paste’ 및 ‘copy-move’ 이용한 정교한 편집에는 적용이 어렵다.[표 1]은 주파수 분석 관련 국내 연구의 특징 및 한계점을 나타낸다.

[표 1] 주파수 분석 관련 국내 연구

[Table 1] Korean Research on Frequency Analysis

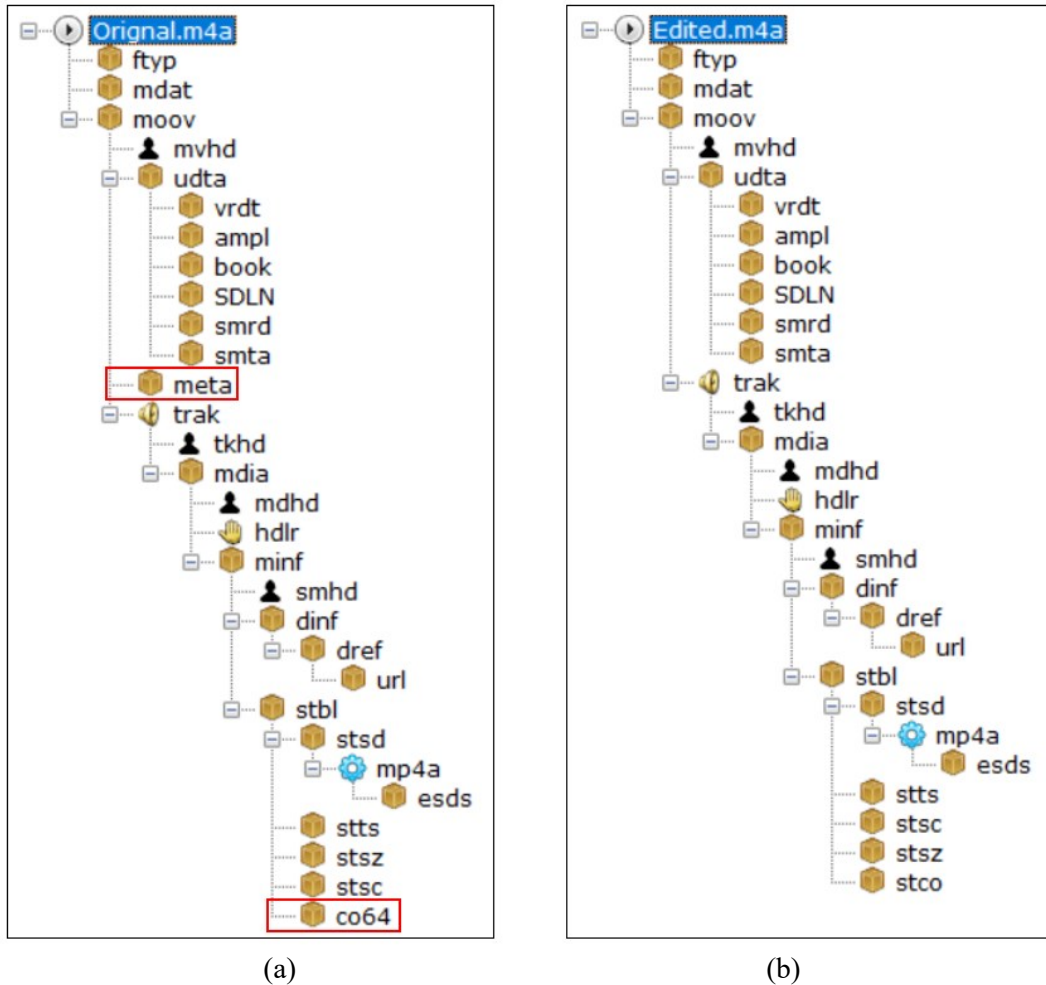
출판년도	논문 제목	저자	특징	한계점
2014	“음성 주파수 분포 분석을 통한 편집 의심 지점 검출 방법”[9]	허희수 외 3인	오디오 스펙트럼을 분석하는 알고리즘을 통해 비음성 구간의 삭제 검출 가능	mix paste 및 copy-move를 이용한 편집시, 검출이 어려움
2018	“이차 보간에 따른 ENF 기반의 위변조 디지털 파일 탐지 기법”[8]	박세진, 윤지원	ENF 신호를 추출하여 위변조 검출, splicing, insertion, 중간영역 deletion에 적용 가능	정교한 편집 검출 불가하며, ENF 신호의 데이터베이스가 필요함. 최신의 녹음 장치에서 ENF 특징을 찾기 어려움
2019	“디지털 오디오 파일의 편집 여부 분석 절차에 대한 연구”[1]	박남인 외 2인	Cut-off Frequency, 주변 잡음을 활용한 불연속 구간 분석	정교한 편집 검출 불가
2022	“삼성 스마트폰의 음성 녹음 어플리케이션에 의해 녹음된 오디오 파일에 대한 위변조 여부 분석 방법”[7]	박남인 외 10인	Zero-padding, Cut-off Frequency	정교한 편집 검출 불가

