



A comparative analysis of metadata structures and attributes of Samsung smartphone voice recording files for forensic use*

Seo-Yeong Ahn¹ · Se-Hui Ryu¹ · Kyung-Wha Kim² · Ki-Hyung Hong^{3,**}

¹Department of Future Convergence Technology Engineering, Sungshin Women's University, Seoul, Korea

²Forensic Science Division, Supreme Prosecutor's Office, Seoul, Korea

³Department of Service Design Engineering, Sungshin Women's University, Seoul, Korea

Abstract

Due to the popularization of smartphones, most of the recorded speech files submitted as evidence of recent crimes are produced by smartphones, and the integrity (forgery) of the submitted speech files based on smartphones is emerging as a major issue in the investigation and trial process. Samsung smartphones with the highest domestic market share are distributed with built-in speech recording applications that can record calls and voice, and can edit recorded speech. Unlike editing through third-party speech (audio) applications, editing by their own builtin speech applications has a high similarity to the original file in metadata structures and attributes, so more precise analysis techniques need to prove integrity. In this study, we constructed a speech file metadata database for speech files (original files) recorded by 34 Samsung smartphones and edited speech files edited by their built-in speech recording applications. We analyzed by comparing the metadata structures and attributes of the original files to their edited ones. As a result, we found significant metadata differences between the original speech files and the edited ones.

Keywords: audio metadata, voice recording, speech forgery detection, MPEG (Moving Picture Experts Group)-4

1. 서론

스마트폰의 대중화로 인하여 근래 범죄의 증거자료로 제출되는 녹취 파일은 대부분 스마트폰을 통하여 생산되고 있으며 (Kim, 2021), 스마트폰을 기반으로 한 녹음 파일의 무결성(위변조) 여부가 수사와 재판 과정에서 주요 쟁점으로 떠오르고 있다.

증거 녹음 파일의 위변조 식별 방식에는 정밀 디지털 신호 분석을 통한 음성/음향 신호의 연속성(continuity) 분석 방식, 녹음 장치에 따른 메타데이터 분석[파일 구조(file structure), 오디오 포맷(audio format) 및 특징 확인] 등이 존재한다(Park et al., 2019). 음성/음향 신호의 연속성 분석 방식은 개별 녹음 신호를 정밀 분석하는 방식으로서, 많은 시간과 노동력이 있어야 한다. 증거

* This work was supported by the Supreme Prosecutor's Office Research Grant in 2021 (research title: Development of Analysis Technique of Metadata for Forgery of Smartphone Recording Files).

** khhong@sungshin.ac.kr, Corresponding author

Received 19 August 2022; Revised 15 September 2022; Accepted 15 September 2022

© Copyright 2022 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

녹음 파일의 녹음 길이가 늘어날수록 노동력과 시간 소모 또한 비례하여 늘어나며, 음성/음향 신호에 대한 전문 지식을 가진 인력이 투입되어야 한다는 단점이 있다. 또한 Park et al.(2019)의 연구에 따르면 문장 또는 어절 단위로 녹음 파일을 편집한 경우에는 음성/음향 신호의 연속성을 통한 위변조의 검출에 어려움이 있다.

메타데이터 분석 방식은 증거 녹음 파일의 메타데이터와 동일한 환경 및 장치에서 녹음된 파일의 메타데이터를 비교하여 해당 파일의 무결성(위변조) 여부를 판단하는 방식이다. Grigoros & Smith(2017)는 반드시 파일 형식 및 메타데이터 분석을 디지털 오디오 인증 과정의 일부에 포함할 것을 주장하였다. 메타데이터 분석 방식은 반드시 동일한 환경 및 장치에서 생성된 원본 음성 파일을 필요로 한다는 단점이 존재하지만 음성 파일의 크기와 녹음 길이와 무관하게 처리 속도가 일관적이며, 컨테이너와 코덱 규정을 따르는 반정형 데이터를 다루는 것이기 때문에 비교 로직의 일반화가 가능하다. 또한 대조를 위한 원본 음성 파일을 수집하여 데이터베이스를 구축함으로써 메타데이터 분석을 통한 위변조 판별 방식을 더욱 가속화 할 수 있다.

음성 파일의 메타데이터란 음성 파일 중 실제 음성 데이터를 제외한 데이터로 파일의 처리 및 녹음 정보 등으로 구성된다. 메타데이터의 서술 규칙과 포함 정보는 음성 파일의 포맷, 컨테이너, 코덱 등에 따라 상이하게 나타나며, 일부 컨테이너의 경우 메타데이터 상에 녹음 소프트웨어 인코더의 자율적인 추가 정보 기술을 허용한다. MPEG-4의 udtA 부분이나, mp3의 id3가 그 예시이다(Apple, 2016; Dan, 2020). 메타데이터 기술 과정에서 허용하는 인코딩 소프트웨어의 자율성은 해당 인코딩 소프트웨어의 원본 및 편집 파일에 대한 메타데이터의 특징을 만들어 내며, 이를 통하여 대조된 파일의 인코딩 소프트웨어 일치 여부와 위변조 여부를 판별할 수 있도록 한다.

본 연구에서는 2015년부터 2020년 사이에 출시된 삼성 안드로이드 스마트폰 34개 기종의 자체 탑재 음성 녹음 및 편집 어플리케이션을 이용한 음성 원본과 편집 파일의 메타데이터 데이터베이스를 구축하고, 편집 파일의 메타데이터 구조 및 속성상 특징을 원본과 비교 분석하여 음성 파일의 무결성(위변조) 여부를 판별하는 결정 트리를 개발하였다.

