

<http://dx.doi.org/10.17703/JCCT.2023.9.2.545>

JCCT 2023-3-67

스마트폰 음성 녹음 파일 위변조 검출을 위한 스펙트로그램 분석의 한계점

Limitations of Spectrogram Analysis for Smartphone Voice Recording File Forgery Detection

한상민*, 손영민**, 박재완***

Sangmin Han*, Yeongmin Son**, Jae Wan Park***

요약 오늘날 누구나 디지털 정보를 용이하게 활용할 수 있게 됨에 따라 디지털 증거의 채택이 증가되고 있다. 하지만 다양한 음성 파일 편집 도구를 보급과 함께 정교한 편집 과정을 거친 음성 녹음 파일의 경우 위변조 진위 여부를 판단하는 것은 사실상 불가능하다. 본 연구는 음성 녹음 파일에 삽입, 삭제, 연결 및 합성 편집 기술을 활용해 원본 파일과 구별하기 어려운 위변조가 가능함을 증명하고자 한다. 본 연구는 위변조 된 음성 파일을 원본과 동일한 확장자로 인코딩하는 작업을 통해 위변조 검출의 어려움을 제시한다. 또한 특징점이 발생한 실험에 한 하여 추가적으로 천이대역의 삭제 및 2차 인코딩 작업을 수행할 경우 위변조 검출은 불가능함을 나타냈다. 이를 통해 본 연구는 음성 녹음 파일을 디지털 증거로 채택하기 위한 더 엄격한 증거능력 판단 기준 수립에 공헌할 것으로 기대된다.

주요어 : 위변조 기법, 스펙트로그램, 천이 대역, 디지털 증거, 음성 녹음 파일

Abstract As digital information is readily available to everyone today, the adoption of digital evidence is increasing. However, it is virtually impossible to determine the authenticity of forgery in the case of a voice recording file that has gone through a sophisticated editing process along with the spread of various voice file editing tools. This study aims to prove that forgery, which is difficult to distinguish from the original file, is possible by using insertion, deletion, linking, and synthetic editing technologies in voice recording files. This study presents the difficulty of detecting forgery by encoding a forged voice file with the same extension as the original. In addition, it was shown that forgery detection is impossible if additional transition band deletion and secondary encoding are performed only for experiments in which features occurred. Through this, this study is expected to contribute to the establishment of more stringent evidence admissibility criteria for adopting voice recording files as digital evidence.

Key words : Forgery Technique, Spectrogram, Transition band, Digital Evidence, Voice Recording File

I. 서론

최근 정보화 사회의 가속화가 진행됨에 따라 디지털 형태의 데이터가 방대해지고 있다. 이와 함께 사진, 녹취록 및 동영상 파일 등 다양한 디지털 기록을 법정 증

*준회원, 송실대학교 글로벌미디어학부 학사과정 (제1저자)
**준회원, 송실대학교 미디어학과 석사과정 (참여저자)
***정회원, 송실대학교 글로벌미디어학부 부교수 (교신저자)
접수일: 2023년 1월 31일, 수정완료일: 2023년 3월 1일
게재확정일: 2023년 3월 9일

Received: January 31, 2023 / Revised: March 1, 2023

Accepted: March 9, 2023

***Corresponding Author: jaewan.park@ssu.ac.kr
Global School of Media, Soongsil Univ., Korea

거로 제출되는 사례가 증가하고 있다[1]. 이러한 디지털 증거가 채택되기 위해 디지털 증거는 무결성이 입증될 필요가 있다[2]. 하지만 디지털 증거 중 녹취 음성 파일의 경우 누구나 손쉽게 어도비 오디오션(adobe audition)을 비롯한 다양한 음성 파일 편집 도구를 이용한 편집이 가능하다[3]. 이러한 편집 도구를 통해 음성 파일이 정밀하게 편집될 경우 위변조 여부를 검증하는 것은 불가능하다[4].