### 3.2 메타데이터 및 파일 구조 분석

녹음장치를 통해 녹음된 음성 파일에는 녹음 일시, 녹음장치 정보, 음성신호와 관련된 다양한 오디오 정보 등의 메타데이터가 기록된다. 녹음된 음성 파일은 녹음장치의 종류 및 운영체제(operating system) 등에 따라 메타데이터가 상이하고, 음성 파일의 구조 또한 달라질 수 있다. 음성파일을 위변조한 후에 원본 파일과 동일한 파일 포맷(format)을 위해 인코딩할 경우, 메타데이터와 파일 구조는 달라진다. 이러한 특성에 의해 M4A 파일 분석이 가능한 도구인 MP4 Inspector[10] 등을 이용해 동일한 녹음기와 운영체제에서 녹음된 음성 파일과의 대조를 통해 음성 파일의 편집 여부를 검출할 수 있게 된다.

최근 스마트폰에 내장된 녹음 앱들이 편집 기능을 가지기 시작하면서, 원본 파일과 동일한 메타데이터와 파일구조를 갖는지에 대한 연구가 진행되었다[1][11]. 예를 들어,

삼성 갤럭시 Z Flip 3 스마트폰(안드로이드 버전 12)에 내장된 음성녹음 앱으로 녹음한 파일을 내장된 음성녹음 앱으로 편집할 경우, 그림 1과 같이 파일 구조가 변화하게 된다. 하지만 아이폰(iOS 12.xx) 내장된 보이스 메모(Voice Memo)에서 녹음된 파일을 보이스 메모에서 편집하고 다시 저장될 경우, 녹음 파일 구조는 동일하게 된다[1][3]. 이러한 원리를 위변조 검출에 활용하기 위해 스마트폰의 제조사, 기종, 운영체제, 녹음 모드 등에 따른 스마트폰 녹음 데이터베이스를 구축의 필요성이 제시되었다[12]. 또한 스마트폰은 다양한 녹음 애플리케이션을 설치할 수 있기 때문에, 이를 활용하여 편집 시 발생하는 파일 구조의 변화를 분석하는 연구가 진행되었다[13].



[그림 1] 삼성 갤럭시 스마트폰 내장 앱에서의 (a) 녹음된 파일과 (b) 편집된 파일의 비교

[Fig. 1] Comparison of (a) Recorded and (b) Edited Files in Samsung Galaxy Smartphone Built-in Apps

메타데이터 및 파일 구조의 분석은 위변조 검출 업무의 효율성 증대, 객관적 위변조 정황 파악 등의 공헌을 기대할 수 있다. 하지만 파일의 메타데이터 및 구조는 바이너리(binary) 파일을 16진수로 표현하고 수정할 수 있는 헥스 에디터(hex editor)에 의해 조작될 수 있기 때문에 메타데이터 및 파일 구조의 분석은 보조적 수단으로 고려될 필요가 있다. [표 2]는 메타데이터 및 파일 구조 분석 관련 국내 연구의 특징 및 한계점을 나타낸다.

[표 2] 메타데이터 및 파일 구조 분석 관련 국내 연구

[Table 2] Korean Research on Metadata and File Structure Analysis

출판년도	논문 제목	저자	특징	한계점
2019	“디지털 오디오 파일의 편집 여부 분석 절차에 대한 연구”[1]	박남인 외 2인	스마트폰에 내장된 녹음 앱에서 녹음된 파일과 편집된 파일 비교	헥스 에디터에 의해 메타데이터 및 파일 구조 변경 가능
2021	“스마트폰 녹음 데이터베이스의 법과학적 활용에 대한 연구”[12]	김경화	스마트폰 녹음 데이터베이스 구축의 필요성 및 활용 방안 제시	
2022	“어플리케이션의 음성녹음 파일에 대한 위변조 탐지 연구”[13]	송유진, 김기범	다양한 모바일 애플리케이션에서의 편집 시, 파일 구조의 변화를 분석	
2022	“법과학적 활용을 위한 삼성 스마트폰 음성 녹음 파일의 메타데이터 구조 및 속성 비교 분석 연구”[11]	안서영 외 2인	업무의 효율성을 증대	