2. 관련 연구

Koenig & Lacey(2017)는 Apple사의 iOS 10.2 iPhone 7에서 iPhone 자체에 탑재된 음성 녹음 및 편집 어플리케이션을 포함한 4개의 모바일 어플리케이션을 통하여 녹음을 수행한 후, 4개의 오디오/비디오 편집 프로그램을 이용하여 편집할 경우 편집된 파일의 메타데이터 구조와 특정 메타데이터 속성(American Standard Code for Information Interchange 정보)에서 원본 m4a 파일과 비교했을 때 명확한 차이를 보임을 발견하였다. Grigoros & Smith(2017)는 Apple iPhone 및 iPod에 내장된 ‘음성 메모’ 어플리케이션을 이용하여 녹음된 음성 파일을 식별하는 의사결정 트리를 제안한다. 해당 의사 결정 트리는 외부 소프트웨어로

편집된 모든 편집본을 식별하지만, 아이폰 내장 음성 녹음 및 편집 어플리케이션인 ‘음성 메모’로 편집 파일한 파일의 경우 일부 파일에 있어 편집 여부를 판별하지 못한다. Zeng et al.(2019)은 iOS 8.1-12.2의 Apple iPhone 5개 기종을 통하여 iPhone의 ‘음성 메모’ 어플리케이션을 통하여 생성한 원본 녹음 파일이 파일 구조, 시간 관련 파일 및 메타데이터 속성 정보, 어플리케이션 데이터베이스 데이터 패턴에 있어 특정한 패턴을 가지고 있음을 발견하였다. Grigoros & Smith(2017)는 다양한 휴대용 음성 녹음 장치로부터 수집한 125개 이상의 오디오 파일과 40개 이상의 상용 소프트웨어에서 수집된 mp3, wav, wma 파일 구조와 형식에 대한 광범위한 분석을 수행하였으며, 같은 제조사의 상이한 모델에서 동일한 메타데이터 구조 및 파일 형식이 나타날 수 있음과 같은 제조사의 상이한 모델에서 파일 형식이 동일하더라도 메타데이터 구조가 상이할 수 있음을 발견하였다.

Michalek(2019)은 안드로이드 기반 삼성 스마트폰 Galaxy J3 Dual SIM을 이용하여 삼성 내에서 자체적으로 제공하는 음성 녹음 및 편집 프로그램과 9개의 외부 음성 전문 소프트웨어를 이용하여 녹음 및 편집을 수행한 음성 파일에 대한 분석을 수행하였다. 이를 통해 삼성 자체 탑재 녹음 및 편집 어플리케이션을 이용해 녹음 및 편집을 수행할 경우, 이를 특정할 수 있는 메타데이터 속성이 존재함과 MPEG 멀티미디어 컨테이너의 경우 음성 파일의 위변조 여부 판별에 유의미한 정보가 많이 포함되어 있음을 확인하였다. Kim(2021)은 12개 안드로이드 삼성 스마트폰의 녹음 방식과 편집 기능을 조사하고 원본 음성 파일과 편집 음성 파일의 메타데이터 속성 측면에서의 차이점을 확인하였다. 해당 연구에서 메타데이터 구조 측면의 비교 분석은 수행되지 않았다. Park et al.(2022)은 11대의 안드로이드 삼성 스마트폰을 이용하여 미디어 로그 및 통화 내역, 통화 내역 파일 분석(음성 지연 구간, 파일 구조 변화)을 통하여 통화 파일의 진본 여부를 확인하였다. 하지만 이는 통화 녹음에 한정되어 있으며 미디어 로그 분석을 위해 녹음을 수행한 기기가 요구된다.

메타데이터 비교를 통한 증거 녹음 파일의 위변조 판별 방식에 대한 연구는 대부분 애플의 iPhone을 기반으로 한 녹음 파일에 대한 연구(Grigoros & Smith, 2017; Koenig & Lacey, 2017; Zeng et al., 2019)이거나, 자체 탑재되어 제공되는 음성 녹음 및 편집 어플리케이션이 아닌 외부 음성 전문 소프트웨어(3rd party software)를 통한 편집 파일과 원본 음성 파일을 비교한 사례(Grigoros & Smith, 2017)가 주를 이루고 있다. 2021년 기준 국내 스마트폰 시장은 삼성 제조 스마트폰이 64.12%의 점유율을 차지함으로써(Statcounter GlobalStats, 2021), 삼성 스마트폰에서 생성된 음성 파일의 메타데이터 분석에 대한 연구가 필요하다. 또한 스마트폰 출시 시에 탑재되어 제공되는 음성 녹음 및 편집 어플리케이션을 통하여 편집된 편집 파일의 경우, 원본 파일과 편집 어플리케이션이 동일하므로 외부 음성 전문 소프트웨어를 통한 편집 파일에 비하여 상대적으로 원본 파일과의 메타데이터상 차이가 적어 무결성 분석에 어려움이 있으며(Grigoros & Smith, 2017), 스마트폰 OS의 업데이트에 따라 자체 탑재 어플

리케이션의 업데이트가 수반될 수 있기에 신기종 출시 및 OS의 업데이트에 따라 음성 파일 메타데이터 분석은 지속적으로 수행되어야 한다(Grigoras & Smith, 2017). 이에 따라 본 연구에서는 삼성 스마트폰을 통하여 녹음된 원본 음성 파일과 자체 탑재 어플리케이션을 이용한 편집 음성 파일에 대한 메타데이터 분석을 수행하였다.

3. 실험 방법

3.1. 실험 환경

실험 환경은 다음과 같다. 음성 파일로부터 메타데이터를 파싱하기 위하여 MediaInfo CLI(CommandLineInterface) 21.03 버전(MediaArea, 2018)과 Java JDK 1.8 버전을 이용하였으며, 데이터베이스 구축을 위해 MySQL 8.0을 이용하였다.

표 1에 정리한 대로 2014~2020년 출시된 안드로이드 버전 6.0.1~10의 삼성 스마트폰 34개 기종을 이용하였다.

표 1. 실험 사용 기종 정보
Table 1. Smartphone models used in this study

Model name	Model number	Release year	Android version
Galaxy Note Edge	SM-N915	2014	6.0.1
Galaxy J3	SM-J320	2015	6.0.1
Galaxy A7	SM-A700	2015	6.0.1
Galaxy S6 Edge Plus	SM-G928	2015	6.0.1
Galaxy Wide	SM-G600	2015	6.0.1
Galaxy Note 5	SM-N920	2015	7.0
Galaxy A8	SM-A810	2016	6.0.1
Galaxy S7 Edge	SM-G935	2016	8
Galaxy S7	SM-G930	2016	8
Galaxy A5	SM-A720	2017	7
Galaxy J7	SM-J730	2017	7
Galaxy Note Fan Edition	SM-N935	2017	7
Galaxy Wide 2	SM-J727	2017	7
Galaxy A8	SM-A530	2017	7.1.1
Galaxy Note 8	SM-N950	2017	7.1.1
Galaxy A5	SM-A520	2017	8
Galaxy S8	SM-G950	2017	9
Galaxy S8 plus	SM-G955	2017	9
Galaxy J2 pro	SM-J250	2018	7.1.1
Galaxy Jean (A6)	SM-A605	2018	8
Galaxy A7	SM-A750	2018	8
Galaxy J6	SM-J600	2018	8
Galaxy Wide 3	SM-J737	2018	8
Galaxy Note 9	SM-N960	2018	8.1.0
Galaxy S9	SM-G960	2018	9
Galaxy A50	SM-A505	2019	9
Galaxy Note 10	SM-N971	2019	9
Galaxy Wide 4	SM-A205	2019	9
Galaxy S10	SM-G973	2019	9
Galaxy S10e	SM-G970	2019	10
Galaxy A80	SM-A805	2019	10
Galaxy A90 5G	SM-A908	2019	10
Galaxy S20 5G	SM-G981	2020	10

3.2. 실험 과정

실험 과정은 그림 1과 같으며, 실험 과정은 크게 샘플 수집 과정(collecting samples), 메타데이터 데이터베이스 구축 과정(speech file metadata database construction), 메타데이터 비교 분석 과정(metadata comparative analysis)으로 구성하였다.