음성 파일의 위변조 기술 발전과 함께 이를 검출해 내는 연구가 시도되었다. 객관적 지표로 여겨지는 음성 파일의 메타데이터, 파일의 구조에 의존한 연구들이 진행되었다[5][6][7]. 또한 음성 녹음 파일의 메타데이터와 파일 구조는 위변조가 가능함을 밝힌 연구가 최근에 수행되었다[8]. 원본 파일과 위변조된 파일의 메타데이터와 파일 구조가 동일하다면 스펙트로그램(spectrogram) 기반 음성 신호의 분석을 통해 위변조 여부를 분석하게 된다. 하지만 이 또한 위변조 여부를 판단하기 위한 객관적 지표가 되기에 충분하지 않다.

따라서 본 연구는 다양한 실험을 통해 스마트폰에서 녹음된 음성 파일의 위변조 수행 시 스펙트로그램을 통해 검출이 불가능하다는 것을 증명하고자 한다. 본 연구를 위해 먼저 음성 파일의 편집 기법들과 스펙트로그램에 대한 이론적 고찰을 수행하고, 편집 기법이 적용된 음성 파일의 스펙트로그램을 이용하여 다양한 실험을 수행한다.

본 연구의 실험 결과는 정교하게 편집된 스마트폰 음성 파일과 원본 데이터와의 유의미한 차이가 없음을 나타낸다. 본 연구는 위변조 검출이 불가능함을 증명함으로써 형사소송법 제 307조 2항의 ‘범죄 사실의 인정이 합리적 의심이 없는 정도의 증명에 이르러야 한다’고 규정하고 있듯이[9], ‘음성녹취 파일이 디지털 증거로 채택됨에 있어서 더 엄격한 증거능력 판단 기준이 필요함을 주장하는 바이다’[8].

II. 이론적 고찰

1. 음성 파일 편집 기법

음성 파일의 편집 방법은 그림 1과 같이 ‘삽입(insert), 삭제(delete), 연결(connect), 합성(overlap)’으로 구분된다[8].

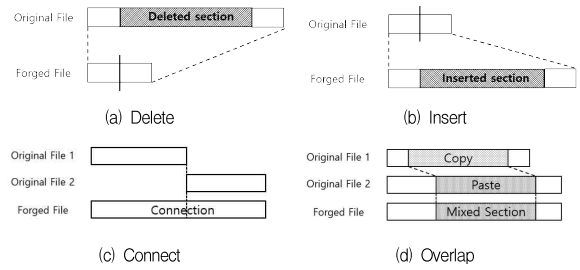


그림 1. 삭제, 삽입, 연결 및 합성 편집 방법
Figure 1. Techniques of Deletion, Insertion, Connection, and Overlap

첫 번째, 삽입은 기존의 음성 파일에 타 영역의 음성 프레임을 혹은 다른 파일의 음성 프레임을 복사(copy)하고 붙여넣기(paste)를 통해 수행된다. 사용자가 삽입을 수행하게 되면 삽입된 구간에 새로운 시간 프레임이 더해진다. 두 번째, 삭제는 기존의 음성 파일에 시간 프레임을 제거하는 방법이다. 세 번째 연결 편집 방법은 서로 다른 두 개 이상의 음성 파일을 멀티트랙(multi-track) 기능을 활용하여 연결하는 방법이다. 마지막으로 합성은 음성을 오버랩(overlap) 할 수 있는 기능으로써 기존 시간 프레임을 유지하며 크로스페이드(crossfade) 및 볼륨 조절을 이용한 편집이 가능하다.

2. 스펙트로그램

스펙트로그램이란 비가시적인 음성 데이터의 효율적인 분석을 위해 시간(time), 주파수(frequency), 진폭(amplitude)의 정보를 담아 시각화한 그래프이다. x축은 시간, y축은 주파수, 색상은 진폭을 나타낸다[10](그림 2).