### 3.3 인공지능 활용

최근 음성 위변조 검출에 인공지능 기술을 적용해 기존 방법의 한계를 보완하는 연구가 이루어지고 있다. 머신러닝 모델의 학습은 음성 파일에서 Mel-Frequency Cepstral Coefficient(MFCC) 및 Linear-Frequency Cepstral Coefficient(LFCC)와 같은 신호 정보를 추출하여 활용한다. 이 정보는 한 음성 파일 내 어느 시점에서나 신호 정보가 유사하게 나타난다는 점을 이용해 위변조 구간을 구분하는 머신러닝 모델이 제안되었다[14]. 이 모델은 음성 파일의 오디오 신호를 일정한 길이의 프레임별로 구분하여 MFCC와 LFCC 특징을 추출하고 특징 분포를 계산한 후, 전체 음성 파일에서 특징 분포가 상이한 구간을 검출하여 점수화한 앙상블(ensemble) 기법이 적용되었다[14]. 이는 insertion의 검출은 가능하나, 동일한 음성 파일의 특정 구간을 삽입하는 copy-move나, insertion을 제외한 위변조 방식은 검출이 불가능하다.

딥러닝 모델의 학습은 주로 음성 신호를 스펙트로그램(spectrogram)으로 변환하여 데이터 셋으로 활용한다. 스펙트로그램 이미지는 시각 처리 과정을 모방한 Convolutional Neural Network(CNN) 구조를 통해 특징 벡터로 추출되어 노이즈(noise)에 대한 안정성을 가지게 된다. 이를 통해 기존 주파수 분석 기법의 한계인 녹음 환경에 따른 영향을 줄이는 것이 가능해졌다. 최근 음성 분류 기법에서 주로 쓰이는 Mel-Spectrogram을 이용한 위변조 검출 기법이 제안되었다[15-17]. Mel-Spectrogram은 음성 신호에 Mel-filter bank를 적용하여 저주파의 특성을 추출할 때 사용되는 기법으로[18], Mel-Spectrogram은 주로 CNN의 특징 벡터를 추출하는 과정에 활용된다. 이를 통해 음성 파일이 스마트폰에서 녹음되었는지 여부를 검출하는 딥러닝 모델이 제안되었다[15]. 이 연구에서 제안된 모델은 학습에 사용되지 않은 스마트폰 기종에 대해서도 스마트폰 녹음 여부 검출이 가능하나[15], 학습에 사용된 ‘기타기기 녹음’ 데이터인 YouTube 영상 오디오 파일은 인간이 스마트폰 녹음 파일과 구분될 수 있는 다양한 정보를 지니고 있다.

최근 딥 보이스와 같은 음성 합성 기술의 진보로 이를 탐지하려는 연구가 수행되었다. 이에 CNN과 MKDE (Multivariate Kernel Density Estimation) 모델을 직렬로 연결하여 합성 음성을 분류하는 연구가 수행되었고, 높은 정확도로 음성 합성에 사용된 알고리즘을 분류

하는 것이 가능해졌다[16]. 또한 딥 보이스를 탐지하는 연구가 진행되었다. 이 연구는 Mel-Frequency Cepstral Coeffic (MFCC)와 Mel-Spectrogram을 활용하여 음성 특징을 추출하고, 이를 CNN 모델과 BiLSTM 모델을 통해 학습시키고 앙상블 기법을 적용한 모델을 구축하였다[17]. 제안된 모델은 높은 정확도로 딥 보이스를 탐지하는 것이 가능하다. [표 3]은 인공지능 관련 국내 연구의 특징 및 한계점을 나타낸다.