샘플 수집과정은 원본 음성 파일과 편집 음성 파일 생성을 통하여 수행하였고, 생성된 원본 및 편집 음성 파일에서 메타데이터를 추출하여 메타데이터 데이터베이스를 구축하였다. 구축된 메타데이터 데이터베이스를 이용하여 비교 분석을 실행하였다.

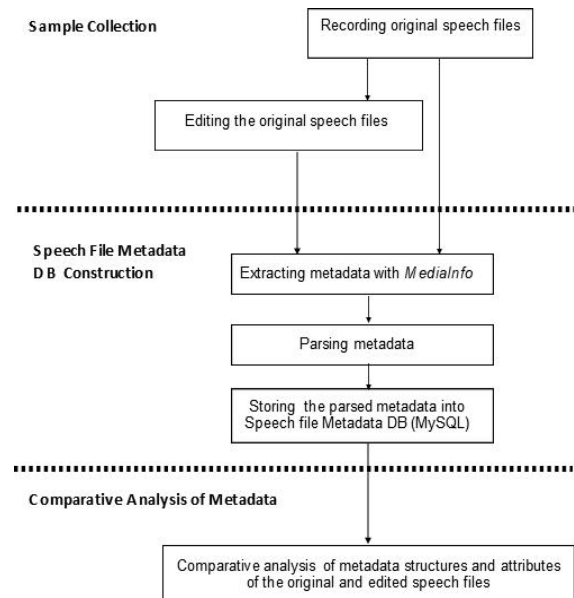


그림 1. 음성 파일의 메타데이터 분석 과정
Figure 1. Metadata analysis process of speech files

3.2.1. 원본 음성 파일 생성

실험에 사용한 삼성 스마트폰의 기종과 안드로이드 버전에 따라 스마트폰 자체 음성 녹음 및 편집 어플리케이션에서 지원하는 음성 녹음 모드에 차이가 있었다. 실험에 사용한 삼성 스마트폰에서 확인 가능한 음성 녹음 모드는 다음과 같다.

- ① 통화 녹음 모드: 실험에 사용한 스마트폰 모두에서 지원하고 있고 통화 중 음성 녹음 모드이며 단일 음질만 지원한다.
- ② 일반 음성녹음 모드: 2015년 초반에 출시된 갤럭시 A7(SM-A700)에서는 일반 및 고품질의 2가지 음질을 선택할 수 있으며, 이후 출시된 스마트폰에서 일반음성녹음모드에서는 3가지 음질(저, 일반, 고품질)을 선택할 수 있었다. 2015년 출시 갤럭시 A7(SM-A700)과 2018년 출시 A7(SM-750)은 다른 기기이다.
- ③ 인터뷰 모드: 음성 파일의 Recording channel이 2개라는 점에서 일반음성녹음모드와 차이가 있으며, 3가지 음질(저, 일반, 고품질)을 지원하였다.
- ④ 회의 모드: 안드로이드 6.0.1 초반에 지원했다 사라진 모드로 3가지 음질(MMS 첨부 모드, 일반, 고품질)을 지원하였다.

실험에 사용한 각 스마트폰에서 지원하는 가능한 모든 모드에서 음성 녹음을 진행하여 원본 음성 파일을 생성하였다. 34개 스마트폰에서 총 137개의 원본 음성 파일을 생성하였다. 이는 위변조 판별 외에 검증에 의해 제출되는 음성 파일의 녹음에 사용된 기종, 안드로이드 버전, 녹음 모드, 그리고 녹음 음질을 추가로 식별하기 위하여 필요한 메타데이터를 확보하기 위한 것이다.

3.2.2. 편집 음성 파일 생성

원본 음성 파일과 원본을 편집한 편집 음성 파일의 메타데이터 차이를 판별하기 위하여 편집 음성 파일의 수집이 필요하다.

갤럭시 스마트폰의 자체 음성 녹음 및 편집 어플리케이션에서 지원하는 편집 방식은 ‘선택한 영역 삭제’와 ‘선택되지 않은 영역 삭제’의 두 가지가 존재하였다.

편집 방식과 원본 파일의 녹음 모드 및 음질에 따라 메타데이터에 미치는 영향을 확인하기 위하여 2016년 출시 안드로이드 버전 8.0 갤럭시 S7, 2017년 출시 안드로이드 버전 9.0 삼성 갤럭시 S8, 2019년 출시 안드로이드 버전 11 갤럭시 S10e를 이용하여 사전 실험을 진행하였다.

사전 실험 결과, 상이한 녹음 모드로 생성된 음성 파일은 메타데이터의 `udat(User data)` 블록 내부 구조에서 유의미한 차이를 보였으며 상세한 결과는 ‘4.2.4. `udat(User data)` 블록 내부 구조 차이’절에서 상세히 설명하였다. 서로 다른 녹음 음질은 평균 비트 속도(average bitrate), 샘플 속도(sampling rate) 등에 영향을 미치지만, 이들이 메타데이터를 통한 위변조 판별에 유의미한 영향을 미치지 않음을 확인하였다. 또한, 편집 방식과 횟수에 있어서 편집 여부만 메타데이터에 유의미한 영향을 미칠 뿐 편집 방식과 횟수는 메타데이터에 유의미한 영향을 미치지 않는 점이 나타났다.

사전 실험 결과에 따라, 생성한 모든 원본 음성 파일에 대하여 편집 음성 파일을 생성하지 않고 일반 음성 녹음 모드에서 일반음질로 녹음된 원본 음성 파일에 대하여만 편집 음성 파일을 생성하였다.