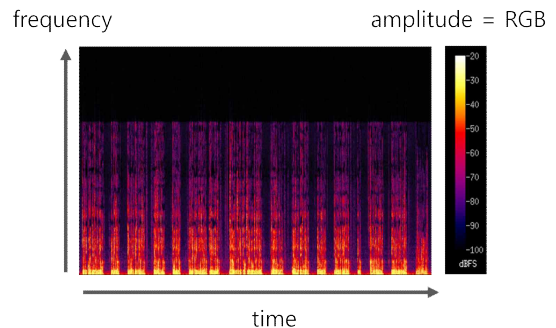


그림 2. 스펙트로그램 이미지
Figure 2. Spectrogram image

III. 실험 방법

본 연구에서는 녹음 시 샘플레이트(sample rate) 48kHz를 갖는 아이폰 XS MAX(IOS 16.02)와 갤럭시 S10(안드로이드 버전 12)을 이용한다. 아이폰에서 녹음된 음성 파일을 기반으로 어도비 오디션(adobe audition) 분석 도구를 이용하여 다양한 오디오 편집 기능을 적용한 후, 인코딩된 파일과 원본 음성 파일의 천이 대역과 가청 주파수 영역에서 스펙트로그램을 비교해 보는 실험을 진행했다. 천이 대역의 스펙트로그램을 효과적으로 분석하기 위해 증폭한다.

아이폰과 갤럭시 기기를 사용하여 녹음된 파일의 경우 m4a(MPEG-4 Audio) 확장자로 저장된다. 하지만 어도비 오디션을 통해 편집된 오디오 파일은 압축되지 않은 원시 오디오를 저장하기 위해 wav(Waveform audio file format) 확장자가 이용된다. 따라서 편집된 음성 파일은 원본 음성 파일의 확장자와 동일한 확장자가 되어야하므로 원본 포맷과 동일한 m4a 형태로 인코딩하는 과정을 갖는다(그림 3).

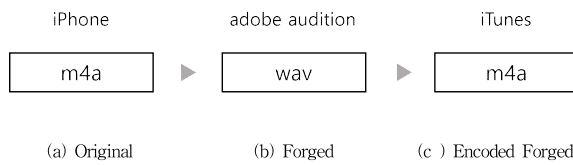


그림 3. 파일 변환 과정
 Figure 3. File Conversion Process

해당 실험을 위해 1분 전후 길이의 음성을 녹음한 후 음성 편집 프로그램인 어도비 오디션에서 원본 파일을 불러온다. 그리고 해당 파일을 이용하여 삭제, 삽입, 연결 그리고 합성 편집을 수행한 후 wav 파일로 저장한다. wav 파일은 다시 맥북의 ‘음악(Music)’ 소프트웨어(버전 1.1.6.37)를 활용해 인코딩 작업을 진행한다. 인코딩 작업 시, 아이폰 인코딩과 동일한 ‘AAC 인코더’, ‘좋은 품질(128bps)’로 설정한 후 m4a 파일로 변환하는 작업을 수행한다. 이로써 원본에 해당하는 기존 m4a 파일이 wav 파일을 거쳐 다시 최종적으로 원본 파일과 동일한 m4a 파일로 변환된다.

또한, 삭제 편집을 제외한 나머지 삽입, 연결 및 합성 편집 수행 시 갤럭시에서 녹음된 음성 파일을 활용하여 편집이 가능한 점을 고려하여 아이폰의 동일 파일

내 편집과 아이폰과 갤럭시 파일 간 편집으로 나누어 실험을 구성한다(그림 4).

