



2023 November Vol. 9

11



# 융합연구리뷰

Convergence Research Review

현장형 마약 검출 기술 및 연구 동향

정호상(한국재료연구원 나노바이오융합연구실 선임연구원)

합성 및 변조 음성 탐지 기술 동향

정수환(숭실대학교 전자정보공학부 교수)



미래융합전략센터  
Future Convergence Strategy Center



# CONTENTS

---

- 01 편집자 주
- 03 현장형 마약 검출 기술 및 연구 동향
- 37 합성 및 변조 음성 탐지 기술 동향
- 69 국가R&D 현황 분석

**융합연구리뷰** | Convergence Research Review

2023 November | Vol. 9 No. 11

**발행일** 2023년 11월 13일

**발행인** 임혜원

**발행처** 한국과학기술연구원 미래융합전략센터  
02792 서울특별시 성북구 화랑로 14길 5  
Tel. 02-958-4973 | <https://kist.re.kr/fcsc>

**펴낸곳** 공간기획 Tel. 044-863-0978

# 편집자 주

## ●● 현장형 마약 검출 기술 및 연구 동향

지난 4월 강남 학원가에서 무료 시음회를 가장하여 학생들에게 마약을 탄 음료를 나눠준 사건은 사회 전체를 큰 충격에 빠뜨렸다. 누구든지 마약범죄의 피해자가 될 수 있다는 두려움이 우리 사회를 짓누르고 있다. 대한민국은 마약 안전지대라는 말이 무색하게 마약사범이 빠르게 증가하고 있다. 2023년 8월 마약류 월간동향에 따르면, 올해 1월부터 8월까지 국내에서 마약류 사범으로 단속된 사람은 총 1만 8,187명으로 이는 작년 전체 단속 인원인 1만 8,395명에 육박하는 수치다. 그러나 그동안의 경찰의 마약 수사는 마약 투약 여부를 현장에서 확인할 수 없어 제약이 있었다. 이에 검거 현장에서 신속하게 마약을 검출할 수 있는 방법 마련이 시급하다.

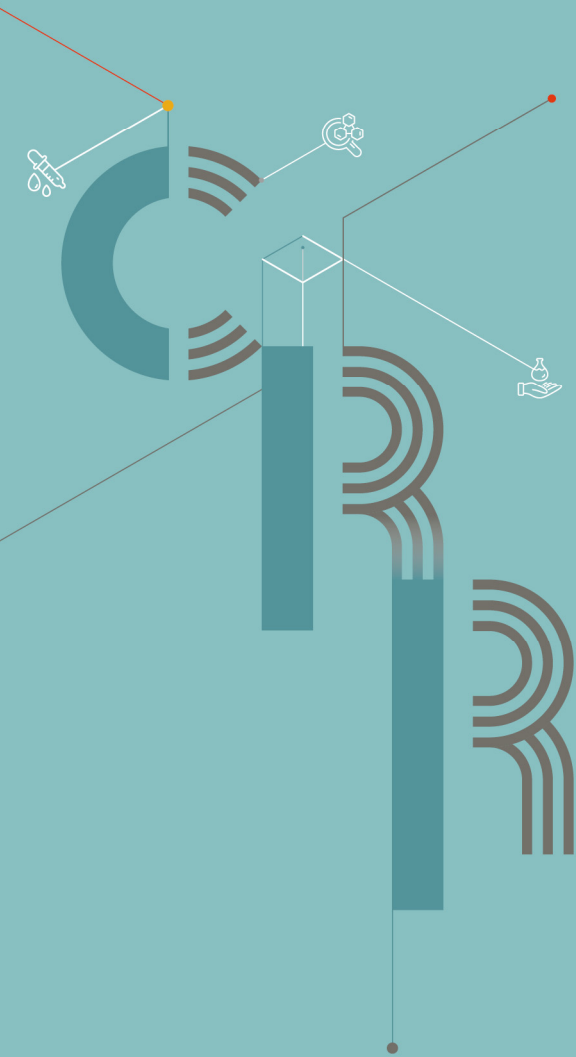
본 호 1부에서는 기존의 마약 검출에 활용되는 이온스캐너, 라만분광기, 촉방유동면역 분석 키트 등에 대해 알아보고, 또한 비색 센서, 전기화학 센서, 광학 센서를 이용한 현장형 마약 검출 기술을 소개한다.

마약과의 전쟁은 미국이 먼저 시작했다. 리처드 닉슨 대통령이 베트남 전쟁 시기에 처음 선포한 이래 마약과의 전쟁은 꾸준히 이어지고 있다. 미국 질병통제예방센터(CDC, Centers for Disease Control and Prevention)에 의하면 2020년 5월부터 2021년 4월까지 1년 동안 마약중독 사망자 수가 처음으로 10만 명을 넘었는데, 이는 총기 사고로 1년에 사망한 사람 수인 약 4만 명 대비 2.5배 많은 수치이다. 우리 정부도 올해 11월 전국 4개 검찰청에 마약범죄 특별수사팀을 설치하고 마약 범죄에 엄정 대응하기로 했다. 첨단 과학기술을 기반으로 하는 현장형 검출 기술 개발을 통해 마약과의 전쟁에서 반드시 승리할 수 있기를 기대해 본다.

## ●● 합성 및 변조 음성 탐지 기술 동향

이제는 수화기 너머의 가족의 목소리도 한번쯤은 의심해봐야 할 날이 멀지 않은 것으로 보인다. 인공지능 기술을 기반으로 한 음성 합성 기술로 범죄 발생이 증가하고 있기 때문이다. 대표적인 사례로 2019년 영국에서는 어느 에너지기업 대표가 AI 음성 생성 기술로 제작된 모회사 대표의 전화를 받고 22만 유로(약 2억 9천만 원)에 달하는 금액을 송금하는 사건이 발생했다. 피해자에 따르면, 전화 목소리가 진짜 CEO와 똑같았고, 독일 사투리까지 정확히 구사해 속을 수밖에 없었다고 한다. 우리나라도 금융감독원에 따르면 2018년부터 2022년까지 5년간 보이스피싱으로 인한 피해 건수는 22만 7,000건, 피해 금액은 1조 6,000억 원에 달한다고 한다. 이렇듯 합성 음성 생성 기술을 악용한 범죄들이 날로 다양해지고 교묘해져 이에 대응하기 위한 기술개발이 필요한 때이다.

딥러닝 기술의 발전으로 음성 합성 기술은 비약적인 성장을 이루었다. 딥러닝 모델을 통해 다양한 사람의 발화 음성에서 나타나는 강세, 높낮이, 음색 등의 특성과 호흡과 공기 중 소리의 공명과 같은 세부적인 부분까지 학습해 예측하고 적은 양의 데이터로도 자연스러운 음성을 구현할 수 있어 무궁한 가치를 가지고 있고 관공서, 방송국 등에서 음성 합성 기술 시장이 빠르게 부상하고 있다. 그러나 이와 동시에 음성 합성 기술이 악용될 수 있기 때문에 합성 및 변조된 음성을 탐지하는 기술도 개발 중이다. 본 호 2부에서는 현존하는 음성 합성 기술에 대해 먼저 알아보고, 이에 대응하기 위한 합성 음성 탐지 기술도 소개한다.



# 융합연구리뷰

Convergence Research Review

# 01

---

## 현장형 마약 검출 기술 및 연구 동향

---

정호상(한국재료연구원 나노바이오융합연구실 선임연구원)

# 01

정호상(한국재료연구원)

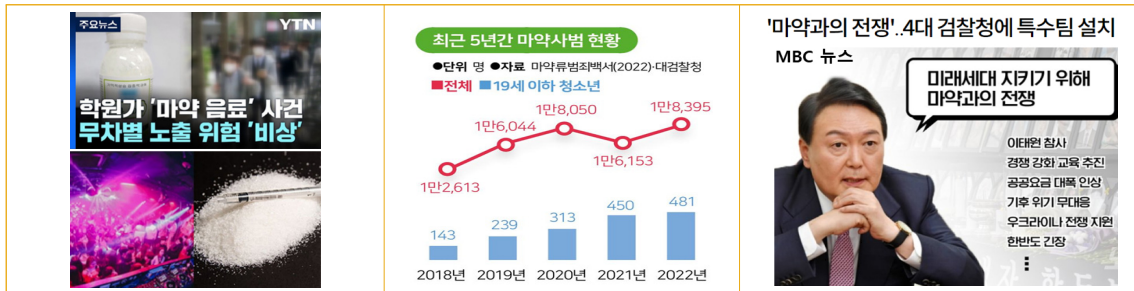
## 현장형 마약 검출 기술 및 연구 동향

### I. 서론

#### 1. 마약 범죄의 사회적 심각성

최근 연예인, 인플루언서, 유명인들의 마약 범죄가 사회적 문제로 심각하게 떠오르고 있다. 최근에는 청소년들이 다니는 학원가에서 마약이 포함된 음료를 무작위 배포되는 사건이 발생하면서 전 국민을 충격과 공포에 떨게 하였다. 이제 누구나 쉽게 텔레그램, SNS를 통해 마약을 구할 수 있는 시대가 되었고, 한 해 마약 압수량은 2022년 기준 804kg이나 된다(대검찰청, 2023). 필로폰 기준 1kg의 마약은 약 3만 3천명이 동시에 투약할 수 있는 양으로 불법 마약류의 국내 유입이 심각한 상황에 이르렀다. 또한, 검찰에 따르면 2022년 마약사범은 1만 8,395명으로 집계되었고, 5년 전 대비 30% 증가하였다(대검찰청, 2023). 드러나지 않은 범죄 비율을 의미하는 ‘암수율’이 29배인 점을 고려하면 국내 전체 마약사범은 52만 명에 달할 것으로 추정된다. 특히, 청소년 마약사범은 2022년 기준 481명으로, 5년 동안 304%나 급증하였다. 마약을 구매해 투여한 경우뿐만 아니라 직접 마약을 판매, 유통한 사례까지 등장하고 있어, 미래 세대를 위해 마약 범죄를 뿌리 뽑을 수 있는 해법이 필요한 시점이 되었다.

그림 1. 국내 마약 관련 범죄 증가 현황 및 정부의 입장



\* 출처: (좌) YTN 뉴스(2023.04.07), (중) 대검찰청(2023), (우) 노동자 연대 기사(2023.02.15), MBC 뉴스(2022.10.14)

현 정부는 2022년 말 전국 4개 검찰청에 ‘마약범죄특별수사팀’을 설치하여 ‘마약과의 전쟁’을 선포하였다. 또한, 2023년 9월, 마약류 대응 범정부 예산을 전년 대비 약 2.5배 수준인 602억 원으로 대폭 확대 편성하였다. 마약 범죄를 근절하기 위해서는, 다양한 마약 범죄 현장에서 과학적으로 대응할 수 있는 현장형 신속·고감도 마약류 검출 기술개발이 필요하다. 마약의 유통, 소지, 가공품 내 마약류 성분 확인뿐만 아니라, 생체시료 내 복용 마약 검출, 이를 기반으로 한 마약사범의 재범률 저하를 위한 관리 기술개발 또한 필요한 시점이다.

## II. 마약 검출 기술 현황

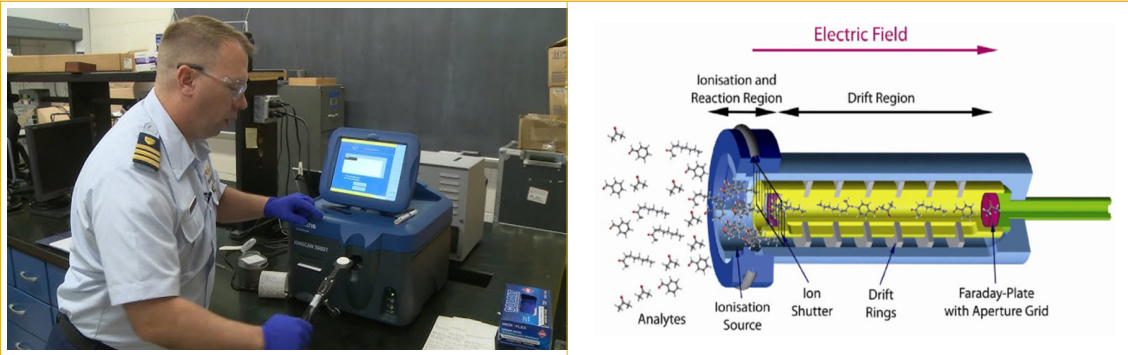
### 1. 마약 검출 기술의 국내외 현황

현재, 소지·유통 마약의 경우 이온스캐너, 라만분광기를 이용하여 현장에서 해당 약물의 신호를 판별하여 압수하고 있다. 복용 약물의 경우 임신진단키트와 같은 원리로 측방유동면역분석 키트(Lateral-Lateral Flow Assay Kit)를 활용하여 소변 내 복용 마약류를 검사하고 있다. 시장에 공급되고 있는 이온스캐너(Ion Scanner or Ion-Mobility Spectrometer)와 라만분광기(Raman Spectrometer)는 대부분이 유럽과 미국 제품으로, 국산 장비의 부재로 인해 국내 시장이 일부 업체에 의해 독점되고 있다. 해외에서 수입되는 제품은 성능이 우수하지만, 고가이고, 안정성 및 유지보수 한계 등의 문제점이 있다. 측방유동면역분석 키트 또한 전량 해외 제품을 수입하여 사용하고 있다. 따라서 국내 원천 기술개발을 통해 기술 수입의존도 감소와 더불어 신종 마약류에 대응할 수 있는 현장형 마약 탐지 기술개발이 절실한 상황이다.

#### 1.1 이온스캐너

현재 관세청과 일부 마약 수사대에서 활용하고 있는 외산 이온스캐너는 대당 5천만 원 수준이다. 공항에서 소지품 검사를 할 때 <그림 2>와 같은 도구를 가방이나 옷에 문질러서 검사하는 경우가 있는데, 이는 이온스캐너를 통해 마약류나 폭발물의 흔적을 감지하기 위한 것이다. 이온스캐너는 쉽게 질량분석기의 단순화 형태로 생각할 수 있다. 마약류가 감지 부위에 존재하면 방사선원에 의해 장비 진입 시 이온화되고, 이후 전기장에 의해 유동 영역(Drift Region)을 지나가면서 이온화된 마약류의 전하, 형태, 질량 차이에 따라 검출 부위에서 작은 이온부터 검출된다. 이후 검출된 이온과 기기가 보유하고 있는 분자 라이브러리를 비교하여 시료를 탐사하는 원리로 마약류를 검사한다. 전통적인 마약류(코카인 및 헤로인)의 검출 감도는 1~50ng 수준이나 분자량이 작은 마약류에 대한 고감도 검출에는 한계가 있다. 또한, 고체 시료 이온화원으로 방사성동위원소 사용에 따른 안전성 이슈, 이온화원의 장시간 warm up 시간(15분), 고가의 유지비용(~300만 원/yr) 등으로 인해 전국적인 보급 확대에 어려움이 있다.