편집 과정은 다음과 같다. 실험 기기를 통해 원본 음성 파일을 녹음한 후 해당 기기에 탑재되어있는 음성 녹음 및 편집 어플리케이션을 이용하여 원본 음성 파일의 중간 부분에서 앞뒤로 10초가량을 잘라내어 삭제하였으며, 이를 다른 이름으로 저장하였다.

3.2.3. 메타데이터 데이터베이스 구축

녹음 및 편집 과정을 통하여 생성된 원본과 편집 음성 녹음 파일에 대하여 그림 2와 같이 MediaInfo 프로그램(MediaArea, 2018)을 이용하여 XML 문서 형식으로 메타데이터를 추출하였다. 실험에 사용된 원본 녹음 파일 및 편집 파일은 모두 MPEG-4 컨테이너를 준수하는 파일이었다.

MPEG-4 메타데이터 구조(itom tree)는 메타데이터 블록과 속성이 나타나는 순서를 의미한다.

```
<block name="File Type" offset="0" size="24">
  <block name="Header" offset="0" size="8">
    <data name="Size" offset="0">24</data>
    <data name="Name" offset="4">ftyp</data>
  </block>
  <data name="MajorBrand" offset="8">3gp4</data>
  <data name="MajorBrandVersion" offset="12">0</data>
  <data name="CompatibleBrand" offset="16">icbm</data>
  <data name="CompatibleBrand" offset="20">3gp4</data>
</block>
<block name="Data" offset="24" size="509969">
  <block name="Header" offset="24" size="8">
    <data name="Size" offset="24">509969</data>
    <data name="Name" offset="28">mdat</data>
  </block>
  <data name="Data" offset="32">(509961 bytes)</data>
</block>
<block name="File header" offset="509993" size="6279">
  <block name="Header" offset="509993" size="8">
    <data name="Size" offset="509993">6279</data>
    <data name="Name" offset="509997">moov</data>
  </block>
```

그림 2. MediaInfo를 통해 XML 형태로 추출한 메타데이터 예시
Figure 2. A metadata example in XML format extracted by MediaInfo

블록(block)은 ‘청크(chunk)’, ‘박스(box)’, ‘아톰(itom)’으로 불리기도 하며 그림 2(1)과 같이 “<block ..>”으로 나타난다. 하나의 블록은 블록의 이름(name), 위치(또는 주소, offset), 그리고 크기(size)로 표시된다. 블록 내부에는 하위 블록과 속성이 나타날 수 있다. 본 연구에서는 편의상 A 블록의 하위에 존재하는 B 블록을 ‘A>B’로 표기한다.

속성(attribute)은 그림 2(2)와 같이 “<data ..>”로 나타나며, 속성의 이름(name)과 위치(offset), 그리고 속성의 값이 표시된다.

반정형 형태의 데이터인 음성 파일 메타데이터를 정형화하여 MySQL를 통하여 구축한 데이터베이스에 저장하였다. 해당 과정은 JAVA를 통하여 자체적으로 구축한 소프트웨어를 통하여 진행되었다.

4. 메타데이터 비교 분석 결과

4.1. 안드로이드 6.0.1 이상 음성 녹음 파일의 특징

안드로이드 6.0.1 이상의 음성 파일들은 `m4a` 파일 확장자를 가진다는 점과, MPEG-4 컨테이너 구조를 따른다는 점의 공통점이 있다. MPEG-4 컨테이너는 실제 오디오 정보 외 다양한 메타데이터를 저장함으로써 무결성 분석에 유용한 메타데이터 정보를 다수 포함하고 있다(Marcin, 2019). MPEG-4 컨테이너의 구조 및 속성의 해석은 MediaInfo(MediaArea, 2018)와 QuickTime File Format Specification(Apple, 2016)을 따랐다.

4.2. 메타데이터 구조 분석

구조 분석에서는 블록의 유무, 순서를 통해 메타데이터를 비교한다. 동일 인코딩 소프트웨어 및 설정을 통하여 녹음하였을 경우, `mdat(Media data)` 하위에 존재하는 실제 오디오 데이터와 오디오 데이터 분할 단위인 샘플에 대한 정보를 나타내는 `stbl(Sample table)`의 하위 블록 및 속성을 제외하고 메타데이터 구조 면에서 일치해야 한다.

그림 3은 삼성 갤럭시 노트 8(안드로이드 버전 7.1.1)의 일반 음성 녹음 모드로 생성한 원본 음성 파일의 메타데이터 구조를 MP4 Inspector(Codeine, 2016)를 이용하여 추출한 것이다.

원본과 편집 음성 파일의 메타데이터 비교를 통하여 다음과 같은 메타데이터 구조 및 속성값의 차이가 존재함을 알 수 있었다.

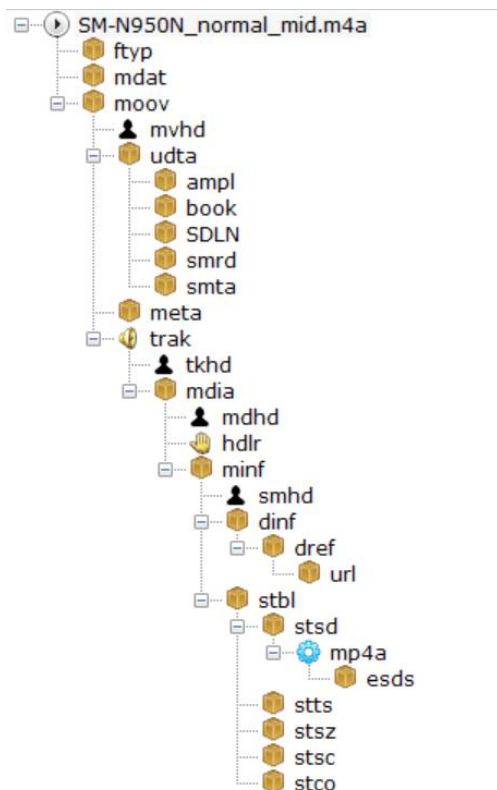


그림 3. 삼성 갤럭시 노트 8 일반 음성녹음 모드 일반 음질 원본 파일 메타데이터 구조

Figure 3. Metadata structure of the original speech file recorded by Samsung Galaxy Note 8

4.2.1. 최상위 블록 구조의 변화

삼성 스마트폰에서 생성한 원본 음성 파일 메타데이터의 최상위 블록은 ftyp(File type), mdat(Media data), moov(File header)로 구성되어 있다(그림 3 참고). 2015년 이전 출시된 안드로이드 6.0.1 기종에서 편집된 음성 녹음 파일의 메타데이터에서는 원본 파일의 메타데이터에는 존재하지 않는 skip(Skip) 블록이 최상위 블록으로 생성된다.