[표 3] 인공지능 활용 관련 국내 연구

[Table 3] Korean Research on Artificial Intelligence Utilization

출판년도	논문 제목	저자	특징	한계점
2014	“법음성학에서의 오디오 신호의 위변조 구간 자동 검출 방법 연구”[14]	양일호 외 5인	다른 녹음 환경에서 생성한 음성을 insertion한 위변조 구간 검출	deletion, copy-move, mix paste 편집시 검출 불가
2022	“스마트폰 녹음 여부 검출을 위한 딥러닝 기반 오디오 포렌식”[15]	엄소정 외 2인	Mel-Spectrogram 과 CNN 구조를 이용해 스마트폰 녹음 여부 검출	구분 가능한 다양한 정보가 존재
2023	“CNN과 MKDE 모델의 직렬 연결을 기반한 합성 음성 분류”[16]	심영준 외 2인	CNN-MKDE 모델을 활용하여 음성 합성에 사용된 알고리즘을 분류	합성 여부에 대한 검출 불가
2023	“음성 특징 추출 및 딥러닝 기반 딥 보이스 탐지 시스템”[17]	한승우 외 4인	CNN 모델과 BiLSTM 모델을 이용한 앙상블 기법 제안	deletion, copy-move, mix paste 편집시 검출 불가

#### 4. 토론

음성파일 위변조 검출과 관련된 국내 논문을 분석한 결과, 음성파일 위변조 검출의 기법은 크게 (1) 주파수 분석, (2) 메타데이터 및 파일 구조 분석, (3) 인공지능 활용으로 유형화되었다. [표 4]는 3가지 유형에 따른 세부 분석 요소를 나타낸다.

기존의 위변조 검출 작업은 음성파일의 메타데이터 및 구조 분석을 수행하고 주파수 특징을 분석하는 방법으로 진행되었다. 그러나 음성 파일의 메타데이터 및 구조를 원본과 동일하게 구축하는 것이 가능하고 주파수 영역에서의 편집이 가능하기 때문에 정교하게 편집된 음성 파일의 위변조를 검출하는 것은 사실상 불가능하다. 또한 최근 인공지능을 활용하여 음성 파일의 위변조를 찾으려는 시도가 이루어지고 있으며, 이러한 연구들은 딥러닝 기술을 활용하여 진행 되고 있다. 딥러닝 연구를 위해 다량의 양질의 데이터셋의 확보가 필요하다. 이러한 이유로 인해 비교적 데이터 셋 확보가 쉬운 딥 보이스 검출과 관련된 연구가 진행되고 있는 것으로 판단된다.

현재 멀티미디어 위변조를 검출하는 연구가 활발히 진행되었으나, 대부분 사진 및 영상을 대상으로 한 연구가 주로 이루어 졌으며 상대적으로 음성 파일 위변조 검출 연구는 부족한 상황이다[19]. 음성 파일 위변조 검출을 위한 심도 있는 연구를 위해, 우선적으로 안티포렌식(anti-forensic) 기법에 대한 연구 필요하며, 이를 기반으로 한 검출 알고리즘 개발이 필요하다.

최근 딥러닝 기술의 발전은 음성 파일만을 가지고 위변조를 정확히 판별할 수 있는 가능성을 열고 있다. 이를 위해 다양한 녹음 장치 및 운영체제에서 녹음된 음성 파일의 수집과 다양한 음성 편집 소프트웨어 따른 편집 기법이 적용된 오픈(open) 데이터 셋 구축이 선행될 필요가 있다. 이는 보편적 데이터 셋은 아니지만 사회적 문제해결을 위해 정부의 적극적 지원이 필요하다. 이를 활용하여 음성 파일 위변조를 검출할 수 있는

다양한 딥러닝 알고리즘 및 모델이 출현하게 될 것이다. 이는 위변조된 음성 파일이 법원에 증거물로 제출되기 전에 무결성을 입증하기 위한 인증 시스템 구축을 가능하게 할 것이다.

[표 4] 국내 논문에 나타난 음성파일 위변조 검출 세부 분석 기법

[Table 4] Detailed Analysis Technique for Detecting Forgery Voice Files Presented in Korean Academic Journals