그림 2. 이온스캐너를 이용한 마약 검출 사진 및 검출 원리

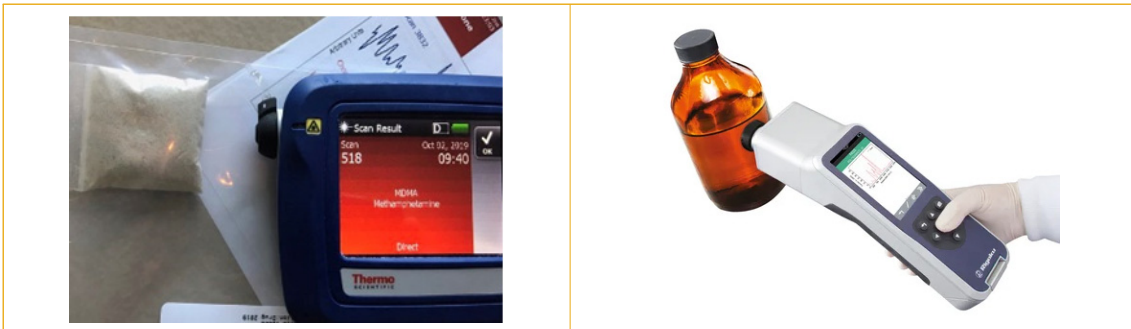


\* 출처: (좌) Theday 기사(2019), (우) Hauschild et al.(2012)

## 1.2 라만분광기

최근 초소형 레이저 및 광학 기술의 급속한 발전으로 전 세계적으로 휴대용 라만분광기가 개발되어 시판 중이다. 라만분광기술은 물질 고유의 분자 진동에 의한 화학정보(Molecular Fingerprint)를 제공하는 분석기술에 해당한다. 따라서, 추가적인 분석 없이 획득한 스펙트럼을 바탕으로 해당 물질이 무엇인지 직관적으로 알 수 있다는 장점이 있다. 또한, 대부분의 화학물질에 대한 라만 스펙트럼 라이브러리가 존재하기 때문에 데이터베이스 활용에도 장점이 있다. 라만분광기는 레이저 조사를 통한 비파괴 검사가 가능하여 별도의 유지비용이 필요 없고, 1분 이내의 warm up이 가능하며, 안전성 이슈가 없어 휴대용 마약 탐지 장비로써 파급력이 있다. 그러나 라만분광기 단독으로는 파우더 및 원액 마약류만 분별할 정도로 검출 민감도가 낮아 극미량의 마약 흔적을 분석하기에는 무리가 있다. 또한, 현재 국내 수사기관에서 사용하는 모든 라만분광기는 전량 외산 제품을 사용하고 있다.

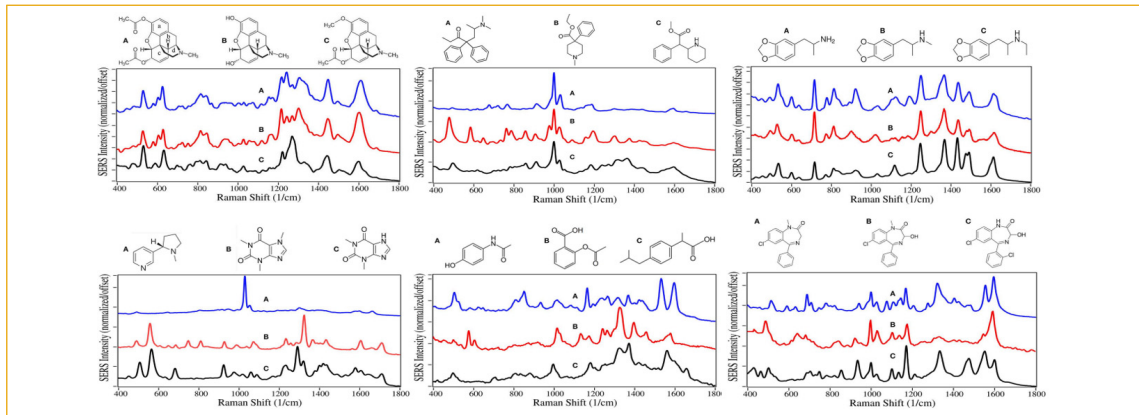
그림 3. 휴대용 라만분광기를 이용한 마약류 검사



\* 출처: (좌) Thermo Fisher Scientific社 홈페이지, (우) Rigaku社 홈페이지



그림 4. 마약별 분자지문 라이브러리 예시

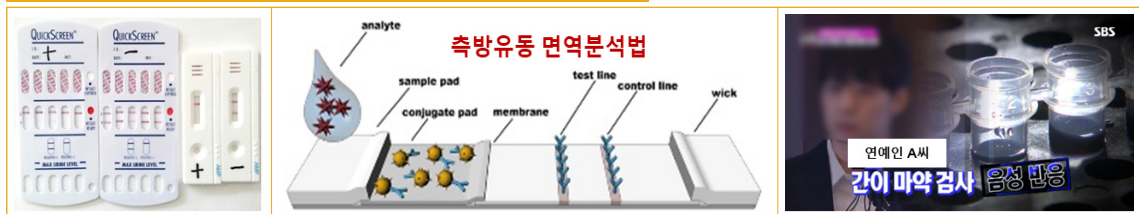


\* 출처: Farquharson et al.(2019)

### 1.3 측방유동면역분석법

측방유동면역분석법은 임신진단키트 혹은 코로나19 검사키트와 같은 원리로, 타겟 마약이 있는 곳에서 나타나는 붉은 선으로 마약 복용에 대한 음성 및 양성 여부를 판단한다. 타겟 마약과 특이적으로 결합할 수 있는 항체를 이용하여 항체가 마약을 포집한 경우, 금 나노입자나 색이 있는 비드(bead)가 검사 선에 모여 색이 나타나게 된다. 해당 방법은 경우에 따라 경쟁적 반응을 활용하기도 하는데, 마약이 있는 경우 검사 선에서 오히려 색이 나타나지 않게 하면서 마약이 검체 내 존재하였는지를 추정하기도 한다. 해당 방법은 간단하고 신속하게 마약 복용 여부를 확인할 수 있지만, 색 변화를 눈으로 확인하여 판별하기 때문에 주관적 판단이 개입될 수 있다. 또한, 저농도 마약류의 경우 색 변화량이 미미하여 고감도 검출이 불가능하다는 단점이 있다. 가장 큰 한계는, 단일 항체는 하나의 마약류만 검출할 수 있기 때문에 하나의 키트는 하나의 마약만 검사할 수 있다는 점이다. 예를 들어, 대마 성분을 검사하는 키트를 이용하여 마약검사를 하면, 코카인을 한 사람에게 대해서는 음성의 결과를 보여준다. 따라서, 여러 종의 마약류를 동시에 검출하기 위해 개별 키트를 병렬로 연결하여 스크리닝 하는 방법이 시도되고 있지만 키트 가격이 개당 수십만 원(i.e. 20개 동시 검사)에 이르러 현장에서 활용하기에 부담이 있다. 또한, 신종 마약류의 경우 분자구조가 달라 현장에 적용하는데 여전히 한계를 나타내고 있다.

그림 5. 측방유동면역분석 마약키트 및 작동원리, 검출 정확도 이슈



\* 출처: (좌) Countrywied Testing社 홈페이지, (중) Miočević et al.(2017), (우) SBS 본격 연예 한밤(2019.04.23)

한편, 측방유동면역분석법은 소변뿐만 아니라 타액에도 적용되어 마약 검출 키트 형태로 활용하는 국가들이 있다. 경구투여한 마약은 구강세포에 의해서 한동안 재배출되어 타액에 섞여 있기 때문에, 체외 배출이 모두 완료되기 전에 검출할 수 있다. 미국, 유럽, 호주 등에서는 roadside-drug-driving test (거리에서의 마약 운전 검사)를 통해 행인 혹은 운전자의 침을 키트에 흐르게 하여 대상 마약류가 발견되면 경찰이 체포할 수 있도록 시행하고 있다.

그림 6. 침을 이용한 마약 검출 키트 및 국외 운전자 마약검사



\* 출처: (좌) TransMed社 홈페이지, (중) Daily Mail 기사(2018.04.27), (우) Daily Mail 기사(2019.10.12)

### 1.4 땀 속 마약류 검출 기술

미국에서는 비침습(혈액을 뽑지 않는) 방법으로 용의자 혹은 마약사범의 마약 복용 여부를 검사하기 위해 인체 피부에 땀을 수집할 수 있는 패치를 붙이는 기술을 활용하고 있다. 복용한 마약류는 혈액과 함께 체내를 순환하다가, 땀을 통해 일부 배출된다고 알려져 있다. 인체의 다양한 생체시료(땀, 침, 눈물, 소변, 모발)의 배출 기간은 모두 다르게 나타나는데, 모발이 가장 오랜 시간 마약 흔적을 남기고 그 다음은 땀인 것으로 알려져 있다(이는 약물의 반감기와 대사 루트에 따라 일부 차이가 있다). 미국 Pharmcheck社는 땀을 흡수할 수 있는 패드 형태의 패치를 상용화하여 미국 내에 보급하고 있다. 마약을 복용한 경우 땀과 함께 마약이 배출되어 패치에 축적되고, 일정 시간 이후 패치를 제거하여 실험실에 배송하면, 크로마토그래피-질량분석기를 통해 마약 검출 여부를 리포트 하는 서비스를 하고 있다. 소변을 검사하는 경우 검사자가 용의자가 용변을 보는 것을 봐야 하므로 윤리적 문제가 발생하는데, 땀을 이용하는 경우 이를 피할 수 있어서 미국에서는 해당 기술을 받아들이고 50개 미연방 대법원에서 증거로 인용한 판례도 존재한다.

그림 7. 미국 Pharmcheck社의 땀 속 마약 검출 패치 및 검사가능 마약류, 분석 방법



\* 출처: (좌, 중) Pharmcheck社 홈페이지, (우) Hemochrom社 홈페이지

## 1.5. 표면증강라만산란 센서 및 휴대용 라만분광기 융합시스템

미국 Metrohm Raman社は 2018년 휴대용 라만분광기(Mira DS)를 이용한 마약 검출기기를 상용화하였다. 시료의 라만스펙트럼을 효과적으로 획득하는 광학 설계기술과 내장된 약물 라이브러리와 매칭하는 통계분석기법을 탑재하여 마약류에 대한 검출이 가능하도록 휴대용 라만분광기를 제작하였다. 특히, 마약류 고감도 검출을 위해 귀금속 나노입자(Ag, Au)를 종이 위에 잉크젯 프린팅한 표면증강라만산란(SERS, Surface Enhanced Raman Scattering) 스트립을 활용하였다. SERS 기술은 분자의 고유 구조 신호를 100억 배 이상 증폭하기 때문에 추가적인 분석 없이 마약류를 판별할 수 있다는 장점이 있다. 본 기술은 대상 시료를 스트립에 용적 후 건조하여 라만분광기를 찍는 방식으로 활용이 제안되었다. 이후, Metrohm Raman社は 라만분광기에 액체, 기체, 고체 등을 손쉽게 검출할 수 있는 어댑터를 추가로 개발하여 2020년부터 보급하고 있고, 마약류뿐만 아니라 폭발물, 유해물질 등으로 응용 범위를 확장하고 있다.

그림 8. 미국 Metrohm Raman社の SERS 기반 마약 검출 키트 및 SERS 원리



\* 출처: (좌) Metrohm社 브로슈어, (우) Rajapandiyam 홈페이지

최근, 스웨덴의 SERSTECH社에서 미국 사례와 비슷하게 휴대용 라만분광기에 어댑터를 연결하여 헤로인, 펜타닐, 메스암페타민, 케타민, 코카인, 암페타민을 검출할 수 있는 기술을 상용화하여 보급하고 있다. 고감도 검출을 위해 SERS 플레이트(plate)라는 센서를 패키지로 제공하는데, 마약류를 녹일 수 있는 용매인 메탄올을 함께 제공하여 사용자가 의심 마약류를 용매에 녹이고, SERS plate에 용적하여 건조 후 사용하도록 한다. 해당 제품은 국내 경찰청에도 일부 보급되었지만, 휴대용 라만분광기 사용에 대한 이해 부족과 다양한 마약 현장을 반영하기에 기술적 한계가 있어 활용도가 거의 없는 수준이다.

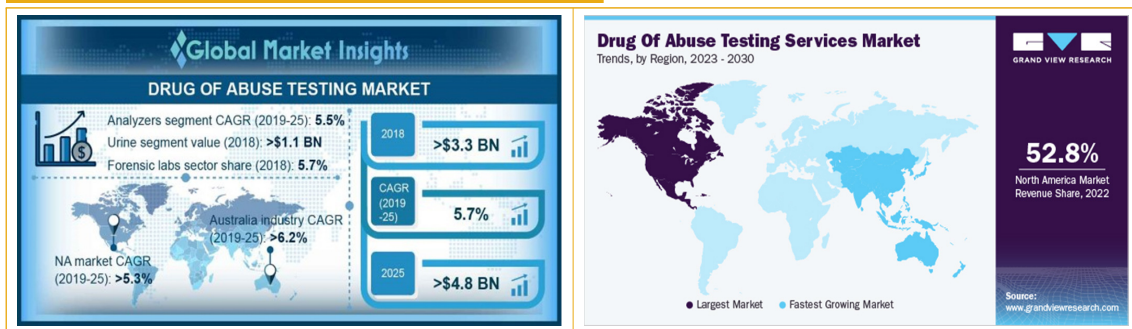
## 2. 마약 검출 시장 현황

### 2.1 세계 약물 검출 시장 현황

세계 마약류 검출 테스트 시장은 2018년 기준 33억 달러 이상으로 집계되었고, 2019년부터 2025년까지 연평균 성장률 5.5%를 기록하여 2025년까지 48억 달러 이상으로 증가할 것으로 예상된다. 지역적으로는

북미, 유럽, 아시아, 라틴 아메리카, 중동 및 아프리카로 분류할 수 있으며, 북미가 약 70%로 가장 큰 시장을 차지한다. 아시아 지역은 약 10% 정도에 해당하지만, 최근 빠른 성장세를 나타내고 있다. 세계 시장을 선도하는 주요 업체로는 Alere, Inc., Drugerwerk AG & Co., KGAA, Abbott Laboratories, Hoffmann-La Roche AG, Biorad Laboratories, Inc., Siemens AG 등이 있으며 세계 시장의 80% 이상을 일부 업체들이 독점하고 있다. 국내 기업이 마약 검출 키트를 개발한 사례는 있지만, 현재 수사 현장에서는 전량 외산 제품을 수입하여 검사를 시행하고 있다.