4.2.2. trak(Track)과 udta(User data) 블록의 위치 변경

2015년 이전 출시된 안드로이드 6.0.1 기종에서는 음성 파일을 편집할 경우, moov(File header) 블록 내부 trak(Track)과 udta(User data) 블록의 순서가 원본 순서인 udta, meta(Metadata), trak인 것과 반대로 trak, udta순으로 나타난다.

4.2.3. meta(Metadata) 블록의 유실

그림 4와 같이 meta(Metadata) 블록은 녹음 파일에 대하여 편

집이 수행될 경우 사라진다. meta 블록 내부에는 안드로이드 버전 정보가 존재하기 때문에, 6.0.1 이상 삼성 원본 음성 파일은 자체 파일을 통해 녹음 기종의 OS 정보를 확인할 수 있지만, 편집을 수행할 경우, OS 정보를 알 수 없게 된다.

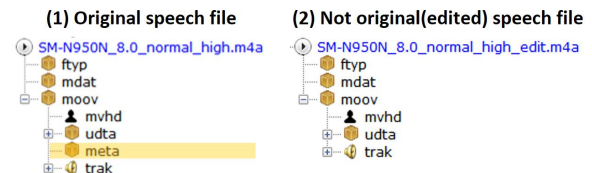


그림 4. 원본 음성 파일과 편집 음성 파일의 meta(Metadata) 블록 차이
Figure 4. meta (Metadata) block difference between the original and edited speech file

4.2.4. udta(User data) 블록 내부 구조 차이

udta(User data) 블록은 음성 녹음 및 편집 소프트웨어 인코더 단에서 자율적으로 확장이 가능한 블록으로서 인코딩 소프트웨어의 특성이 반영된다. udta 블록 내부에서 나타나는 SDLN, smrd, smta 블록은 삼성 스마트폰의 자체 탑재 음성녹음 및 편집 소프트웨어에서 녹음 수행 시 파일 내부에서 나타나는 블록으로 알려져 있으며(Michalek, 2019), udta 블록은 그 자체가 제조사 고유 메타데이터를 표시할 수 있는 블록이다. 삼성 자체 음성녹음 및 편집 소프트웨어에서 어떤 음성 녹음 모드(기종 및 버전에 따라 설정 가능한 모드는 상이)로 생성하였는지에 따라 udta 블록 내부는 상이한 구조와 속성을 가지게 된다.




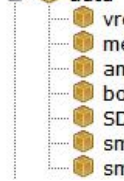

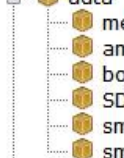
표 2는 현재 식별한 녹음 모드에 따른 udta 블록의 내부 구조의 차이를 정리한 것이다. 안드로이드 버전 5.0.1에서 10까지의 통화 녹음 모드(call recording)에서 생성한 파일의 udta 블록은 하위 블록으로 SDLN, smrd, smta의 세 개가 순서대로 나타난다. 또한 갤럭시 노트 에지(안드로이드 6.0.1)의 일반 음성녹음 모드(normal mode recording)로 생성한 파일도 동일한 구성을 가진다. 그러나 갤럭시 노트 에지(안드로이드 6.0.1)의 인터뷰 모드(interview mode)와 회의 모드(meeting mode)에서 생성한 파일은 udta 블록 내부에 metd 하위 블록이 추가로 나타남을 알 수 있다.

녹음 모드에 따른 이와 같은 차이는 위변조 분석에 유의미하게 작용하지는 않지만, 증거로 제출된 파일이 어떤 기기의 어떤 모드에서 생성되었는지를 유추하는 데 중요한 요소이다.

4.3. 메타데이터 속성 분석

속성 분석에서는 원본 음성 파일과 편집 음성 파일 간의 속성의 일치 여부를 검사한다. MPEG-4에는 매우 다양한 속성이 존재한다. 원본 파일과 비교를 통한 녹취 파일의 위변조 판별에는 유의미한 속성과 무의미한 속성이 존재하며, 무의미한 속성의 예시에는 음성 파일의 길이를 나타내는 Duration 속성, mdat(Metadata)의 Size 속성 등과 편집을 수행함에도 변하지 않는 속성값 등이 있다. 다음은 위변조 판별에 유의미한 편집 시의 속성 변화를 나열하였다.

표 2. udta(User data) 블록 내부 구조
Table 2. udta (User data) block internal structure

udta Structure	Device and recording mode
	Android version 5.0.1-10 call recording, Android version 6.0.1 Galaxy Note Edge normal mode recording
	Android version 6.0.1 Galaxy Note Edge interview and meeting mode recording
	Android version 6.0.1-8.1.0 normal mode recording
	Android version 8.0-10 interview mode recording and, Galaxy Note 10 normal mode recording
	Android version 8.0-10 normal mode recording
	Android version 7.0-8.1.0 interview mode recording

4.3.1. Major brand와 Compatible brand 속성의 변화

ftyp(File type) 블록 내부에서 나타나는 속성과 순서는 모든 원본 음성 파일에서 동일하게 나타난다. Major brand 속성의 값은 '3gp4'이며, Major brand version의 속성은 0, Compatible brand 속성이 2개 존재하며 각각 순서대로 'isom', '3gp4'로 나타난다. 편집을 수행할 경우, 안드로이드 6.0.1 이상에서는 해당 값들이 동일하게 유지되지만, 일부 안드로이드 6.0.1과 안드로이드 6.0.1 미만 기종에서는 Major brand의 값은 '3gp7', Major version의 값은 768, Compatible brand의 값은 '3gp6'과 'isom' 또는 '3gp7'과 'isom'으로 변경되었다.