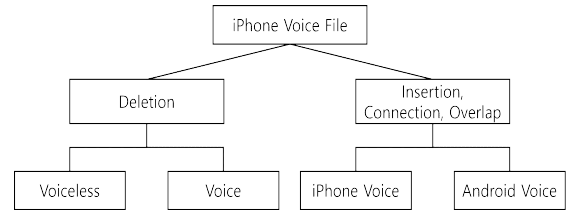


그림 4. 편집 방법 별 적용 대상 파일 유형
 Figure 4. Applicable File Types by Editing Method

IV. 실험 및 분석

1. 삭제 편집 시 스펙트로그램 분석

삭제 편집 기법의 경우 삭제하고자 하는 프레임 내의 발화 유무로 나누어 실험을 했다. 발화영역에서의 삭제 편집을 진행한 실험에서 원본, wav 파일 그리고 최종 인코딩된 m4a 파일 간 스펙트로그램의 특이현상이 발생하지 않았다. 반면 무음 구간의 프레임을 삭제하였을 때, 편집된 wav 파일 내 가청주파수 영역의 스펙트로그램에서 특이점이 존재하는 경우가 발생했다. 삭제된 지점의 가청주파수 대역에 원본에 없던 신호가 생성됐다(그림5). 하지만 인코딩 작업을 수행한 이후, 해당 특이점은 사라진다. 즉, 삭제 편집을 수행한 최종 m4a 파일로 저장될 경우 위변조 여부를 식별하기 어렵다.

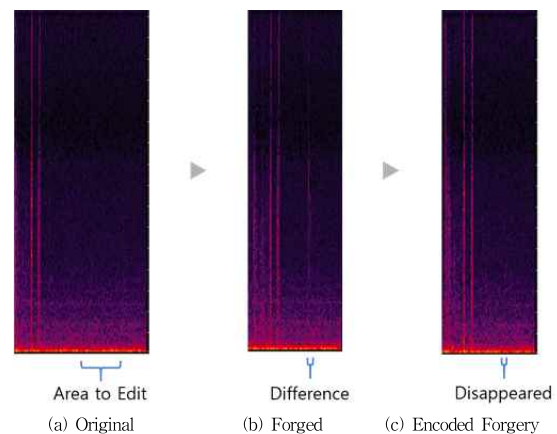


그림 5. 무음 구간 삭제 시 스펙트로그램 변화 과정
 Figure 5. Spectrogram Change Process When Silent Section Is Deleted

2. 삽입 편집 시 스펙트로그램 분석

삽입 편집 실험은 아이폰 파일 간의 삽입과 아이폰 파일에 갤럭시 파일의 삽입이 실험되었다. 아이폰 파일 간 삽입 편집의 경우 wav 파일에서 삽입 시작 부분의 천이 대역에 붉게 튀는 현상이 나타난다. 하지만 wav 파일을 m4a 파일로 변환한 후 해당 대역을 확인해 보면 튀는 현상들이 줄어들었으며, 편집을 수행하지 않는 영역에서 붉게 튀는 현상도 줄어들게 된다(그림6).

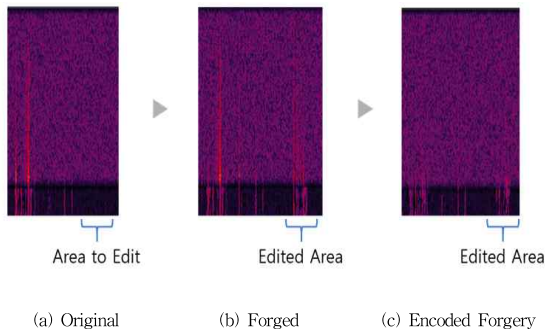


그림 6. 아이폰 파일 내 삽입 시 스펙트로그램 변화 과정
Figure 6. Spectrogram Change Process When Inserting into an iPhone File

반면 아이폰 음성 파일에 갤럭시 음성 파일을 삽입하는 경우, wav 파일 생성 시 그림 7와 같이 천이 대역에서의 주파수 차이에 의한 특이현상이 나타난다. 이는 아이폰과 갤럭시에 적용된 로우패스 필터(Low-pass filter)의 차단 주파수(cut-off frequency)의 차이에 의해 나타난다(아이폰은 16khz, 갤럭시는 20khz). 하지만 인코딩 작업을 거친 이후 해당 특이점은 없어진다. 따라서 삽입 편집 시 최종 인코딩 작업을 수행한 음성 녹음 파일의 스펙트로그램은 원본과 명확히 구분되는 특이 현상을 보이지 않는다.