그림 9. 세계 약물 검사 시장 규모 및 신규 시장 진입 국가들

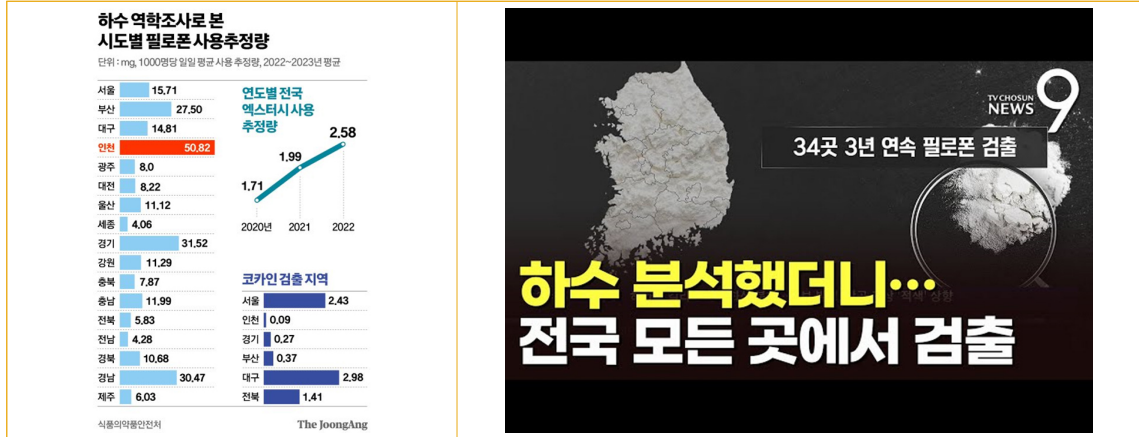


\* 출처: Grand View Research(2023)

## 2.2 국내 약물 검출 시장 현황

국내에서는 유통 마약 적발을 위해 공항, 항만 등에서 외산 이온스캐너를 활용하고, 경찰청, 검찰청 등에서 마약 복용 용의자 검사, 마약사범 투약 여부 조사 등을 목적으로 측방유동면역분석 기반 간이검사 키트를 활용한다. 모든 장비 및 키트는 전량 해외 제품을 수입하여 사용한다. 외산 제품 의존도가 높아 현장 경찰의 수사기법에 상당수 의존하는 경우가 많으며, 국내 기업에서 개발한 제품의 경우 국내 보급은 이루어지지 않고, 일부 동남아시아 지역에 보급된 사례가 있다. 2023년 군(軍) 마약검사 의무화 시행을 검토하겠다는 정부 발표에 의해 국내 시장 규모를 추산한 사례가 있는데, 100억 원에도 못 미치는 것으로 밝혀졌다. 이와 같이 시장 규모가 크지 않은 데에는 군인과 같은 공무 수행 인원뿐만 아니라 마약 용의자들에 대해 법적으로 약물검출 의무화가 이루어지지 않은 데에 영향이 있다. 의무 검사 대상자 수가 적으면 국내 기업이 시장에 뛰어들기 힘든 실정이다. 하지만, 마약류 오남용으로 인한 사회적 범죄가 증가함에 따라 범죄 예방 차원의 마약 복용 규제 및 통제가 강화되고 있으며, 이에 따라 마약 복용자를 판별하기 위한 수단이 필요하고, 관련 시장은 꾸준히 증가할 것으로 예상된다. 대한민국은 오래전 마약 청정국 지위를 상실하였다. 2022년 식품의약품안전처가 전국 27개 하수처리장에서 잔류 마약류 종류와 양을 분석한 결과, 모든 곳에서 필로폰이 배출되었다고 발표하였다. 대한민국이 더 이상 해외 기술에 의존하지 않고, 국내 기술로 현장에서 마약류를 검사할 수 있는 기술력 확보에 총력을 기울여야 한다고 판단된다.

그림 10. 국내 하수 역학조사 결과 및 마약 검출량

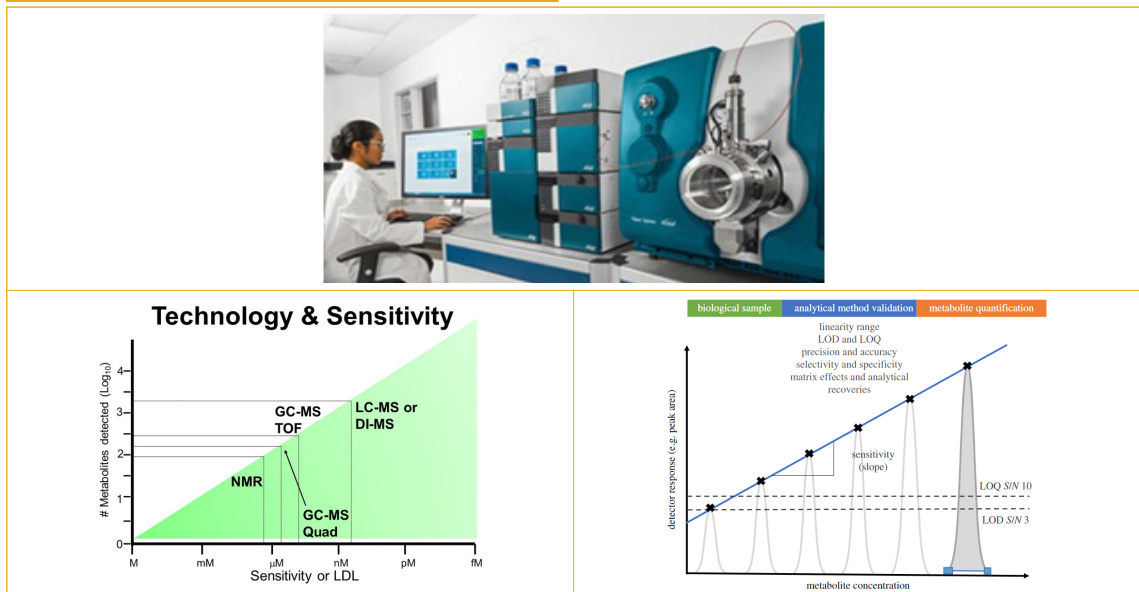


\* 출처: (좌) 중앙일보 기사(2023.06.09), (우) TV 조선(2023)

### 3. 현장형 마약 스크리닝 기술의 필요성

앞에서 국내외 다양한 마약 검출 기술 현황과 종류를 소개하였다. 소지 및 복용 마약의 정밀 분석은 크로마토그래피와 질량분석기가 결합된 액체크로마토그래피 텐덤 질량분석기(LC-MS/MS, Liquid Chromatography-tandem Mass Spectrometry), 기체크로마토 그래피 질량분석기(GC-MS, Gas Chromatography-tandem Mass Spectrometry)가 현존하는 가장 우수한 기술이다.

그림 11. LC-MS/MS 장비 및 기술별 검출 민감도 비교



\* 출처: (상단) SCIEX社 홈페이지, (하단 좌) Goldansaz et al.(2017), (하단 우) Jorge et al.(2016)

그러나 분석기기가 고가이고, 숙련된 작업자가 필요하며, 표준물질 확보와 복잡한 전처리 과정이 필요하다는 점에서 현장 활용성이 떨어진다. 국립과학수사연구원에서 검사하는 방법이 상기 기술들에 해당하는데, 현재 국립과학수사연구원 1일 평균 마약검사 의뢰 수가 300건에 이를 정도로 업무 부담이 크다. 마약검사 의뢰가 많아질수록 검사 결과가 나오는데 더 많은 시간이 소요되어 수일에서 일주일까지 소요되는 경우도 있다. 이는 코로나-19 초창기에 모두 PCR 검사를 통해 확진을 받고 격리 및 행정조치가 이루어졌기 때문에, 바이러스 확산을 막기에 어려웠던 것과 유사한 상황이다. PCR 검사를 위해 검체 채취, 검사기관에 전달, 검사 결과가 나오기까지 만나질 정도 소요되었다. 코로나-19 감염자가 폭발적으로 나타난 시기에는 국가에서도 이를 감당하기 어려울 정도로 검사 시간과 비용이 소요되었고, 결국 자가진단 키트를 긴급 승인하고 보급함으로써 현장 스크리닝 기술의 효과를 체감할 수 있게 되었다. 이와 마찬가지로 마약 검출의 경우도, 정량분석까지 필요하지는 않지만 다양한 마약류를 동시에 검사할 수 있고, 또한 마약 현장에서 마약 소지 및 복용 여부를 즉각 판별할 수 있는 기술이 개발된다면 마약 관련 범죄를 효과적으로 감소시킬 수 있을 것으로 기대된다. 현장형 기술 중 측방유동면역분석법의 경우 앞서 논의한 대로 정확도가 떨어지고, 동시에 여러 마약을 검사하기에 부족하다는 한계가 있다. 특히, 현장 스크리닝 결과와 최종 국립과학수사연구원 검사 결과가 달라 신속한 경찰 수사에 차질을 주는 경우가 자주 보도된다. 복용 마약뿐만 아니라 소지 마약, 음식 및 음료, 가공품 형태로 유통되는 마약들이 다양화되고 있으므로 복용 전·후 마약류를 판별할 수 있는 기술개발이 필요하다.

그림 12. 다양한 형태의 적발 마약류 및 마약 범죄 현장



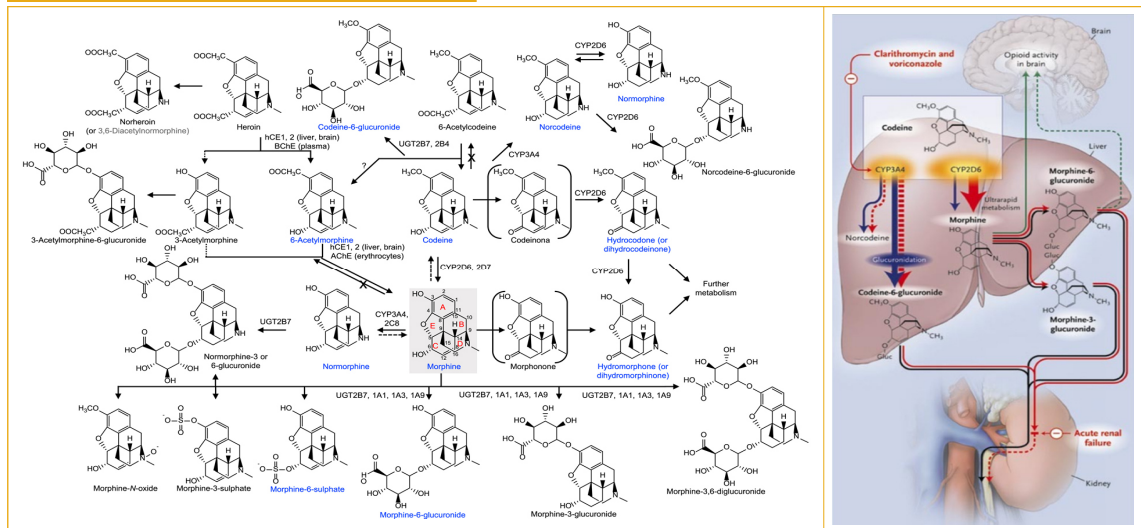
\* 출처: 한국일보 기사(2023.02.07), 한국경제 기사(2022.08.11), 조선일보 기사(2022.07.19)

### III. 생체 내 마약류 대사 및 시료별 특성

#### 1. 인체 내 마약류 대사 과정

마약이 체내에 투여되면 혈관을 통해 뇌를 포함한 모든 장기에 전달되게 된다. 특히, 중추신경계를 통해 신경전달물질의 교란을 일으켜 다양한 정신적 효과를 가져오게 된다. 마약은 아래 <그림 13>처럼 체내에서 다양한 형태로 물질대사 되어 형태가 변화하고, 이후 배설물을 통해 인체에서 빠져나가게 된다. 마약마다 반감기(혈액 내 초기 농도 또는 양이 절반이 되는 시기)가 다르지만, 소변, 대변, 침, 땀 등으로 배출되고 모발에는 축적되게 된다.

그림 13. 마약의 대사, 체내 작용 및 배출 과정



\* 출처: (좌) Dinis-Oliveira.(2019), (우) Gasche et al.(2004)

물붕(GHB, Gamma-HydroxyButyric acid)과 같이 체외 배출이 빠른 경우에는 며칠만 지나도 투약 여부를 판단하기 힘들다. 따라서, 범죄 목적으로 GHB에 노출되면 피해 사실을 인지한 즉시 소변이나 증거품을 가지고 경찰서에 방문해야 한다. 한편, 메스암페타민과 같은 약물은 반감기가 6~15시간 정도 되는데, 상대적으로 체내 오래 남아있으므로 검출이 용이한 편이다. 대마초 성분인 테트라하이드로카나비놀(THC, TetraHydro Cannabinol)의 경우 일반적으로 수일 이상(많게는 2주) 체내에 잔류하게 된다. 이처럼 상대적으로 반감기가 긴 약물들의 경우 측방유동면역분석법으로 검출할 수 있도록 키트가 다른 약물들에 비해 잘 개발된 편이다. 한편, 약물의 반감기에 영향을 미치는 요인들은 나이, 혈액순환, 식단, 성별, 약물 사용기록, 신장 및 간 기능, 체중, 질병, 인종 등 다양하다. 복용 마약의 경우 물질대사와 함께 분해 및 체외 배출되기 때문에 검사 시점에 따라 마약 복용 여부를 결정하는 것에 변수가 작용할 수밖에 없다.

## 2. 생체시료별 마약류 방출 특징

마약 복용 여부를 검사하기 위해서는 생체시료에 마약류 성분이 포함되어 있는지를 검사해야 한다. 따라서, 활용할 수 있는 생체시료를 선정해야 하는데 보통 혈액, 소변, 모발이 가장 많이 활용되고 있다. 앞서 언급한 것처럼 모발의 경우 마약류가 축적되기 때문에 국립과학수사연구원에서 검사하기 좋은 시료에 해당한다. 마약의 반감기마다 차이가 있지만, 보통 소변은 2~3일, 혈액은 12~48시간, 침은 12~48시간, 땀은 수 주, 모발은 수개월~수년까지 마약 흔적 검출이 가능하다고 알려져 있다.