4.3.2. Time scale 속성의 변화

moov 내부 mvhd(Movie header) 블록에서 나타나는 Time scale 속성의 경우 그림 5와 같이 원본 음성 파일에서는 해당 값이 1,000으로 나타나지만, 편집 음성 파일의 경우 해당 속성이 1,000이 아닌 44,100, 48,000 등으로 나타난다. 해당 현상은 안드로이드 6.0.1 이하 또는 안드로이드 11 이상의 경우에서 나타나지 않으며, 안드로이드 6.0.1 이하의 경우 편집 음성 파일의 Time scale이 1,000으로 나타나는 경우가 존재하며, 안드로이드

11 이상의 경우 원본 음성 파일의 Time scale값이 10,000이며 편집 시 Time scale값이 10,000이 아닌 값으로 변경된다.

moov→trak→mdia(Media)→mdhd(Movie header) 블록 내부의 Time scale 속성과 혼동하지 않도록 유의해야 한다.

(1) Original speech file

```

6A0D3 Movie header (108 bytes)
6A0D3 Header (8 bytes)
6A0D3 Size: 108 (0x0000006C)
6A0D7 Name: mvhd
6A0D8 Version: 0 (0x00)
6A0DC Flags: 0 (0x000000)
6A0DF Creation time: 3638148084 (0xD8D98BF4)
6A0E3 Modification time: 3638148084 (0xD8D98BF4)
6A0E7 Time scale: 1000 (0x000003E8) - 1000 Hz
6A0EB Duration: 27307 (0x00006AAB) - 27307 ms
6A0EF Preferred rate: 65536 (0x00010000) - 1.000
6A0F3 Preferred volume: 256 (0x0100) - 1.000
6A0F5 Reserved: (10 bytes)
  
```

(2) Not original(edited) speech file

```

68DE1 Movie header (108 bytes)
68DE1 Header (8 bytes)
68DE1 Size: 108 (0x0000006C)
68DE5 Name: mvhd
68DE9 Version: 0 (0x00)
68DEA Flags: 0 (0x000000)
68DED Creation time: 3676020249 (0xDB1B9E19)
68DF1 Modification time: 3676020249 (0xDB1B9E19)
68DF5 Time scale: 44100 (0x00000C44) - 44100 Hz
68DF9 Duration: 1183744 (0x000121000) - 26842 ms
68DFD Preferred rate: 65536 (0x00010000) - 1.000
68E01 Preferred volume: 256 (0x0100) - 1.000
68E03 Reserved: (10 bytes)
  
```

그림 5. mvhd(Movie header) 블록 일부
Figure 5. Part of the mvhd (Movie header) block

4.3.3. tkhd(Track header) 블록 내부 속성들의 변화

moov>trak>tkhd(Track header) 블록 내부에는 Flags, Create time, Modification time, Volume 등의 속성이 그림 6과 같이 존재한다. Flags 속성의 값은 원본 음성 파일에서 7로 나타나지만, 편집 파일의 경우 1 또는 15로 나타난다. Volume 속성은 원본 음성 파일에서는 256으로 나타나지만, 편집 음성 파일에서는 0으로 변경된다. 원본 음성 파일의 경우 Creation time과 Modification time이 일치하지만, 편집 음성 파일의 경우 Modification time의 속성이 변경되어 Creation time과 Modification time이 일치하지 않는다. Volume과 Creation time, Modification time 변경 현상은 2015년 이전 출시된 안드로이드 6.0.1의 경우 나타나지 않는다.

(1) Original speech file

97446	Track Header - 1 (0x1) - 36525 (0x8EAD) ms (92 bytes)	
97446	Header (8 bytes)	
97446	Size:	92 (0x0000005C)
97446	Name:	tkhd
9744E	Version:	0 (0x00)
9744F	Flags:	7 (0x00000007)
97452	Track Enabled:	Yes
97452	Track in Movie:	2 (0x0000000000000002)
97452	Track in Preview:	4 (0x0000000000000004)
97452	Track in Poster:	0 (0x0000000000000000)
97452	Creation time:	3676019251 (0xDB1B9A33)
97456	Modification time:	3676019251 (0xDB1B9A33)
97472	Volume:	256 (0x0100) - 1.000
97474	Reserved:	0 (0x0000)

(2) Not original(edited) speech file

7124A	Track Header - 1 (0x1) - 26842 (0x68DA) ms (92 bytes)	
7124A	Header (8 bytes)	
7124A	Size:	92 (0x0000005C)
7124E	Name:	tkhd
71252	Version:	0 (0x00)
71253	Flags:	15 (0x0000000F)
71256	Track Enabled:	Yes
71256	Track in Movie:	2 (0x0000000000000002)
71256	Track in Preview:	4 (0x0000000000000004)
71256	Track in Poster:	8 (0x0000000000000008)
71256	Creation time:	3676019251 (0xDB1B9A33)
7125A	Modification time:	3676020249 (0xDB1B9E19)
71276	Volume:	0 (0x0000) - 0.000
71278	Reserved:	0 (0x0000)

그림 6. tkhd(Track header) 블록 일부

Figure 6. Part of the tkhd (Track header) block

4.3.4. Number of entries 속성의 변화

moov→trak→tkhd→mdia→minf(Media information)→stbl(Sample table)→stsc(Sample to chunk) 블록 내부의 Number of entries 속성은 원본에서는 1이지만, 편집을 수행할 경우 1이 아닌 값으로 변경된다.

4.3.5. Language 속성의 변화

moov→trak→tkhd→mdia→mdhd 블록 내부의 Language 속성은 원본에서 모두 0(English)으로 나타나지만, 편집을 수행한 경우 일부 안드로이드 6.0.1 기종의 경우 해당 값이 21,956(undefined)으로 변경되었다.

4.3.6. Component name 속성의 유실

원본 음성 파일의 경우 그림 7(1)과 같이 moov→trak→tkhd→mdia→mdhd→hdlr(Handler reference)블록 내부에 component name 속성이 존재하며, 해당 값은 SoundHandle로 나타나지만, 편집 음성 파일의 경우 그림 7(2)와 같이 해당 속성이 나타나지 않는다.