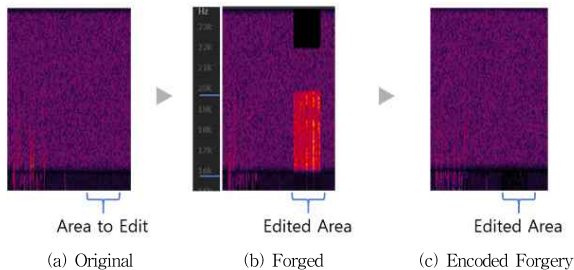


그림 7. 아이폰 파일에 갤럭시 파일 삽입 시 스펙트로그램 변화 과정
Figure 7. Spectrogram Change Process When Inserting a Galaxy File into an iPhone File

3. 연결 편집 시 스펙트로그램 분석

연결 편집 실험은 두 파일이 겹치는 부분이 있도록 실험되었다. 아이폰 파일 내 연결 편집 시 멀티트랙 기능을 이용하여 서로 연결된 프레임이 wav 파일에서 다른 프레임들에 비해 천이 대역에 색이 덧칠해져 있는 현상을 확인할 수 있었다. 하지만 wav 파일을 인코딩하게 되면 기존의 스펙트로그램이 덧칠되는 현상은 사라지고 원본과 유사하게 변환되었다. 하지만 인코딩 이후 천이 대역의 연결 시작 혹은 끝부분의 파형이 튀는 특이점이 몇몇 실험에서 발견되었다(그림 8).

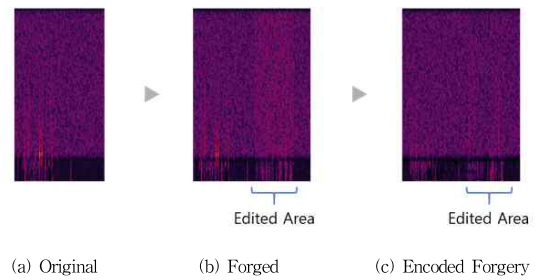


그림 8. 아이폰 파일 간 연결 시 스펙트로그램 변화 과정
Figure 8. Spectrogram Change Process When Linking iPhone Files

아이폰과 갤럭시 파일 간 연결 편집 시 천이 대역에서 삽입 편집 실험과 유사한 특이점이 나타났다. 이 또한 인코딩 이후 원본과 같이 변환되어 특이점이 사라진다(그림 9).

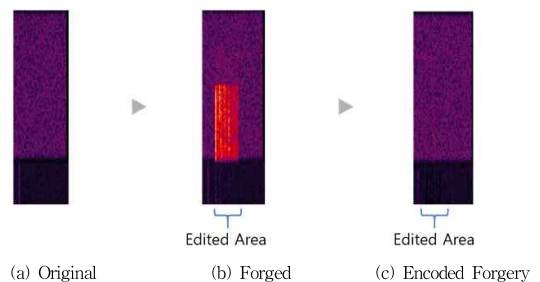


그림 9. 아이폰과 갤럭시 파일 간 연결 시 스펙트로그램 변화 과정
Figure 9. Spectrogram Change Process when Connecting iPhone and Galaxy files

4. 합성 편집 시 스펙트로그램 분석

아이폰 파일 간 합성 편집 시 원본 음성에서 천이 공간 대역의 스펙트로그램을 제외한 0~16kHz(가청주파수) 부근의 대역을 복사하여 혼합 붙여넣기의 과정을 수행한다. 따라서 원본과 편집된 wav 파일 간 천이 대

역에서의 차이는 확인할 수 없다. 하지만 인코딩 이후 그림 10와 같이 m4a 파일에서의 시작 및 끝 지점의 천이 대역에서 붉은 신호가 튀는 현상이 발생한다.