현존하는 기술로는 앞서 언급한 LC-MS/MS, GC-MS 등을 이용하여 모든 생체시료에 대해 전처리 후 마약을 검출할 수 있지만, 현장형 마약 검출 기술의 필요에 따라 혈액 대신 소변, 침, 땀 등 비침습 방법으로 마약류를 검출하는 기술들이 개발되고 있다. 아래 <그림 14>와 같이 소변은 수일 정도 마약 흔적 검사가 가능하고, 또 혈액과 유사한 정도로 대사산물을 배출하기 때문에 마약의 농도가 다른 생체시료에 비해 높아 검사가 용이하다. 하지만, 시료 오염의 위험, 윤리적 문제 때문에 대체 기술이 끊임없이 요구되고 있다. 특히, 마약사범의 관리의무가 있는 법무부에서는, 마약사범들이 재범을 일으키지 않는지 주기적으로 소변검사를 시행해야 하는 의무가 있다. 이 경우 소변검사를 여러 번 검사자 앞에서 실시해야 해서 수치심을 유발하는 등 관리대상자뿐만 아니라 검사자도 어려움을 느끼고 있다. 침의 경우 경구투여한 마약류가 구강세포에서 재분비되기도 하고, 다른 루트로 복용한 마약이 침에 섞여서 검출되는 경우가 있다. 하지만 침은 음식물 섭취를 포함하여 쉽게 시료가 오염될 수 있어 전처리가 여전히 필요하다. 땀은 소변, 침에 비해 매우 순수한(가장 정제된) 시료에 해당한다. 땀의 99%는 물이고 이온들과 질소 함유물, 젖산, 요소 등이 포함되어 있다. 그러나 소변처럼 정상 대사산물이 화학물질로써 다량 섞여 있지 않기 때문에 마치 전처리가 이루어진 시료처럼 마약을 검출할 수 있다. 하지만, 땀으로 배출되는 마약의 농도가 소변과 비교하면 현저하게 낮다는 단점이 있다.

약물마다 반감기와 배출 농도가 상이하여 직접 농도를 언급하는 데 어려움이 있지만, 시판되는 소변검사 키트의 경우 검출 한계치가 수십 ng/mL에 해당하기 때문에 땀 속 마약류를 검출하기 위해서는 더 민감도가 좋은 기술을 개발해야 한다. 이러한 사항 때문에 미국 Pharmcheck社에서도 땀 속 마약을 패치로 수집만 하고, 분석은 LC-MS/MS로 시행하고 있는 이유이기도 하다.

그림 14. 생체시료별 마약 검출 가능 시간



\* 출처: Shamil et al.(2021)



# IV. 현장형 마약 검출 기술 연구 동향

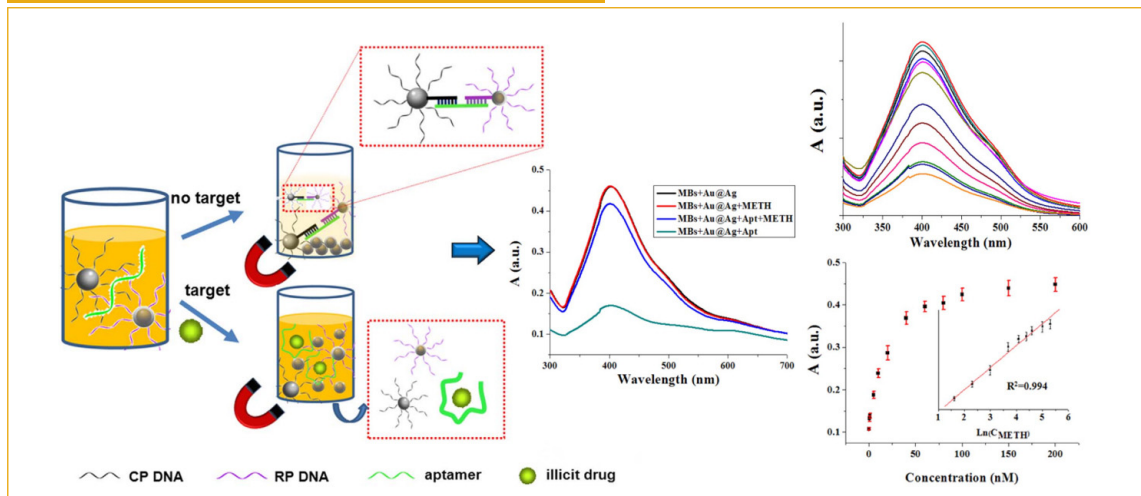
## 1. 기술별 마약 검출 연구 동향

### 1.1 비색(Colorimetric) 센서

비색센서는 마약과 반응한 센서가 발색 혹은 색 변화(변색)를 일으키는 방법으로 마약을 검출할 수 있어서 직관적이고 눈으로 결과를 확인할 수 있다는 장점이 있다. 측방유동면역분석방법이 가장 대표적인 비색센서에 해당하고 이미 상용화가 되어 있다. 앞에서 측방유동면역분석법의 원리는 설명하였기 때문에, 이를 제외한 비색센서 연구 동향을 소개하고자 한다. 색이 나타나거나 변화한다는 것은, 마약과 결합한 물질이 마약 대신 발색을 하거나 분자구조 변화로 색 변화를 일으킨다는 의미이다.

아래 <그림 15>는 중국 우한대학교 연구팀이 귀금속 나노입자에 DNA 서열(DNA sequence)을 부착하여 마약 존재 여부에 따라 색 변화를 유도한 연구이다. 해당 연구에서는 메스암페타민과 결합하는 aptamer (Aptamer, 특정 물질과 결합할 수 있도록 설계된 핵산 물질)와 DNA가 결합된 귀금속 나노입자를 함께 용액에 섞는다. 메스암페타민이 없는 경우, 귀금속 나노입자의 DNA가 aptamer와 결합하여 입자끼리 가까워지도록 유도한다. 반대로 메스암페타민이 있는 경우, aptamer가 귀금속 나노입자의 DNA보다 메스암페타민에 반응하기 때문에 입자들이 서로 붙지 않는 상태로 용액 내에 존재하게 된다. 귀금속 나노입자는 가시광 영역에서 색을 띠게 되는데, 이들이 서로 결합하거나 거리에 따라서 색깔이 달라지는 효과가 나타난다. 이러한 원리로 본 연구에서는 메스암페타민을 0.1nM(14.9ng/L)까지 검출하는 데 성공하였다. 하지만, 용액 기반의 비색센서의 경우 현장에서 피펫(스포이트)이나 용기를 이용해야 한다는 단점이 있고, 나노입자가 장기간 유통되기에 안정성이 떨어진다는 한계가 있다.

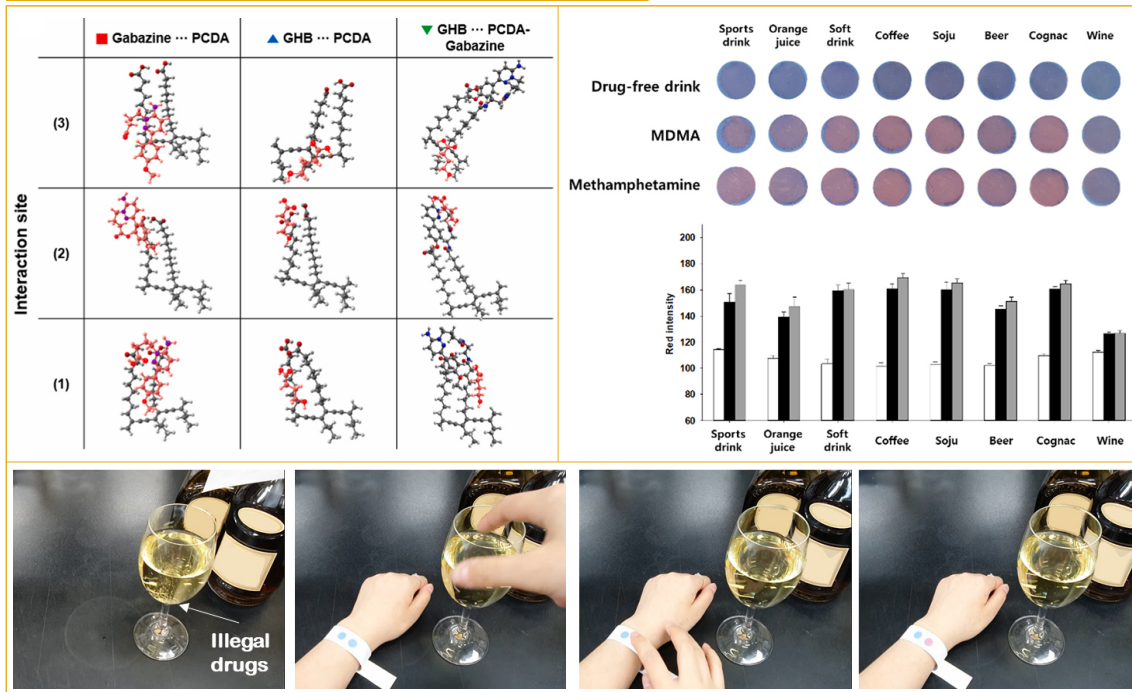
그림 15. 귀금속 나노입자와 핵산을 이용한 비색센서 작동원리



\* 출처: Mao et al.(2017)

한국생명공학연구원에서는 GHB, 암페타민과 결합할 수 있는 분자인식 고분자를 개발하여 종이에 코팅하고 이를 현장에서 사용할 수 있는 키트 형태로 개발하였다. 분자인식 고분자는 타겟 마약성분과 만났을 때 구조적 변화를 일으키는데, 해당 구조변화가 색깔의 변화로 나타나 시료 내에 마약류가 포함되었는지를 판별할 수 있다. 실제 음료 8종에 포함된 마약류 검출을 시도하였으며, 팔찌, 컵 받침 형태로 개발하여 성범죄 목적으로 음료에 마약을 넣었을 경우 당사자가 직접 테스트 할 수 있도록 제안하였다.

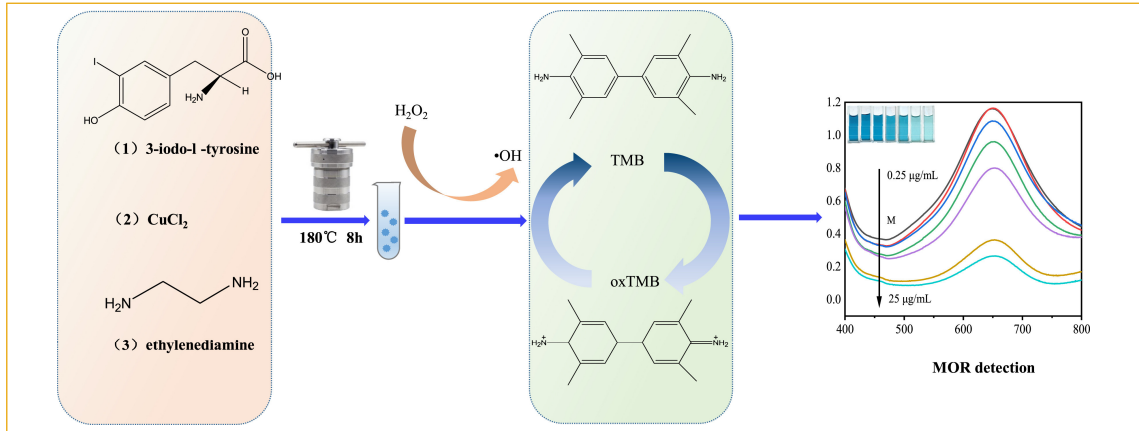
그림 16. 분자인식 고분자를 활용한 비색센서 및 현장형 사용 예시



\* 출처: (상단 좌) Son et al.(2021), (상단 우) Jang et al.(2022), (하단) 한국생명공학연구원 제공

발색을 이용하는 또 하나의 방법은 효소를 이용하는 것이다. 보통 산화환원반응을 일으키는 효소들에 의해 3,3',5,5'-티트라메틸벤지딘(TMB, 3,3',5,5'-Tetramethylbenzidine)과 같은 발색 물질들의 구조가 변화하면서 색을 생성한다. 이러한 방법은 질병 진단에 가장 많이 쓰이는 바이오에세이 방법인 효소결합면역흡착검사(ELISA, Enzyme- Linked ImmunoSorbent Assay)에서 활용한다. 아래 <그림 17>은 중국 군명대학교에서 개발한 기술로, 구리가 도핑된 탄소 양자점을 모르핀과 함께 열을 가해주었을 때, 마치 효소와 유사한 역할로 TMB 구조변화를 유도하는 결과를 얻었다. 모르핀의 농도에 따라서 발색 정도가 달라지고, 발색 된 색의 진하기를 기반으로 모르핀을 64ng/mL까지 검출하는 데 성공하였다.

그림 17. 구리가 도금된 탄소양자점을 이용한 모르핀 검출 기술



\* 출처: Yin et al.(2023)

상기 비색센서 기술들은 발색 혹은 변색을 기반으로 정량평가를 실시해야 하는데, 타겟 마약류의 농도가 낮으면 이를 눈으로 분간하기 어렵다. 마치 코로나-19 감염 초기에 검사선이 희미하게 보여서 확진에 해당하는지 아닌지 구분하기 힘든 것과 같은 현상이다. 따라서, 이러한 정량평가 한계를 극복하기 위해, 이미지를 기반으로 빨간색, 녹색, 파란색(RGB, Red, Green, Blue) 색 비율 변화를 관찰하고 이를 휴대폰에 앱으로 설치하여 분석에 활용한다. 경우에 따라 휴대용 자외선 및 가시선 분광분석기를 활용하여 스펙트럼으로 색 변화를 관찰하기도 한다.