(1) Original speech file

B040C	Handler Reference (44 bytes)	
B040C	Header (8 bytes)	
B040C	Size:	44 (0x0000002C)
B0410	Name:	hdlr
B0414	Version:	0 (0x00)
B0415	Flags:	0 (0x000000)
B0418	Component type:	
B041C	Component subtype:	sound
B0420	Component manufacturer:	
B0424	Component flags:	0 (0x00000000)
B0428	Component flags mask:	0 (0x00000000)
B042C	Component name:	SoundHandle

(2) Not original(edited) speech file

712CE	Handler Reference (33 bytes)	
712CE	Header (8 bytes)	
712CE	Size:	33 (0x00000021)
712D2	Name:	hdlr
712D6	Version:	0 (0x00)
712D7	Flags:	0 (0x000000)
712DA	Component type:	
712DE	Component subtype:	sound
712E2	Component manufacturer:	
712E6	Component flags:	0 (0x00000000)
712EA	Component flags mask:	0 (0x00000000)
712EE	Component name size (not in specs):	0 (0x00)

그림 7. hdlr(Handler reference) 블록 일부

Figure 7. Part of the hdlr (Handler reference) block

4.4. 음성 녹음 파일 원본과 편집본 식별 결정 트리

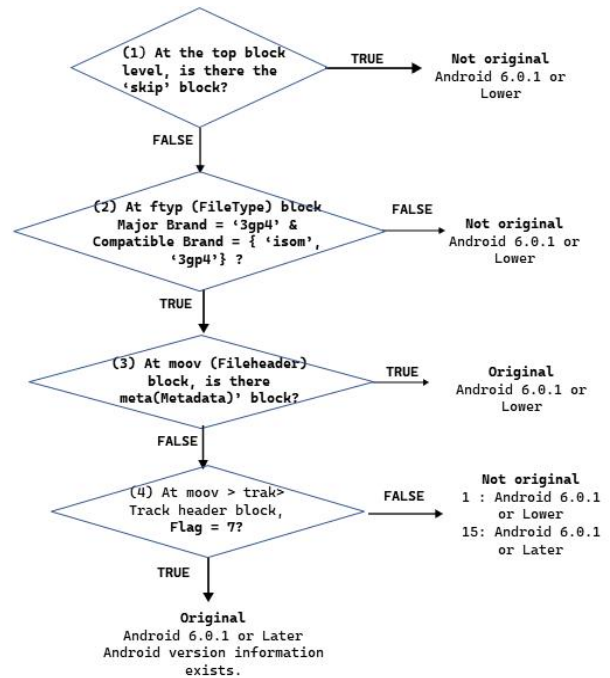


그림 8. 원본, 편집 음성 파일 결정 트리

Figure 8. Decision tree to determine the original speech files

4.2.절 메타데이터 구조 분석에서 도출한 구조 차이 4가지(4.2.1.-4.2.4.절 참고)와 4.3.절 메타데이터 속성 분석에서 도출한 속성 값의 차이 6가지(4.3.1.-4.3.6.절 참고)를 종합하여 2015년부터 2020년 상반기까지 출시된 삼성 스마트폰 원본 음성 파일과 편집 음성 파일의 식별을 위한 결정 트리를 그림 8과 같이 개발하였다. 그림 8의 결정 조건 4가지는 다음과 같다.

- (1) 최상위 블록단에 skip(Skip) 블록의 존재 여부(4.2.1. 최상위 블록 구조의 변화)
- (2) Major brand의 속성값이 3gp4 이며, Compatible brand의 값이 'isom', '3gp4'인지 여부(4.3.1. Major brand와 Compatible brand 속성의 변화)
- (3) File header 블록 내부에 meta(Metadata) 블록의 존재 여부 [4.2.3. meta(Metadata) 블록의 유실]
- (4) Track>Track header 내부 flag값이 7인지 여부[4.3.3. tkhd (Track header) 블록 내부 속성들의 변화]

위 4가지 결정 조건에 포함되지 않은 구조와 속성값의 차이는 모두 위 4가지 조건의 검사로 해결되므로 굳이 별도의 검사가 필요하지 않았다.

(1)의 조건이 참일 경우, 해당 파일은 안드로이드 6.0.1 이하의 편집본으로 판단할 수 있다. 해당 조건이 거짓일 경우 (2)로 진행한다.

(2)의 조건이 거짓일 경우, 해당 파일은 안드로이드 6.0.1 이하의 편집본으로 판단할 수 있다. 해당 조건이 참일 경우 (3)으로 진행한다.

(3)의 조건이 참일 경우, 안드로이드 6.0.1 이상의 원본으로 판별할 수 있다. 또한, 이 경우 별도의 비교 파일이 없더라도 자체적으로 파일 녹음 기종의 OS 버전을 알 수 있다. 해당 조건이 거짓일 경우 (4)로 진행한다.

(4)의 조건이 참일 경우, 6.0.1 이하의 원본으로 판별할 수 있다. 해당 값이 1로 나타날 경우, 안드로이드 6.0.1 이하의 편집본으로 판별할 수 있다. 해당 값이 15일 경우, 안드로이드 6.0.1 이상의 편집본으로 판별할 수 있다.

5. 결론 및 논의

스마트폰의 대중화로 증거로 제출되는 음성 파일의 대부분이 스마트폰에서 생성되고 있으며, 국내 시장의 가장 큰 점유율을 보이는 삼성 스마트폰에서 생성 또는 편집된 음성 파일의 위변조 식별은 수사와 재판에서 중요한 이슈가 되고 있다.

음성 파일의 메타데이터 분석을 통한 위변조 식별은 디지털 음성 신호 데이터의 정밀 분석을 통한 식별에 비하여 노력과 시간에서 큰 장점을 가진다.

본 연구에서는 2015년에서 2020년 사이에 출시된 삼성 스마트폰에서 자체 지원하는 음성 녹음 어플리케이션에서 생성한 음성 녹음 파일 원본과 편집한 음성 파일의 식별을 위한 음성 파일 메타데이터 비교 분석을 수행하였다. 먼저, 삼성 스마트폰의 자체 음성 녹음 어플리케이션에서 지원하는 다양한 녹음 모드(통화녹음, 일반음성녹음, 인터뷰녹음 등)에서 원본 음성 파일을 생성하고, 이를 편집한 파일을 생성한 다음, 생성된 원본 및 편집 음성 파일에서 메타데이터를 추출하여 메타데이터 데이터베이스 구축하였다. 구축한 메타데이터 데이터베이스를 이용하여 메타데이터의 비교 분석을 실시하여, 그 결과로 녹취한 스마트폰의 모델과 녹음 모드, 그리고 편집 여부의 식별에

유용한 메타데이터 상의 유의미한 차이가 존재함을 알아내었다. 이러한 유의미한 차이를 기반으로 편집 여부를 판별할 수 있는 간단한 결정 트리를 도출하였다. 구축한 메타데이터 데이터베이스는 증거 녹음 파일의 녹취에 사용된 스마트폰을 확보하지 못한 경우에도 해당 스마트폰의 기종과 안드로이드 버전을 알 경우, 해당 증거 녹음 파일의 진위를 판단할 수 있도록 한다.