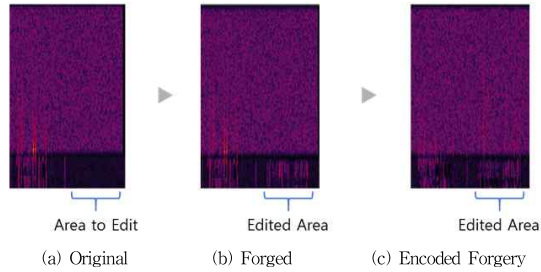


그림 10. 아이폰 파일 내 합성 시 스펙트로그램 변화 과정
 Figure 10. Spectrogram Change Process When Synthesizing within an iPhone File

다음으로 아이폰과 갤럭시 파일 간 합성 편집 시, 그림 11과 같이 wav 파일에서 특이점을 보이지 않고 인코딩 작업 이후 천이 대역에서 튀는 현상을 나타낸다.

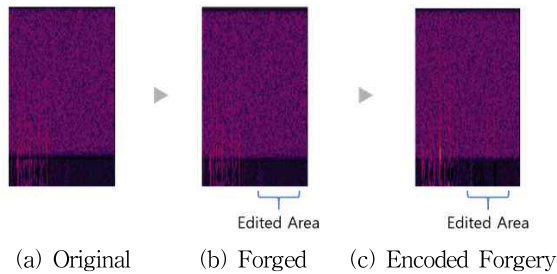


그림 11. 아이폰과 갤럭시 파일 간 합성 시 스펙트로그램 변화 과정
 Figure 11. Spectrogram Change Process When Synthesizing iPhone and Galaxy Files

V. 실험 결과 및 토론

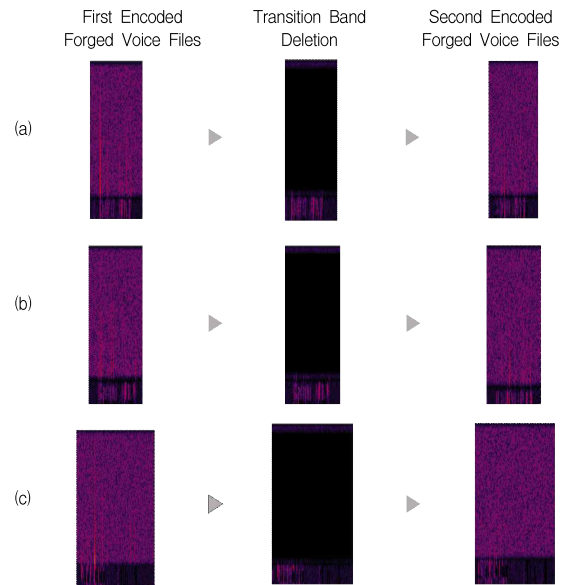
삭제 편집의 경우, 인코딩 작업을 거친 이후 무음 영역과 발화 영역 삭제한 이후 모두 특이점이 사라졌다. 삽입 편집 시 불규칙적으로 천이 대역에 붉은 신호가 생성되는 특이 현상이 발생했지만 인코딩 작업 이후 해당 천이 대역에서의 특이점이 없어짐을 확인했다. 연결 편집의 경우 아이폰과 갤럭시 간 편집은 주파수 대역의 차이로 인해 발생한 천이 대역의 특이현상이 인코딩 작업 이후 사라졌다. 반면 아이폰 파일 내 연결 편집과 합성 실험에서 최종적으로 특이 현상이 발생했다. 즉, 인코딩 과정을 가진 이후에도 천이 대역에서 편집 영역

에서 붉은 신호가 강하게 튀는 현상을 보여 원본 파일과 식별 가능한 차이를 확인할 수 있다. 표1은 전체적인 실험 결과를 나타낸다.