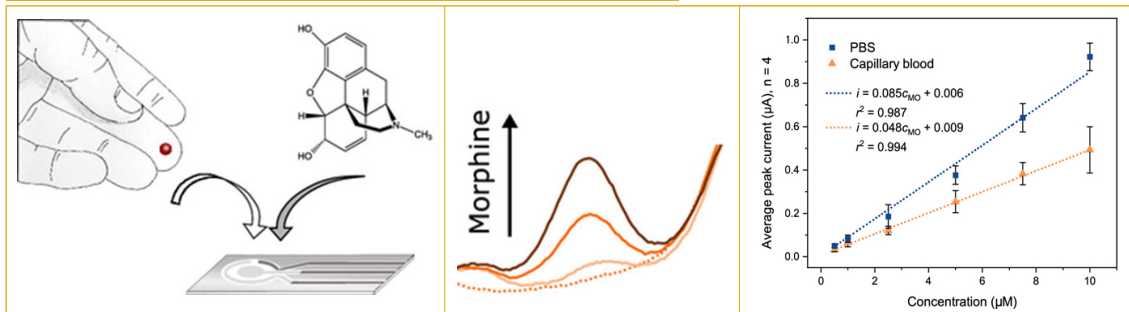
## 1.2 전기화학(Electrochemical) 센서

전기화학 센서는 타겟 물질이 센서 전극 상에서 반응하여 생성되는 신호를 이용하여 물질을 검출하는 센서를 의미한다. 전류, 전압, 전도도, 저항 변화를 감지하여 마약류를 검출할 수 있고 높은 민감도, 간편성, 빠른 검출 속도, 휴대용 검출기기로 시스템화할 수 있다는 장점이 있다. 검출 타겟에 색이 있는 경우, 비색센서는 분석에 한계를 나타내는데 전기화학 센서는 광학적 간섭을 거의 받지 않기 때문에 센서로써 많은 장점이 있다. 다만, 복잡한 방해 물질들이 섞여 있거나, 다양한 마약을 동시에 검출해야 하는 경우에는 신호 분석과 해석에 어려움이 있다는 단점이 있다.

아래 <그림 18>이 나타내는 연구는 핀란드 알토 대학교에서 개발한 전기화학 센서 기술로, 전처리하지 않은 혈액 내에 포함된 모르핀을 선택적으로 검출하였다. 전극 표면에는 탄소나노튜브가 코팅되어 있고, 나피온(Nafion) 고분자(불소계 수소이온 전도성 고분자 전해질)를 센서 최상단에 코팅하여 혈액에서 직접 타겟을 검출할 수 있도록 설계하였다. 모르핀 분자에 대해 차동 펄스 전압전류(DPV, Differential Pulse Voltammetry)를 측정하였고, 모르핀 농도가 증가함에 따라 전류 값이 선형적으로 증가하는 것을 확인하였다. 많은 연구의 경우 혈액 내 존재하는 혈구 세포, 혈장 단백질 등을 분리하기 위해 전처리를 수행하는데, 해당 연구는 전극의 구조를 디자인함으로써 전처리 없이 혈액 내 모르핀을 0.5µM 까지 검출하는 데 성공하였다.

다만, 타겟이 모르핀이 아닌 경우 전류 값이 변화하는 전위(포텐셜, potential) 위치가 바뀔 것이고, 혈액 내 다른 성분들과의 방해 현상 때문에 다양한 마약류에 적용할 수 있을지는 더 연구가 필요하다.

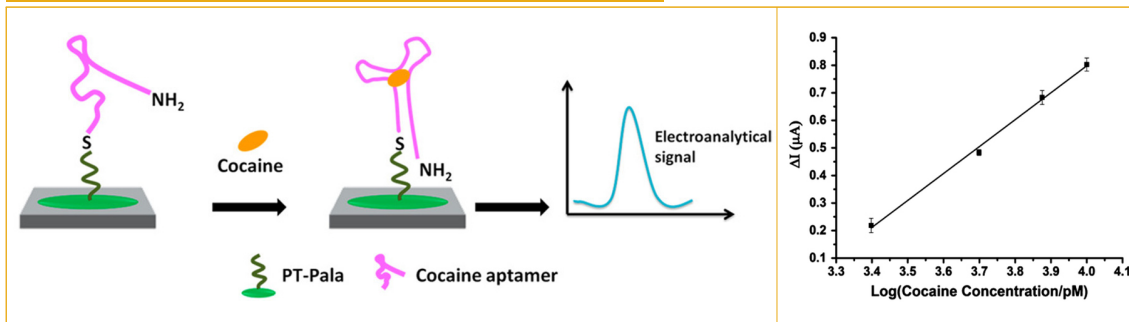
그림 18. 탄소나노튜브가 코팅된 전극을 이용한 DPV형 전기화학 센서



\* 출처: Verrinder et al.(2021)

전기화학 센서의 또 다른 장점은 전극에 타겟 물질과 선택적으로 결합할 수 있는 수용체를 쉽게 활용할 수 있다는 것이다. 모든 센서는 수용체를 활용할 수 있지만, 전기화학센서의 경우 수용체가 전극 표면에 위치해 있어 타겟과 만나 발생하는 산화 환원 반응을 쉽게 유도할 수 있다. 아래 <그림 19>가 나타내는 연구는 터키 이스탄불 공과대학교에서 개발한 기술로, 유리 탄소 전극(GCE, Glassy Carbon Electrode) 전극 표면에 코카인과 선택적으로 결합할 수 있는 압타머 수용체를 고정하였다. 코카인이 전극 위에 존재할 시 압타머의 3차원 구조가 변화하게 되는데, 이때 전극표면에서 산화 환원 프로브인  $Fe(CN)_6^{3-/4-}$ 의 전자 전달을 방해하는 원리로 코카인을 정량하였다. 해당 연구에서는 코카인이 인체 내에서 대사되어 형성되는 대사체 중 하나인 벤조일렉고닌(benzoylcegonin)을 함께 검출하였다는 데 의미가 있다. 보통 인체 내에 마약류가 투여되면 원래의 마약 형태와 일부 대사된 대사체 형태로 함께 배출되는 경우가 많은데, 이를 동시에 검출할 수 있기 때문에 체내 대사 과정을 추적하여 복용 시점을 예측할 수 있는 장점이 있다. 해당 기술은 코카인 대사체를 1.5nM까지 검출하는 데 성공하였다.

그림 19. 압타머를 이용한 코카인 및 코카인 대사체 검출 센서 원리



\* 출처: Dashtian et al.(2023)

다음은, 입을 수 있는 형태인 웨어러블 전기화학 센서를 이용한 약물 모니터링 기술을 소개하고자 한다. 인체는 소변, 침, 눈물, 땀 등 다양한 생체시료를 배출한다. 따라서 몸에 붙일 수 있는 형태의 센서는 배출되는 생체시료 내에 포함된 약물을 직접 검출하고 모니터링할 수 있다는 장점이 있다. 마이크로니들 형태의 센서는 피하에 있는 간질액을 추출할 수 있고, 포함된 약물의 신호를 검출하여 피부를 통한 약물 검사가 가능하다. 반지 형태의 센서는 침에 포함된 테트라하이드로칸나비놀(THC, Tetrahydrocannabinol)이나 알코올을 검사하고, 안경에 미세유체 칩이 포함된 센서는 눈물 속 알코올 측정도 가능하다. 전자회로를 부착한 전기화학 센서는 땀과 함께 배출되는 약물과 알코올을, 손가락 스트립 형태의 센서는 터치하는 방식으로 코카인 존재 여부를 확인할 수 있다. 이러한 방법들은 모두 타겟 물질들이 전극에 닿았을 때 나타나는 산화 환원 반응에 의한 저항, 전류, 전압 변화 등을 확인하는 방법으로 이루어진다. 흥미로운 것은 몸에 부착한 센서가 형성하는 신호를 무선통신으로 전달받아 해석하는 시도들이 이루어지고 있다. 전기화학 센서는 소형화가 용이하기 때문에, 향후 용의자가 인지하지 못하는 상황에서 만지는 것만으로 작은 센서가 마약의 흔적을 찾아내는 데 사용될 수 있을 것으로 판단된다.

그림 20. 웨어러블 전기화학 센서를 이용한 약물검출 기술들



\* 출처: Teymourian et al.(2020)

### 1.3 광학(Optical) 센서

광학 센서는 빛을 이용하여 타겟 분자가 센서 표면에 반응한 신호를 확인하여 검출 여부를 판단하는 방법이다. 비색센서도 가시광 영역을 이용하기 때문에 큰 카테고리에서는 광학 방식으로 간주할 수도 있지만, 이번 장에서는 형광 및 라만분광 방법을 이용한 연구 동향을 소개하고자 한다.

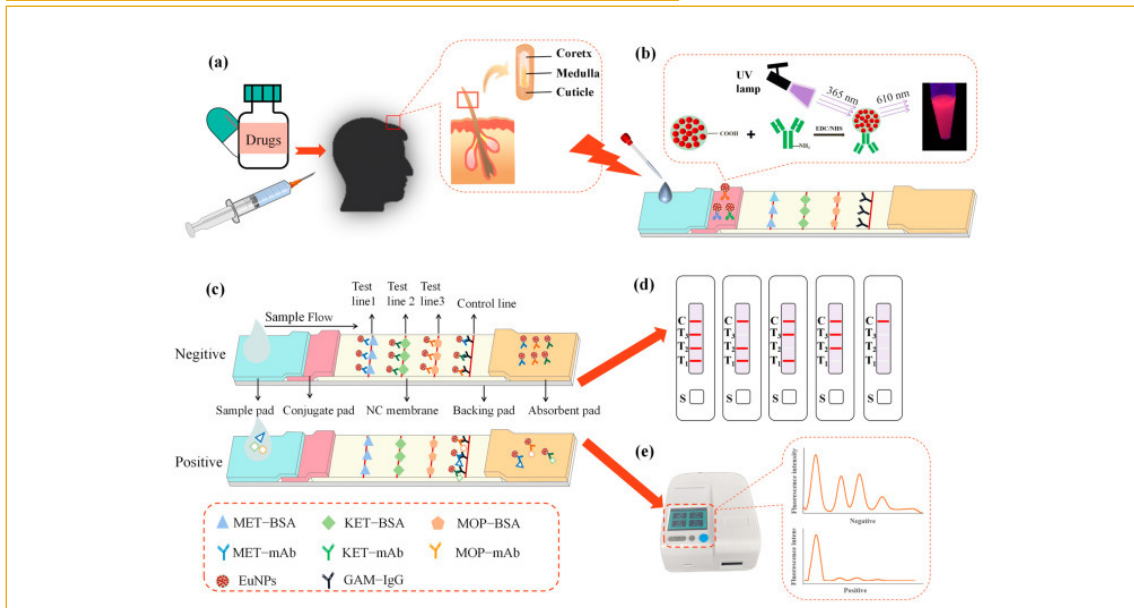
형광은 비색센서에 비해 광학적으로 좋은 장점을 가지고 있다. 비색센서의 단점 중 하나가 타겟 농도가 낮으면 발색 혹은 변색 정도가 작아 이를 직관적으로 판단하기에 어려움이 있다는 것이다. 그러나 형광은 우리 눈으로 직접 확인하지 않고 형광 현미경이나 스펙트로미터로 신호를 확인한다. 장비를 이용해야 한다는 단점은 존재하지만, 형광물질 이외의 성분들은 대부분 형광 신호를 나타내지 않기 때문에 신호의 대비(contrast)가 매우 우수하다. 따라서, 방해 물질들이 존재하는 복잡한 시료 내에서도 타겟 물질에

형광물질을 선택적으로 결합할 수 있다면 극미량의 마약도 검출할 수 있다는 장점이 있다. 형광 검출기 또한 단일 파장만 활용할 때 대당 가격이 100만 원 수준으로 낮아지고 있어 현장 활용성이 더욱 증가하고 있다.

측방유동면역분석법은 앞서 소개한 바와 같이 비색센서로 가장 많이 활용되고 있다. 하지만 검출 민감도를 증가시키기 위해 형광 기반의 센서로 개발되는 사례도 많이 존재한다. 중국 길랴대학교에서는 모르핀, 메스암페타민, 케타민을 동시에 검출할 수 있는 형광 기반 측방유동면역분석기술을 개발하였다.

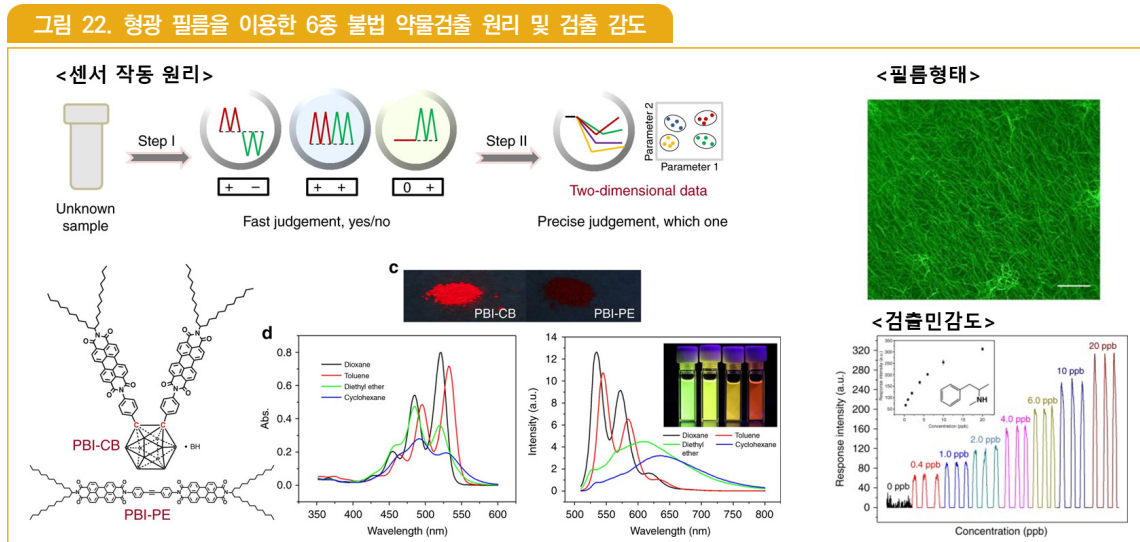
모발에 축적되어있는 세 가지 마약류를 추출한 후, <그림 21>과 같이 샘플 패드(sample pad)에 용적하면 마약 성분이 측방유동에 의해 흐르게 된다. 총 3가지 테스트 라인에는 각각의 마약과 소 혈청단백질이 결합된 복합체, 그리고 마약 물질에 선택적으로 결합할 수 있는 항체에 형광물질인 유로퓸 나노입자(EuNPs, Europium fluorescent NanoParticle)가 결합된 수용체가 마약-혈청단백질에 고정된 상태로 준비하였다. 이후, 타겟 마약이 지나가게 되면, 검출 선에 붙어있던 항체-형광물질 복합체가 흐르는 마약에 더 높은 결합력으로 반응하여 형광 신호가 사라지게 된다. 예를 들어, 메스암페타민이 모발 추출액에 포함된 경우, 메스암페타민 항체-형광물질 결합체가 검사선에 있다가(이미 형광신호가 형성된 상태) 흐르는 메스암페타민을 만나 검사선에서 떨어져 함께 흘러 검사선의 신호가 줄어드는 원리이다. 이는 마약류의 농도에 따라 이미 붙어있던 형광물질 신호가 줄어드는 것으로 정량화가 가능하다. 특히, 총 3개의 테스트 라인을 형성하고 있어서, 모발에 3개 중 어떤 마약이 흐르더라도 줄어드는 신호에 해당하는 마약을 검출할 수 있다. 해당 기술은 모르핀, 케타민, 메스암페타민에 대해 각각 0.219ng/mL, 0.079ng/mL, 0.329ng/mL의 검출 감도를 나타내었다. 기존 비색방법의 측방유동면역분석법의 검출 한계가 수십 ng/mL에 해당하는 것을 고려하였을 때, 형광 기반의 검출법이 우수한 검출 민감도를 나타낸다는 것을 보여주는 사례가 된다.