현재까지 확보한 삼성 스마트폰 기종은 2015년에서 2020년 상반기까지 출시된 것으로 새로운 스마트폰 기종의 출시, 안드로이드 OS의 지속적인 버전 업그레이드가 이루어지고 있다. 본 연구의 분석 결과 및 위변조 식별 결정 트리의 유효성을 확보하기 위하여 스마트폰의 원본 음성 녹음 파일 및 편집 파일의 확보, 메타데이터 데이터베이스 확장, 비교 분석 연구가 지속적으로 필요하다.

References

- Apple. (2016). QuickTime file format specification. Retrieved from <https://developer.apple.com/library/archive/documentation/QuickTime/QTFF>
- Codeine, W. (2016). MP4 Inspector [Computer program]. Retrieved from <https://sourceforge.net/projects/mp4-inspector/>
- Dan, O. N. (2020). ID3 the audience is informed. Retrieved from <https://id3.org/ID3v1>
- Grigoras, C., & Smith, J. (2017, June). Large scale test of digital audio file structure and format for forensic analysis. *Proceedings of the 2017 AES International Conference on Audio Forensics*. Arlington, VA.
- Kim, K. W. (2021). A study on the forensic application of smartphone recording database. *Journal of Digital Forensics*, 15(1), 26-42.
- Koenig, B. E., & Lacey, D. S. (2017, June). Forensic authenticity analyses of the metadata in re-encoded iPhone M4A files. *Proceedings of the 2017 AES International Conference on Audio Forensics*. Arlington, VA.
- MediaArea. (2018). MediaInfo (version CLI.21.03) [Computer program]. Retrieved from <https://mediainfo.net/ko/MediaInfo>
- Michalek, M. (2019). The characteristics of popular audio recording applications installed on smartphones with an android operating system in relation to forensic audio analysis. *Problem of Forensic Sciences*, 120, 335-361.
- Park, N. I., Lee, J. W., Kim, J. H., Lim, J. S., Na, G. H., & Jeon, O. (2022). Forensic analysis method for forgery detection of call recordings generated by Samsung smartphones. *Journal of Digital Forensics*, 16(1), 142-159.
- Park, N. I., Shim, K. S., & Jeon, O. Y. (2019). A study on authentication analysis procedure of digital audio files. *Journal of Digital Forensics*, 23(4), 257-269.
- Statcounter GlobalStats. (2021). Mobile vendor market share in

Republic of Korea: February 2021. Retrieved from <https://gs.statcounter.com/vendor-market-share/mobile/south-korea/#monthly-202102-202102-bar>

Zeng, J., Lian, Q., & Shi, S. (2019). Forensic originality identification of iPhone's voice memos. *Journal of Physics: Conference Series*, 1345(2019), 052053.

• 안서영 (Seo-Yeong Ahn)

성신여자대학교 미래융합기술공학과 석사과정

서울시 성북구 보문로 30길 50

Tel: 02-920-7895

Email: 98_0731@honglab.org

관심분야: 음성 사용자 인터페이스, 보완 대체 의사소통

• 유세희 (Se-Hui Ryu)

성신여자대학교 미래융합기술공학과 석사과정

서울시 성북구 보문로 30길 50

Tel: 02-920-7895

Email: rsh725@honglab.org

관심분야: 음성 사용자 인터페이스, 보완 대체 의사소통

• 김경화 (Kyung-Wha Kim)

대검찰청 과학수사부 음성감정관

서울시 서초구 반포대로 157

Tel: 02-3480-2150

Email: savoix@spo.go.kr

관심분야: 법음성학, 화자 인식

• 홍기형 (Ki-Hyung Hong) 교신저자

성신여자대학교 서비스디자인공학과 교수

서울시 성북구 보문로 30길 50

Tel: 02-920-7525

Email: khhong@sungshin.ac.kr

관심분야: 음성 사용자 인터페이스, 디지털 정보시스템

법과학적 활용을 위한 삼성 스마트폰 음성 녹음 파일의 메타데이터 구조 및 속성 비교 분석 연구*

안 서 영¹ · 유 세 희¹ · 김 경 화² · 홍 기 형³

¹성신여자대학교 미래융합기술공학과, ²대검찰청 과학수사부, ³성신여자대학교 서비스디자인공학과

국문초록

스마트폰의 대중화로 인하여 근래 범죄의 증거자료로 제출되는 녹취 파일은 대부분 스마트폰을 통하여 생산되고 있으며, 스마트폰을 기반으로 한 녹음 파일의 무결성(위변조) 여부가 수사와 재판 과정에서 주요 쟁점으로 떠오르고 있다. 가장 높은 국내 시장 점유율을 가진 삼성 스마트폰은 통화 및 음성 녹음, 그리고 편집이 가능한 자체 음성 녹음 편집 어플리케이션이 탑재되어 유통되고 있으며, 자체 어플리케이션을 통한 편집은 외부 어플리케이션을 통한 편집과 다르게 원본 파일과의 유사성이 높기에, 무결성을 입증하기 위해 더 정밀한 분석 기법 개발이 필요하다. 본 연구에서는 삼성 스마트폰 34개 기종에서 생성된 원본 녹음 파일과 자체 제공 음성녹음 편집 어플리케이션을 통한 편집 파일의 메타데이터 구조와 속성을 분석하여, 원본과 편집본 사이의 음성 파일 메타데이터 구조 및 속성 값에서 유의미한 차이가 있음을 확인하였다.

핵심어: 음성 메타데이터, 음성 녹음, 음성 위변조 탐지, MPEG-4

참고문헌

- 김경화(2021). 스마트폰 녹음 데이터베이스의 법과학적 활용에 대한 연구. *디지털포렌식연구*, 15(1), 26-42.
- 박남인, 심규선, 전옥엽(2019). 디지털 오디오 파일의 편집 여부 분석 절차에 대한 연구. *디지털포렌식연구*, 13(4), 257-270.
- 박남인, 이지우, 김진환, 임재성, 나기현, 전옥엽(2022). 삼성 스마트폰에서 생성된 통화녹음파일에 대한 위변조 검출을 위한 법과학적 분석 방법. *디지털포렌식연구*, 16(1), 142-159.

* 이 논문은 2021년 대검찰청 연구 용역의 지원으로 수행되었습니다(과제명: 스마트폰 녹음 파일의 위변조 여부 분석 기법 개발).