위변조 검출 기법	세부 분석 요소	대표적 활용 기술 및 도구	검출 편집 기법	검출의 어려움 및 불가
주파수 분석	Zero-padding 분석	스펙트로그램	deletion, splicing	정교한 편집
	Cut-off Frequency 분석	스펙트로그램	splicing, insertion	정교한 편집
	주변 잡음을 활용한 불연속 구간 분석	스펙트로그램	insertion	mix paste 및 copy-move 편집
	ENF 신호 분석	ENF 신호 데이터베이스	splicing, insertion, 중간 영역 deletion	동일한 시간과 장소의 파일, mix paste 편집
	편집 의심 지점 검출	스펙트럼	deletion	동일한 sample rate, deletion 외의 편집 기법
메타데이터 및 파일 구조 분석	메타데이터 분석	Mediainfo 동일한 원본 녹음 파일 녹음 데이터베이스	편집 후 원본과 동일한 파일 포맷으로 인코딩	헥스 에디터를 이용한 메타데이터 및 파일 구조의 위변조
	파일 구조	MP4 Inspector Hex Editor 동일한 원본 녹음 파일 녹음 데이터베이스		
인공지능 활용	위변조 구간 자동 검출	MFCC, LFCC, GSV, SVC 양상불	insertion 검출	동일한 환경의 녹음 파일 합성
	녹음 장치 검출	Mel-Spectrogram, CNN,	스마트폰 녹음 여부 검출	스마트폰 이외의 녹음 장치 활용
	음성 합성 알고리즘 분류	Mel-Spectrogram, CNN, MKDE	음성합성 알고리즘 분류	음성 합성 여부
	딥 보이스 검출	MFCC, Mel-Spectrogram CNN, BiLSTM, 양상불	딥 보이스 검출	학습한 데이터에 존재하지 않는 기법

### 5. 결론 및 향후 연구

본 연구는 국내 학술지를 기반으로 위변조 검출 기법에 대한 사례 분석을 통한 한계점 분석 및 발전 방향을 도출하는 것을 목적으로 했다. 최근 음성 파일 편집 도구 발전과 인공지능 기술의 발전에 따라, 이를 악용하는 사례가 증가하고 있다. 이에 법정에서 증거물로 제출되는 음성 파일에 대한 무결성을 입증하는 것이 중요시되고 있다.

현재 음성 파일 위변조를 탐지하는 연구는 다양하게 수행되고 있지만, 현재의 기술로 정교하게 편집된 음성 파일의 위변조를 검출하는 것은 거의 불가능하다. 하지만 딥러닝 기술의 진보로 인해 이러한 한계를 극복하는 것이 가능할 것으로 판단된다. 이를 위해 다양한 음성 편집 기법이 적용된 데이터 셋의 구축이 필요하며, 이를 검출하기 위한



알고리즘 개발이 요구된다.

현재 향후 연구로 다양한 음성 편집 기법을 적용한 데이터 셋 구축 및 이의 검출을 위한 알고리즘 및 딥러닝 모델을 개발 중이다. 본 연구는 음성 파일 위변조 검출 기법에 대한 한계점 분석 및 발전 방향을 제시함으로써 위변조 검출 기법의 발전에 공헌할 것으로 기대된다.

## 6. 감사의 글

이 논문은 서울시 산학연 협력사업 2022년도 제 6회 서울 혁신 챌린지 결선(IC220010)의 지원을 받아 수행된 연구임

## References

- [1] Nam In Park, Kyu-Sun Shim, Oc-Yeub Jeon, A Study on Authentication Analysis Procedure of Digital Audio Files, *Journal of Digital Forensics*, (2019), Vol.13, No.4, pp.257-270.  
DOI: <http://doi.org/10.22798/kdfs.2019.13.4.257>
- [2] [https://www.hani.co.kr/arti/society/society\\_general/1070506.html](https://www.hani.co.kr/arti/society/society_general/1070506.html), Jul 4 (2022)
- [3] Jae Wan Park, Won Jun Kwak, John Sanghyun Lee, A Study on Forgery Techniques of Smartphone Voice Recording File Structure and Metadata, *Journal of the Convergence on Culture Technology*, (2022), Vol.8, No.6, pp.807-812.  
DOI: <https://doi.org/10.17703/JCCT.2022.8.6.807>
- [4] Ustubioglu, Arda, Beste Ustubioglu, and Guzin Ulutas, Mel spectrogram-based audio forgery detection using CNN, *Signal, Image and Video Processing*, (2023), Vol.17, No.5, pp.2211-2219.  
DOI: <https://doi.org/10.1007/s11760-022-02436-4>
- [5] MATT REYNOLDS, Courts and lawyers struggle with growing prevalence of deepfakes, *ABA Journal*, (2020)  
Available from: <https://www.abajournal.com/web/article/courts-and-lawyers-struggle-with-growing-prevalence-of-deepfakes>
- [6] Hasan Fayyad-Kazan, Ale Hejase, Imad Moukadem, and Sondos Kassem-Moussa, Verifying the Audio Evidence to Assist Forensic Investigation, *Computer and Information Science*, (2021), Vol.14, No.3, pp.25-37.  
DOI: <https://doi.org/10.5539/cis.v14n3p25>
- [7] Nam In Park, Ji Woo Lee, Seong Ho Lim, Jin-Hwan Kim, Jae Sung Lim, Jung Hwan Lee, Do Joon Jung, Ji-Hun Kim, Jun Seok Byun, Oc-Yeub Jeon, and Gi-Hyun Na, Forgery Analysis Method for Audio Recordings Generated by Voice Recorder Application from Samsung Smartphones, *Korean Journal of Forensic Sciences*, (2022), Vol.23, No.2 pp.77-80.  
DOI: <https://doi.org/10.53051/ksfs.2022.23.2.10>
- [8] Se Jin Park, Ji Won Yoon, ENF based Detection of Forgery and Falsification of Digital Files due to Quadratic Interpolation, *Journal of KIISE*, (2018), Vol.45, No.3 pp.311-320.  
DOI: <https://doi.org/10.5626/JOK.2018.45.3.31>
- [9] Heo Hee-Soo, So Byung-Min, Yang IL-Ho, and Yu Ha-Jin, A Speech Waveform Forgery Detection Algorithm Based on Frequency Distribution Analysis, *Phonetics and Speech Sciences*, (2015), Vol.7, No.4 pp.35-40.  
DOI: <http://dx.doi.org/10.13064/KSSS.2015.7>
- [10] <https://sourceforge.net/projects/mp4-inspector/>, Sept 11 (2023)
- [11] Seo-Yeong Ahn, Se-Hui Ryu, Kyung-Wha Kim, and Ki-Hyung Hong, A comparative analysis of metadata structures and attributes of Samsung smartphone voice recording files for forensic use, *Phonetics and Speech Sciences*, (2022), Vol.14, No.3, pp.103-112.  
DOI: <https://doi.org/10.13064/KSSS.2022.14.3.103>