그림 21. 3종 마약 동시 검출이 가능한 형광 기반 측방유동면역분석법



\* 출처: Xu et al.(2023)

중국 산시사범대학교는 형광 신호를 발생하는 필름 물질을 개발하여 메스암페타민, 엑스터시, 케타민과 같은 마약류뿐만 아니라, 메스암페타민·카페인 복합물, 카페인, 페노바비탈(운동선수들이 복용하는 불법 약물)을 검출하는 데 성공하였다(〈그림 22〉 참고). 해당 필름은 페릴렌 비스이미드 유도체(PBIs, Perylene Bisimide derivatives)로 구성되어 있는데, 유리나 플라스틱 표면에 코팅하여 센서 부재로 사용할 수 있다는 장점이 있다. 총 6개의 약물에 대해 필름이 반응 시 형성되는 형광 신호의 변화, 농도별 인텐시티(Intensity) 변화로 약물을 정량 검출할 수 있다. 6개의 약물이 서로 다른 신호 변화를 형성하는데, 이를 통계적 방법을 이용하여 신호 차이를 구분해 내었다. 다만, 해당 기술은 마약류를 기체화시켜서 측정해야 하는 한계가 있다. 기체화시키기 위해서는 열을 가하거나 시료를 증발시켜야 하는데, 이때 시간이 소요되기 때문에 현장에서의 활용에는 한계가 있다. 그럼에도 ppb(Parts Per Billion) 수준까지 마약류를 검출할 수 있다는 점에서 고감도 검출 센서 개발의 원천기술로 활용될 것으로 기대된다.



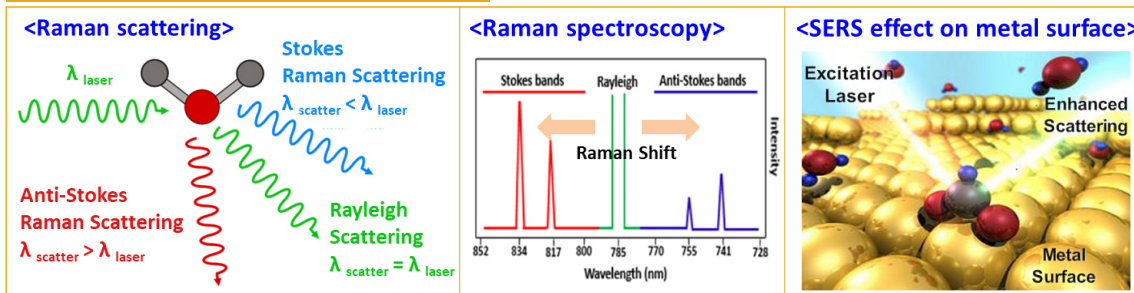
\* 출처: Liu et al.(2018)

광학 센서 중에는 플라즈모닉 소재를 이용한 광학신호 변화를 통해 타겟 분자를 검출하는 표면 플라즈몬 공명(SPR, Surface Plasmon Resonance), 플라즈몬 강화 형광(PEF, Plasmon-Enhanced Fluorescence), 표면증강라만산란(SERS, Surface-Enhanced Raman Scattering) 등의 기술들이 존재한다. 플라즈모닉 소재란, 금, 은과 같이 자유전자를 많이 포함하는 귀금속 중에서, 빛에 의해 자유전자의 치우침으로 국소적인 전자기장을 형성할 수 있는 소재들을 의미한다. 이중 SPR의 경우 가장 전통적인 방법으로, 타겟 물질이 플라즈모닉 소재 표면에 존재하는 수용체에 결합되면 변화하는 굴절률 차이로 해당 타겟의 결합 상수, 검출을 실현할 수 있다. 하지만, 미세유체와 소형화된 장비를 구현해야 한다는 한계가 여전히 존재하여, 현장형 센서보다는 타겟과 수용체의 결합력을 측정하는 목적으로 많이 사용된다. PEF의 경우 앞서 언급한 형광 신호를 증폭하는 현상에 해당하는데, 플라즈모닉 소재 표면에 일정 거리 이상 형광물질이 존재할 경우,

플라즈모닉 소재의 전자기장 증폭 현상에 의해 여기 전자가 존재할 수 있는 상태밀도함수(DOS, Density Of State)가 증가하여 마치 형광물질의 발광(emission)이 늘어나 광량 증가, 즉 형광 신호의 세기가 커지는 형태로 나타난다. 이는 형광 기반 센서에서 더 극저농도의 타겟을 검출할 때 용이하게 쓰이는 기술에 해당한다. 하지만, 검출 원리상 앞서 언급한 형광 기반 센서와 동일하기 때문에, 현재 가장 주목받고 있는 SERS 기술을 이용한 마약 검출 연구동향에 대해 소개하고자 한다.

SERS 기술은 기본적으로 라만분광신호를 증폭하는 기술이다. 라만분광법은 분자가 빛을 조사받았을 때 극히 드문 확률로 에너지를 조금 얻거나 잃는 현상, 즉 비탄성 산란(inelastic scattering)에 의해 변화하는 에너지를 스펙트럼으로 나타낸 것이다. 라만스펙트럼은 분자의 고유 지문에 해당할 정도로 분자에 특이적인 스펙트럼 특성을 나타내는데, 예를 들어, 특정 마약마다 가지고 있는 라만스펙트럼은 고유하다. 따라서, 라만분광신호를 얻기만 하면 추가적인 분석 없이도 데이터베이스의 라만분광신호 패턴과 비교하여 해당 물질이 무엇인지 직관적으로 판단할 수 있다. 라만분광기술이 좋은 분자 정보를 가지고 있음에도 불구하고 센서로 많이 사용되지 못하였던 이유는, 비탄성 산란이 극히 드물게 일어나기 때문에 해당 신호가 매우 작다는 단점이 있기 때문이다. 현재 현장에서 라만분광기를 사용하여 마약류를 검출하는 사례는, 전부 파우더나 액상과 같이 고농도의 마약류 시료의 성분을 분석할 때만 사용되고 있다.

그림 23. 라만분광신호 형성 원리 및 SERS 효과



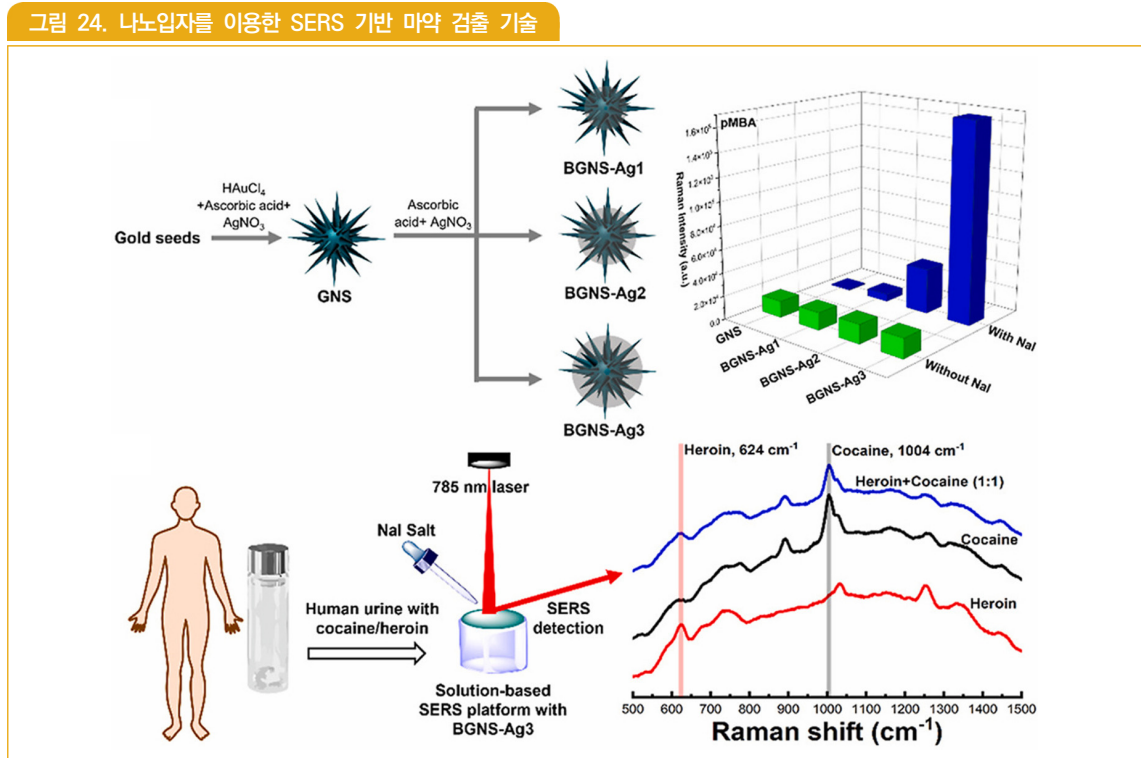
\* 출처: (좌) Bishop(2023), (중) Lohumi et al.(2017), (우) Vrije Universiteit Brussel 홈페이지

하지만, 우리가 센서를 활용하는 이유는 검출하기 어려운 농도나 영역의 미량의 분자를 검출하기 위해서이다. 오감이 느낄 수 있는 정도의 많은 양이라면 굳이 센서를 개발할 필요가 없기 때문이다. 이에 라만분광신호를 증폭하기 위해, 귀금속 나노구조체를 형성하여 SERS 효과를 유도할 수 있다. 라만분광신호의 정보는 유지 하면서, 신호의 크기는 100억 배 이상 증폭되기 때문에 비로소 검출하지 못했던 신호를 기기로 획득하여 센서로 활용할 수 있게 되는 것이다. 다음은 SERS 효과를 이용한 마약 검출 연구를 소개하고자 한다.

미국 듀크대학에서는 사람의 소변으로 배출되는 코카인과 헤로인을 SERS 소재를 이용하여 극미량 검출에 성공하였다(그림 24) 참고). SERS 효과를 나타낼 수 있는 금-은 나노입자 복합체를 소변과 함께 섞어 균집(aggregation)을 유도한다. SERS 효과는 플라즈모닉 소재가 서로 가까이 붙어있을 때 효과가 증가하기 때문에 염 성분을 넣어 균집 형성을 유도하였다. 이후 형성되는 코카인과 헤로인의 고유 라만분광신호를



기반으로 극미량 검출에 성공하였다. 기존 기술들에 비해 획기적인 검출 민감도를 나타내었는데, 코카인의 경우 10pg/mL, 헤로인의 경우 100pg/mL까지 검출에 성공하였다.

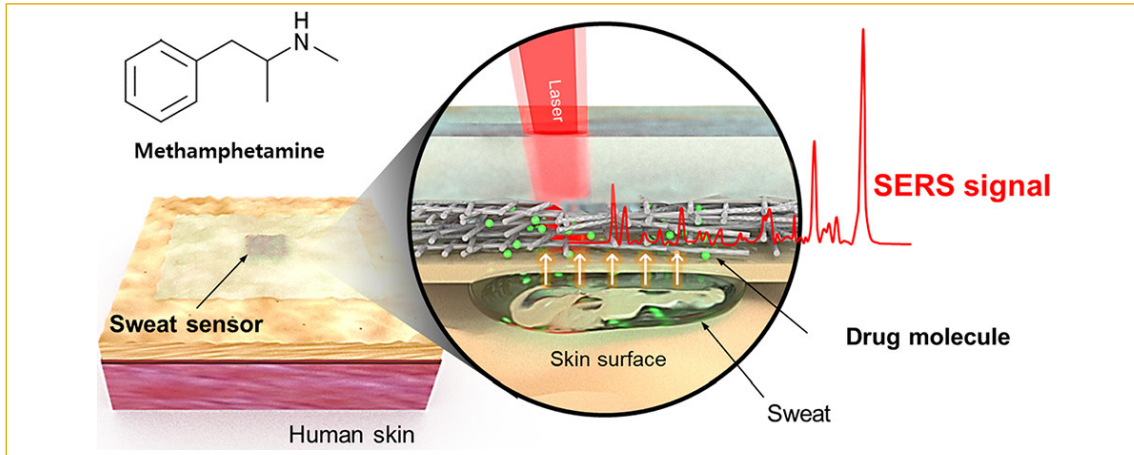


\* 출처: Atta et al.(2023)

한국재료연구원은 땀 속에 존재하는 미량의 마약류를 패치 형태로 검출할 수 있는 SERS 기술을 개발하였다(〈그림 25〉 참고). 땀은 다른 생체시료에 비해 상대적으로 적은 농도의 마약류를 포함하고 있어서 검출 감도가 충분히 높은 기술을 활용해야만 땀 속 마약 흔적을 추적할 수 있다. 피부 부착 시 면역반응을 일으키지 않는 단백질 필름을 저분자(대부분 마약류에 해당하는 분자량 500 이하) 화학물질만 투과할 수 있도록 준비하고, SERS 효과를 나타낼 수 있는 은 나노와이어를 해당 단백질 필름 위에 코팅하였다. 이후 의약품 패치에 전사하여 피부에 부착할 수 있는 형태로 개발하였다. 본 기술은, 기존 땀에서 마약을 검출하는 기술들이 대부분 센서나 패치를 제거한 후에 다시 마약 성분을 추출하여 검사하는 방법과 달리, 피부에 부착된 상태로 센서를 제거하지 않고 휴대용 라만분광기로 메스암페타민을 검출할 수 있다는 것에서 큰 장점이 있다. 본 연구에서는 메스암페타민을 50ng/cm<sup>2</sup>까지 검출하는 데 성공하였다. 마약사범 관리의무가 있는 법무부에서는, 소변검사를 통해 일정 기간 내에 마약을 재투여하였는지 소변검사를 통해 실시한다. 하지만, 정기적 검사 간격이 수주에서 수달 정도이기 때문에 반감기가 짧은 마약류의 경우 다음 검사까지 몸에서 모두 배출되어 흔적을 찾을 수 없게 되는 경우가 발생한다. 하지만, 해당 기술의 경우 패치 센서를

붙이고 있는 기간 동안 단 한 번이라도 마약을 투여했다면, 땀으로 배출되는 마약 성분이 패치에 축적되어 남아있으므로 관리기간 동안 마약사범의 재범 여부를 판단하는데 용이한 기술로 활용될 수 있다. 해당 기술은 마약뿐만 아니라 올림픽이나 국제 스포츠 대회에서 선수들이 불법 약물 투여 여부에 대해 상시 모니터링이 가능한 기술로 활용될 수 있음을 제안하였다.