표 1. 실험 결과
 Table 1. Experiment Result

편집 기법	실험 대상	특이 현상(m4a)	
삭제	아이폰	무음	X
		발화	
삽입	아이폰 - 아이폰	X	
	아이폰 - 갤럭시	X	
연결	아이폰 - 아이폰	O	
	아이폰 - 갤럭시	X	
합성	아이폰 - 아이폰	O	
	아이폰 - 갤럭시	O	

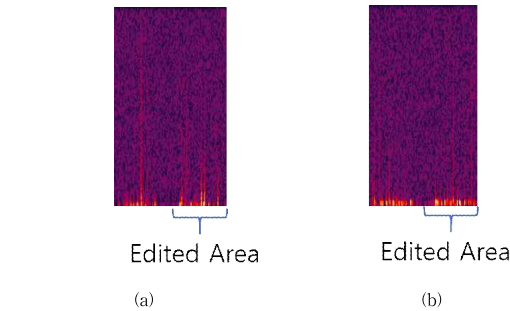
음성 파일 편집 수행 후 최종적으로 인코딩된 m4a 파일에 특이점이 발생한 경우에 한 해 인코딩된 m4a 파일의 천이 대역 전체를 삭제하고 다시 인코딩했다. 즉, 아이폰 파일 내 연결 편집, 아이폰 파일 내 합성 편집 그리고 아이폰과 갤럭시 파일 간 합성 편집 이렇게 세 경우의 천이 대역 전체를 삭제 후 인코딩하여 2차 편집 파일을 생성하고 분석을 수행했다. 2차 편집된 m4a 파일은 기존의 특이현상이 모두 나타나지 않고 원본 파일과 유사하게 변환됨을 확인할 수 있다(그림 12).



(a) Spectrograms When Linking iPhone Files
 (b) Spectrograms When Synthesizing within an iPhone File
 (c) Spectrograms When Synthesizing iPhone and Galaxy Files

그림 12. 천이 공간 삭제 후 재 인코딩
 Figure 12. Re-encoding after Deleting the Transition Band

그림 13과 같이 천이 대역 전체를 삭제 후 재인코딩을 수행하면 편집 영역에서 튀는 현상들이 감소한 것을 나타내며, 편집을 수행하지 않는 부분에서의 붉게 튀는 현상도 감소되었다. 하지만 이는 인간이 편집에 의해 발생인지 발화에 의한 자연스러운 생성인지를 명확하게 판단하는 것은 거의 불가능하다. 즉, 스펙트로그램 분석만으로 음성파일 위변조를 검출하는 것은 한계를 지닌다.



(a) Original Voice File
 (b) Second Encoded Forged Voice File after deletion transition bands
 그림 13. 천이 공간 삭제 후 재 인코딩 했을 때 원본과 비교
 Figure 13. Comparison with the Original when Re-encoding after Deleting the Transition Band

VI. 결 론

IT 기술의 급속한 발전과 확산에 따라 법률 분야에서 디지털 증거 채택 여부가 큰 관심사로 떠오르고 있다[11]. 이에 본 연구는 음성 녹음 파일의 편집 후 스펙트로그램의 천이 대역의 편집 및 인코딩 작업을 통해 위변조가 이론상 가능하다는 것을 실험을 통해 증명하는 것을 목표로 했다.

음성 파일을 편집할 때 최대한 실제 위변조 작업이 요구되는 상황을 가정하여 실험을 진행하였다. 이를 위해 삭제 편집 시 발화 지점 유무로 나누어 편집을 진행하고, 삽입, 연결 및 합성 편집 시 아이폰(iOS) 기종과 갤럭시(Android) 기종을 함께 사용하여 다양한 위변조 예시를 적용했다.

본 연구에서 아이폰 파일 내 연결 편집, 아이폰 파일 내 합성 편집 그리고 아이폰과 갤럭시 파일 간 합성 편집에 해당하는 이 세 가지 실험을 제외하고 원본의 m4a 파일과 구분하기 어려운 편집된 m4a 파일을 생성하는 것이 가능했다. 더욱이 특이 현상이 발생한 세 종류의 실험도 이후 진행한 스펙트로그램의 천이 대역 삭

제 후 인코딩 작업을 통해 스펙트로그램 분석의 한계점을 나타냈다. 즉, 스펙트로그램 상 특이현상을 확인할 수 없는 음성 파일도 위변조되지 않은 파일로 단정되어 판단할 수 없다.