- [12] Kyung-Wha Kim, A study on the forensic application of smartphone recording database, *Journal of Digital Forensics*, (2021), Vol.15, No.1, pp.26-42.  
DOI: <https://doi.org/10.22798/kdfs.2021.15.1.26>
- [13] YouJin Song, Gibum Kim, A Study on the Detection of Falsification of Voice Recording Files in an Application, *Journal of Digital Forensics*, (2022), Vol.16, No.3, pp.65-76.  
DOI: <https://doi.org/10.22798/kdfs.2022.16.3.65>
- [14] Il-Ho Yang, Kyung-Wha Kim, Myung-Jae Kim, Rock-Seon Baek, Hee-Soo Heo, Ha-Jin Yu, An Automatic Method of Detecting Audio Signal Tampering in Forensic Phonetics, *Phonetics and Speech Sciences*, (2014), Vol.6, No.2, pp.21-28.  
DOI: <https://doi.org/10.13064/ksss.2014.6.2.021>.
- [15] So-Jeong Eom, Hyun-Soo Kim, and Hae-Yeoun Lee, Audio Forensics for Smartphone Recording Detection using Deep Learning, *The Journal of Korean Institute of Information Technology*, (2022), Vol.20, No.7, pp.103-109.  
DOI: <https://doi.org/10.14801/jkiit.2022.20.7.103>
- [16] Youngjun Sim, Jungyu Choi, and Sungbin Im, Synthetic Speech Classification based on Cascade Connection of CNN and MKDE Models, *Journal of The Institute of Electronics and Information Engineers*, (2023), Vol.60, No.2, pp.94-101.  
DOI: <https://doi.org/10.5573/ieie.2023.60.2.94>
- [17] Seung-Woo Han, Seong-Hun Han, Seong-Min You, Dong-ho Song, and Chang-Jin Seo, Deep voice detection system based on voice feature extraction and deep learning, *Proceedings of the the 54th KIEE Summer Conference*, Korea Institute of Electrical Engineers, (2023)
- [18] Feng Liu, Tongsheng Shen, Zaili Luo, Dexin Zhao, and Shaojun Guo. Underwater Target Recognition Using Convolutional Recurrent Neural Networks with 3-D Mel-spectrogram and Data Augmentation, *Applied Acoustics*, (2021), Vol.178, 107989.  
DOI: <https://doi.org/10.1016/j.apacoust.2021.107989>
- [19] Jongwon Seok, Microphone Type Classification for Digital Audio Forgery Detection, *Journal of Korea Multimedia Society*, (2015), Vol. 18, No. 3, pp.323-329.  
DOI: <http://dx.doi.org/10.9717/kmms.2015.18.3.323>