그림 25. SERS 효과를 이용한 웨어러블 마약 검출 기술



\* 출처: Koh et al.(2021)

본 장에서는 크게 3가지 검출 방법인 비색센서, 전기화학 센서, 광학 센서를 이용한 마약 검출 기술 연구동향에 대해 소개하였다. 소개된 연구들은 실제 마약류를 검출한 사례를 다루었으나, 마약 외의 화학물질을 검출할 수 있는 연구개발이 폭발적으로 쏟아져 나오고 있다. 이를 마약 검출에 적용하지 않았을 뿐, 연구개발 단계에서 마약을 검출할 수 있는 원천기술들은 이미 많이 존재한다고 판단된다. 마약 검출 기술을 개발할 때 가장 먼저 마주하게 되는 문제는, 연구목적으로 마약을 확보하는 것이다. 국내의 경우 연구자용 마약을 구하기 위해서는 관할 식품의약품안전처에 승인받은 후 구매하여 연구에 사용하게 되는데, 이때 마약과 관련된 연구개발 과제가 존재해야 신청을 할 수 있다. 단순 실험 목적으로 마약을 구하기는 현실적으로 어려운 것이 사실이다. 또한, 마약 검출 기술의 파급효과가 다른 약물검출, 질병 진단용 센서에 비해 상대적으로 비중이 작은 것 또한 해당 연구 분야 활성화가 활발하지 않은 이유 중 하나다. 마약 검출 기술은 공공의 목적으로 개발되어야 하고, 특히 중독된 이후에 중증 질병처럼 삶의 질에 큰 영향을 미치는 만큼 정부가 연구개발에 과감한 투자와 원천기술 확보에 앞장서야 한다고 판단된다.

## V. 현장형 마약 검출 기술 상용화 현황

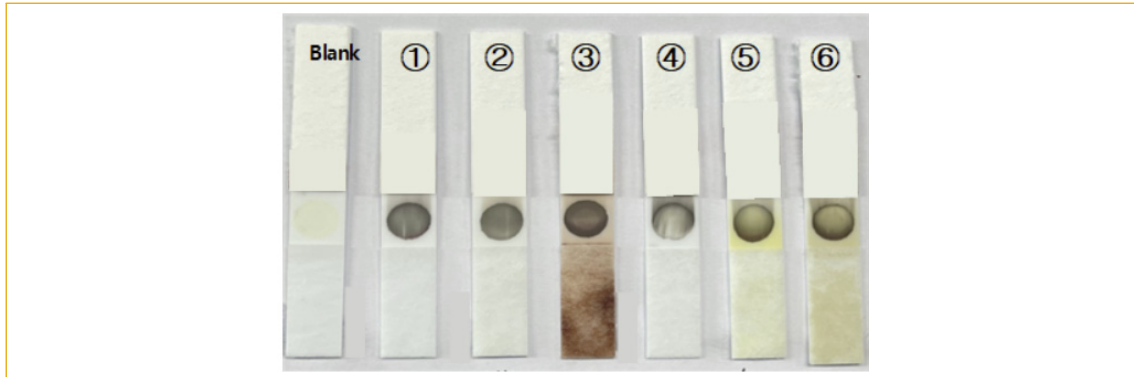
### 1. 국내 상용화 현황

국내에서는 소지 마약을 검사할 수 있는 이온스캐너, 라만분광기 등을 국산화하는 연구개발과 시도들이 이루어졌다. 외산 제품보다 저렴하게 기기를 보급할 수 있다는 장점이 있지만, 마약검사의 특성상 법적인 문제로 귀결되는 만큼, 정확도와 신뢰도가 중요한 지표가 된다. 동일 원리의 검출기기를 국내 기업에서 제품화하더라도, 실제 현장에서는 상기 문제들, 그리고 각 마약 성분들에 대한 검출 라이브러리 확보 등의 어려움 때문에 결국 외산 제품을 선택하게 되는 경우가 많다. 복용 후 마약검사의 경우도 국내 기업들이 측방유동면역분석법을 이용하여 소변에서 마약류를 검사할 수 있는 제품을 상용화한 사례가 있다. 해당 방법은 마약에 대한 항체만 존재하면 얼마든지 검사 마약 종류와 제품화가 가능하므로 상대적으로 진입장벽이 높지 않다. 코로나-19 시기에 국내 많은 기업들이 동시에 앓다투어 자가진단 키트를 개발할 수 있었던 이유도, 동일한 측방유동면역분석법을 이용하고 코로나-19 항원에 대한 항체가 발굴되었기 때문에 가능하였던 것과 마찬가지로이다. 이러한 상황에서는 기업들이 제품을 출시하더라도, 가격 경쟁력 확보를 통한 시장 진출 밖에 방법이 없다. 항체를 포함한 원재료 공급망의 수월성이 제품의 가격 경쟁력으로 그대로 나타나게 된다. 또한, 국내 마약검사 시장이 해외보다 현저히 작은 이유와 그리고 해외에서는 이미 외산 제품들이 시장을 선점하여 국내 기업들이 시장 점유율을 높이기 어려운 상황 때문에, 국내 기업들이 기술력을 충분히 가지고 있다 하더라도 시장 확보가 명확해지지 않으면 상용화에 나서지 않게 된다. 따라서, 국내 원천기술이 상용화되기 위해서는 새로운 검출기술이나 방법을 제시함으로써 기존 기술 대비 성능이 우수하거나 가격이 저렴해야 한다. 예를 들어, 하나의 키트로 여러 종의 마약을 동시에 검사할 수 있는 기술, 다양한 생체시료에서 마약을 검출할 수 있는 키트, 항체보다 더 저렴한 원재료를 이용하여 검사하는 방법 등 기술적 진보가 이루어져야 시장진입이 가능할 것으로 판단된다.

최근, 경찰청이 성균관대학교 과학수사학과와 함께 마약 성분을 측석에서 검사할 수 있는 휴대용 마약 검사 키트를 개발하여 올해부터 실용화 단계에 착수한다. 해당 키트는 음료에 포함된 마약류를 검사하기 위해 음료 속에 담가 사용하는 ‘스트립형’과 음료를 찍어서 종이에 펴 바르는 ‘스티커형’ 두 가지로 개발된다. 언론에 의하면 2023년 9월경 일선 경찰에 보급되어 기술을 검증할 예정이고, 16종의 마약을 검사할 수 있다고 한다. 해당 기술의 공개특허를 살펴보면, 헥사클로로플라틴산칼륨(Potassium hexachloroplatinate)이 측방유동면역분석법에 쓰이는 항체 대신 마약류와 반응하여 색 변화를 이루도록 설계되었다.

공개된 특허의 실시 예를 살펴보면, 소주 안에서 암페타민에 대해 0.05(v/v)% 까지 테스트하였고, 색 변화를 농도별로 관찰하였다. 복잡한 음료 내에서 마약류를 간단하고 신속하게(10~15분) 검출할 수 있는 기술로 판단되나 실제 보급이 이루어진 후 다양한 현장 활용 가능성, 검출 민감도에 대한 평가가 필요한 상황이다.

그림 26. 성균관대학교 개발 마약 간이검사 키트 결과



\* 비색법: 1. 소주, 2. 맥주, 3. 레드와인, 4. 화이트와인, 5. 오렌지주스, 6. 콜라

\*\* 출처: 매일경제 기사(2023.04.16)

최근, 2020년부터 국내 기업 이노릭스社は 소변 마약검사, 타액 마약검사, 모발 마약검사 서비스를 수행하고 있다. 신속 검사와 연구실 검사로 나누어 서비스하고 있고, 신속 검사의 경우 타액 20종, 소변 10종, 모발 14종까지 가능하다. 하지만, 공개된 정보를 바탕으로 종합해보면 새로운 현장형 검사 기술을 개발하였다기보다는 기존 표준 기술들을 활용하여 검사 서비스를 제공하는 것으로 판단된다. 그럼에도, 국내 마약 범죄가 급증하고 있는 상황에 이를 검사할 수 있는 서비스와 기술을 제공한다는 점에서 향후 국내 시장을 주도할 수 있을 것으로 기대된다.

그림 27. 이노릭스社에서 제공하는 신속 및 연구실 마약검사 서비스

	타액	소변	모발	손발톱
신속검사	신속검사 (20종 마약검사)	1종 (개별 약물)  6종 (MET, THC 외 4종)  울트라 센서티브 10종	신속검사 (14종 마약검사)	-
연구실 검사	연구실 검사 20종	연구실 검사 20종	연구실 검사 20종	연구실 검사 20종

\* 출처: 이노릭스社 블로그

## 2. 국외 상용화 현황

앞서 2장에서 소개한 검사장비 및 기기들은 대부분 외산 장비에 해당한다. 장비 성능뿐만 아니라 마약류에 대한 라이브러리, 검사 서비스 등 해외에는 다양한 사업군들이 이미 존재한다. 이는 기술적으로 국내보다 국외 대형 회사들이 이미 시장을 선점한 이유도 있지만, 마약 관련 제품과 서비스가 많다는 것은 마약 관련 제조, 유통, 오남용 등 관련 범죄가 미국과 같은 나라에서 많이 일어난다는 환경적 요인, 그리고 주(state)마다 마약에 관한 법률이 다르게 적용되고 있어 검사의 필요성과 시장이 크다는 이유도 함께 작용한다. 소변, 타액 등을 이용한 현장형 측방유동면역분석 키트의 경우도 해외에서 먼저 상용화되었다. 미국 AdvaCare Pharma社의 AccuQuick 검사 키트, Confirm Bioscience社의 검사 키트와 같이 다종의 마약류를 동시에 검사할 수 있는 제품이 전 세계로 수출되고 있으며, 우리나라의 경우 AccuSign DOA 제품을 간이 마약 검사 키트로 경찰들이 많이 사용한다.

그림 28. 외산 소변 및 타액 검사 키트 제품들



\* 출처: Advacare Pharm 및 Confirm Bioscience社 홈페이지

최근에는 유엔 마약 범죄 사무소(UNODC, United Nations Office on Drug and Crime)에서 200개의 약물과 약물 전구체를 비색방법으로 검사할 수 있는 테스트 키트를 보급하고 있다. 다양한 약물들에 대해 현장에서 테스트할 수 있도록 시약들을 제공하고, 의심 마약류에 대해 검사하는 방법에 대한 가이드라인도 함께 제공한다. 다양한 마약류를 검사할 수 있는 패키지를 제공한다는 장점이 있지만 측방유동면역분석법과 같이 단순히 시료만 용적하여 검사 결과를 판별할 수 있는 기술에 비해 상대적으로 검사의 전문성과 난이도가 높은 편에 해당한다.

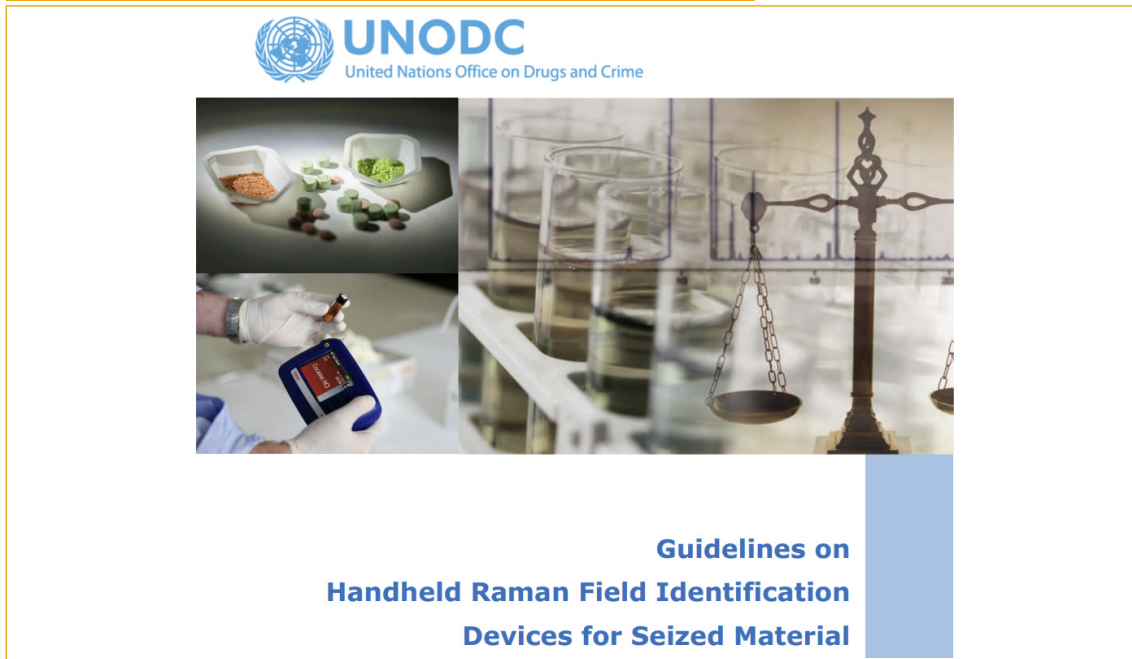
그림 29. UN에서 보급하는 200종의 마약 및 마약 전구체를 검사할 수 있는 키트 패키지



\* 출처: UNODC 브로슈어

UNODC에서는 휴대용 라만분광기를 이용한 마약 검출법에 대한 가이드라인도 제공하고 있다. 아직까지는 파우더, 알약과 같이 벌크(bluk) 시료에 대한 마약 성분과 라이브러리 매칭을 통한 검사법을 표준화하고 있지만 앞으로는 앞서 Metrohm社, SERSTECH社와 같이 SERS 기술과 휴대용라만분광기가 결합된 형태로 액상, 고체상, 기체상, 가공품 형태의 마약류에 대해 고감도로 검출할 수 있는 기술로 나아갈 것으로 기대된다.