본 연구는 음성 파일 편집 도구를 통한 위변조의 용이성과 검출의 어려움을 실험을 통해 밝힘으로써 음성 녹음 파일의 증거 채택에 있어 엄격한 절차를 주장하는 바이다. 하지만 본 연구는 아이폰 XS MAX(iOS 16.02) 기종만을 기반으로 실험을 진행했다는 점에서 한계가 존재한다. 이에 향후 다양한 아이폰 기종 및 갤럭시 기종 음성 파일을 기준으로 실험을 이어갈 예정이다.

References

- [1] K.S. Kim, "A Study on the Proof of the Authenticity of Digital Evidence." Graduate School of Law at Yonsei University, Seoul, 2015.
- [2] J.S. Lee, J.W. Park, "Research on Improving Standards of Admissibility of Digital Voice-Recorded File under Criminal Procedure Law," *Yonsei Law Review*, Vol. 32, No. 3, pp. 31-59, September 2022. DOI: 10.21717/ylr.32.3.2
- [3] H. Heo, B. So, I. Yang, H. Yu, "A Speech Waveform Forgery Detection Algorithm Based on Frequency Distribution Analysis," *Phonetics and Speech Sciences*, Vol. 7, No. 4, pp. 35-40, December 2015. DOI: 10.13064/KSSS.2015.7.4.035
- [4] S.Y. Ahn, S.H. Ryu, Kim, K.W. Kim, K.H. Hong, "A comparative analysis of metadata structures and attributes of Samsung smartphone voice recording files for forensic use*." *Phonetics and Speech Sciences*. The Korean Society of Speech Sciences, Vol. 14, No. 3, pp. 103-112, September 2022. DOI:10.13064/ksss.2022.14.3.103
- [5] N.I. Park, K. Shim, O. Jeon, "A Study on Authentication Analysis Procedure of Digital Audio Files," *Journal of Digital Forensics*, Vol. 13, No. 4, pp. 257-269, December 2019. DOI: 10.22798/KDFS.2019.13.4.257
- [6] Y.J. Song and G.B. Kim, "A Study on the Detection of Falsification of Voice Recording Files in an Application." *Digital Forensics research*, Vol. 16, No. 3, pp. 65-76, September 2022. DOI: 10.22798/KDFS.2022.16.3.65
- [7] I.H. Yang, K.W. Kim, M.J. Kim, R.S. Baek, H.S. Heo, H.J. Yu, "An Automatic Method of Detecting

- Audio Signal Tampering in Forensic Phonetics.”
Phonetics and Speech Sciences. The Korean Society
of Speech Sciences, June 2014. DOI:
10.13064/KSSS.2014.6.2.021
- [8] J.W. Park, W.J Kwak, S.H. Lee, “A Study on
Forgery Techniques of Smartphone Voice
Recording File Structure and Metadata.” The
Journal of the Convergence on Culture Technology,
Vol. 8, No. 6, pp. 807-812, November 2022. DOI:
10.17703/JCCT.2022.8.6.807
- [9] Criminal Procedure Act of the Republic of Korea
Article 307 Paragraph 2, <https://www.law.go.kr/>
- [10] Bracewell, R. N. “The Fourier transform and its
applications”, Vol. 31999, 1986, New York:
McGraw-Hill., pp. 267-272.
- [11] G.R. Shin, “A Study on Selectively Collecting
Digital Evidence in Search and Seizure.” Journal of
Legislative Studies, Vol. 17, No. 2, pp. 221-245,
August 2020. DOI: 10.31536/ylr.2020.17.2.008

※ 이 성과는 정부(과학기술정보통신부)의 재 원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2021R1F1A1063884)