그림 30. UN에서 제공하는 휴대용 라만분광기를 이용한 약물 검사 가이드라인



\* 출처: UNODC 브로슈어

## VI. 현장형 마약 검출 기술 활용방안 및 전망

### 1. 마약류 검출 기술의 활용방안

마약류를 현장에서 신속·고감도로 검출해야 하는 기술이 필요한 이유는, 용의자가 현장에서 법적 구속력을 피해 증거인멸, 도주, 추가 범행으로 이어질 수 있기 때문이다. 특히 마약류의 경우 성범죄를 포함하여 한 번 노출되면 한 개인의 인생이 망가질 수 있을 정도로 중증 범죄로 이어질 수 있어서 사전에 범죄 확산을 막는 것이 매우 중요하다. 앞서 밝힌 대로 국립과학수사연구원에서 이 모든 검사를 담당하기에는 무리가 있다. 국립과학수사연구원이나 국내 수사기관에서 마약검사를 담당하는 전문 인력은 몇 명 되지 않는다. 그러므로 현장에서 객관적이고 과학적인 간이검사를 통해 법적 구속력을 발동할 수 있는 과학기술 개발이 필요한 시점이다.

마약류 검사 기술이 가장 먼저 적용될 곳은 성범죄 목적으로 마약류를 상대가 알아차리지 못하게 투여하게 하여 중범죄로 이어지는 것을 막는 데 쓰일 수 있다. 또한, 극미량의 마약을 신속·고감도로 검출할 수 있는 기술이 개발된다면 던지기 수법을 포함하여 마약류에 노출된 사람들을 추적하고 유통경로를 파악하는 데 도움이 될 수 있다. 이는 경찰, 검찰에서 충분히 활용할 수 있는 기술로, 국립과학수사연구원을 포함한 전문 수사기관의 업무량 감소와 신속한 용의자 검거에 많은 도움을 줄 수 있을 것으로 판단된다. 특히, 이러한 기술이 적용되면 용의자의 진술과 자백을 받아내는데 용이하게 활용될 수 있고, 이러한 기술이 보급된다는 것만으로도 마약 관련 범죄를 효과적으로 예방할 수 있다는 기대효과 또한 존재한다.

한편, 법무부에서는 마약사범들을 일정기간 동안 보호 관찰해야 하는 의무가 있다. 현직 담당자로부터 전해 들은 바에 의하면, 정기 검진 때까지 마약을 하지 않고 참다가, 귀가하여 바로 마약을 하고, 또 검진 전에 마약류가 검출되지 않도록 하여 여전히 마약중독에서 벗어나지 못하는 사례가 존재한다고 한다. 마약류를 주기적으로 관리할 의무가 있는 경우 소변뿐만 아니라 침, 땀, 모발 등에서 국립과학수사연구원을 거치지 않고도 간이검사를 통해 다양한 마약류 투여 여부를 판단할 수 있는 기술개발이 필요하다. 특히, 보호관찰 기간에 1회라도 마약을 투여할 시 그 흔적을 조사할 수 있는 기술이 개발된다면 아래와 같은 사례를 줄일 수 있는 계기가 마련될 것으로 예상된다.

그림 31. 법무부 마약사범 보호관찰 출석 날 마약 범죄 현황

#### “보호관찰 출석 날, 마약하는 날” 보호관찰 마약사범 잇따라

뉴스종합 | 2023-03-22 10:17

f X ㉠ ㉡

‘보호관찰 중 마약 사례 잇따라  
올해만 보호관찰 마약사범 2474건  
‘초기에 집중된 보호관찰, 재범 여지 준다’  
인력 부족도...보호관찰관 1명당 100건 맡아



\* 출처: 헤럴드경제 기사(2023.03.22)

## 2. 마약류 검출 기술의 산업적 응용 분야 및 정부의 정책

마약류를 신속·고감도로 검출할 수 있는 기술은 마약뿐만 아니라, 질병에 걸린 환자들의 약물 반응과 모니터링 목적의 의료용으로도 확대하여 사용할 수 있다. 개인마다 약물에 반응하는 정도와 물질대사가 다르다. 예를 들어, 암 환자에게 항암제를 투여하는데, 어떤 약은 잘 들을 수도 있고, 약물 사용량을 결정해야 하기도 하고, 상황에 따라 예후 관찰을 위해 이를 모니터링해야 하는 필요성이 있다. 이러한 경우, 투여한 약물의 대사 성분이나 인체 반응을 소변, 타액을 포함한 생체시료를 이용하여 간편하게 검사할 수 있다면 의료비용 지출과 사회적 비용을 절감할 수도 있다. 더 발전한다면, 병원에 방문하지 않고도 본인의 질병 예후 관찰을 일반 보건소나 약국, 나아가서는 집에서 코로나-19 키트 처럼 개인적으로 검사할 수 있는 시대도 오지 않을까 기대해 본다.

추가로 마약류 검사 키트는 마약류뿐만 아니라 폭발물, 유해물질, 농약 등 다양한 규제 물질을 현장에서 신속·고감도 검사하는데 활용될 수 있다. 마약도 심각한 문제지만 우리가 매일 섭취하는 먹거리에 농약이나 유해물질이 포함되어 있는지 소비자는 검사하고 싶은 니즈(needs)가 존재한다. PekinElmer社의 중국기업인 Meizheng社에서는 잔류농약을 검사할 수 있는 키트를 개발하여 판매하고 있다. 농약의 경우 검사 대상이 500종 이상이기 때문에 단일키트가 이를 다 검사하거나 500개의 서로 다른 키트를 만들어서 검사하는 것은 무리가 있다. 그럼에도 사용 빈도가 많은 농약이나 제품군에 대해서 검사를 할 수 있는 기술을 보급한다는 점에서 향후 검사 대상 확장이 충분히 용이한 기술군으로 판단된다. 이처럼 현장형 마약류 검출기술은 마약류 뿐만 아니라 다양한 유해물질의 검사 기술로 확장성을 가지고 있기 때문에 원천기술 개발과 신뢰도가 확보된다면 다양한 시장에 적용될 수 있을 것으로 기대된다.

그림 32. 현장형 농약 검사 키트제품 예시

			
Pesticides Carbendazim Quantitative Rapid Test Kit	Milk Pesticides Residues Fipronil Lateral Flow Rapid Test Kit	Acetamiprid Quantitative Rapid Test Kist for Brown Rice and Wheat	Milk Pesticides Residues Carbendazim Lateral Flow Rapid Test Kit

\* 출처: Meizheng社 홈페이지

한편 2023년 9월 28일자 전해숙 의원이 발표한 자료에 의하면 우리나라에 10세 미만의 마약중독자뿐만 아니라 청소년 및 20대의 마약중독자가 상당히 많다고 한다. 마약중독자를 단순히 검거하는 데만 집중하는 것도 중요하지만 치료를 통해 이들을 사회의 건강한 일원으로 복귀시키는 것도 매우 중요하다는 새로운 관점을 제시하였다. 현장형 마약 검출 기술과 함께, 사회적으로 이들을 단순히 사회적으로 격리시키는 것에만 집중할 것이 아니라, 그 이후의 대책에 대해서도 함께 고민해야 할 시점이 왔다고 판단된다.



표 1. 국민건강보험공단의 연령별 마약중독 치료자 현황

(단위: 명)

구분	2018년	2019년	2020년	2021년	2022년
10세 미만	403	341	208	216	210
10대	370	375	284	430	498
20대	893	1,003	1,184	1,340	1,383
30대	1,030	1,037	982	1,006	1,066
40대	1,246	1,125	966	981	971
50대	1,106	1,041	890	879	841
60대	700	655	559	629	600
70대 이상	1,236	1,135	1,025	1,013	1,032
<b>총 합</b>	<b>6,984</b>	<b>6,712</b>	<b>6,098</b>	<b>6,494</b>	<b>6,601</b>

\* 출처: 국민건강보험공단

최근 과학기술계 R&D 예산을 대폭 삭감한다는 이슈 때문에 연구개발 현장에 많은 혼란이 나타나고 있다. 정부가 마약과의 전쟁을 선포하고 난 이후 마약 수사를 비롯한 마약 범죄 대응 관련 예산은 오히려 증가하였다. 그만큼 정부도 우리 사회의 마약 범죄에 대한 심각성을 인식하고, 관련 수사와 R&D에 중요도를 부여하고 있다는 증거이다. 우리나라가 예전의 마약청정국의 지위를 되찾을 수 있도록 과학기술이 검출 기술뿐만 아니라 중독 치료기술과 함께 다양한 각도로 개발되어 국민건강 회복, 국민들이 마약 관련 범죄 위협에 더 이상 노출되지 않는 건강한 사회를 되찾기를 진심으로 바란다.

#### ✎ 저자소개 정호상(Ho Sang Jung)

##### • 학력

포항공과대학교 신소재공학 박사  
포항공과대학교 신소재공학 학사

##### • 경력

現) 한국재료연구원 선임연구원  
現) 과학기술연합대학원대학교(UST) 부교수

## ●● 참고문헌 ●●

### 〈국내문헌〉

- 1) 대검찰청(2023). 2022 마약류 범죄백서.

### 〈국외문헌〉

- 1) Atta, S. & Vo-Dinj, T. (2023). Ultra-Trace SERS Detection of Cocaine and Heroin using Bimetallic Gold-Silver Nanostars(BGNS-Ag). *Analytica Chimica Acta*, 1251, 340956.
- 2) Dashtian, K., Amourizi, F., Shahbazi, N., et al. (2023). Biosensors for Drug of Abuse Detection. *Advanced Sensor Technology*, 125.
- 3) Dinis-Oliveira, R. (2019). Metabolism and Metabolomics of Opiates: A Long Way of Forensic Implications to Unravel. *Journal of Forensic and Legal Medicine*, 61, 128.
- 4) Gasche, Y., Daali, Y., Fathi, M., et al. (2004). Codeine Intoxication Associated with Ultrarapid CYP2D6 Metabolism. *The New England Journal of Medicine*, 351, 2827.
- 5) Goldansaz, S. A., Guo, A. C., Sajed, T., et al. (2017). Livestock Metabolomics and the Livestock Metabolome: A Systematic Review. *Plos One*, 12(5), e0177675.
- 6) Hauschild, A., Schneider, T., Pauling, J., et al. (2012). Computational Methods for Metabolomic Data Analysis of Ion Mobility Spectrometry Data—Reviewing the State of the Art. *Metabolites*, 2(4), 733.
- 7) Farquharson, S., Brouillette, C., Smith, W., et al. (2019). A Surface-Enhanced Raman Spectral Library of Important Drugs Associated With Point-of-Care and Field Applications. *Frontiers in Chemistry*, 7, 706.
- 8) Jang, S., Son, S. U., Kang, B., et al. (2022). Electrospun Nanofibrous Membrane-Based Colorimetric Device for Rapid and Simple Screening of Amphetamine-Type Stimulants in Drinks. *Analytical Chemistry*, 94, 3535.
- 9) Jorge, T., Mata, A., & António, C. (2016). Mass Spectrometry as a Quantitative Tool in Plant Metabolomics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2079), 20150370.
- 10) Liu, K., Shang, C., Wang, Z., et al. (2018). Non-Contact Identification and Differentiation of Illicit Drugs using Fluorescent Films. *Nature Communications*, 9, 1695.
- 11) Koh, E. H., Lee, W., Choi, Y., et al. (2021). A Wearable Surface-Enhanced Raman Scattering Sensor for Label-Free Molecular Detection. *ACS Applied Materials and Interfaces*, 13(2), 3024.
- 12) Lohumi, S., Kim, M. S., Qin, J., et al. (2017). Raman imaging from microscopy to macroscopy: Quality and safety control of biological materials. *Trends in Analytical Chemistry*, 93, 183.

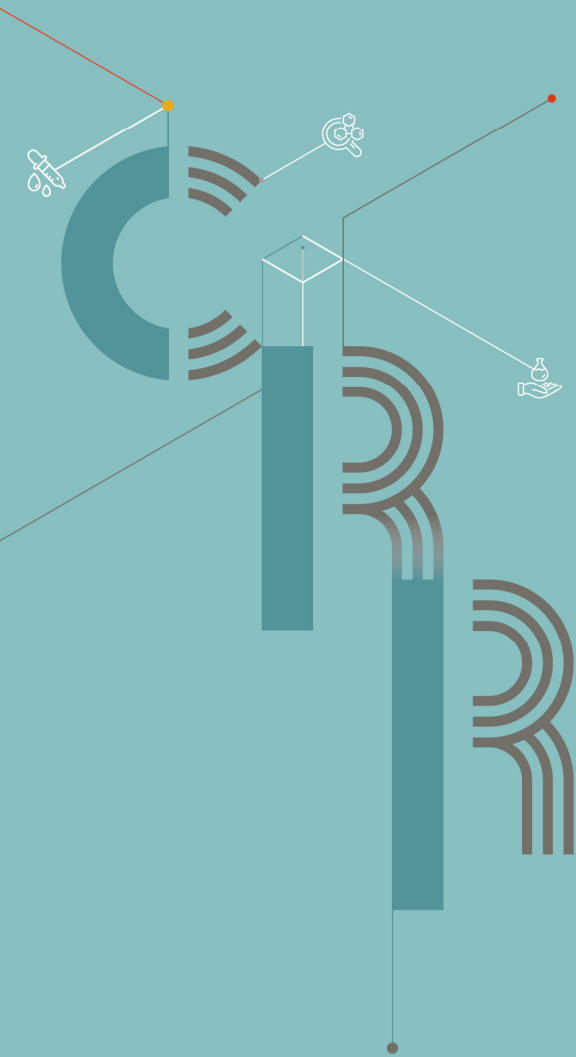
- 13) Mao, K., Yang, Z., Li, J., et al. (2017). A Novel Colorimetric Biosensor based on Non-Aggregated Au@Ag core-shell Nanoparticles for Methamphetamine and Cocaine Detection. *Talanta*, 175, 338.
- 14) Miočević, O., Cole, C. R., Laughlin, M. J., et al. (2017). Quantitative Lateral Flow Assays for Salivary Biomarker Assessment: A Review. *Frontiers in Public Health*, 5, 133.
- 15) Shamil, E., Cook, P., D'Souza, A., et al. (2021). Patient Perceptions of the Use of Cocaine in Sinonasal Surgery. *Clinical Otolaryngology*, 46(6), 1300.
- 16) Xu, S., Ma, B., Li, J., et al. (2023). Europium Nanoparticles-Based Fluorescence Immunochemical Detection of Three Abused Drugs in Hair. *Toxics*, 11(5), 417.
- 17) Son, S. U., Jang, S., Kang, B., et al. (2021). Colorimetric Paper Sensor for Visual Detection of Date-Rape Drug  $\gamma$ -hydroxybutyric acid (GHB). *Sensors and Actuators B: Chemical*, 347, 130598.
- 18) Teymourian, H., Parrilla, M., Sempionatto, J. R., et al. (2020). Wearable Electrochemical Sensors for the Monitoring and Screening of Drugs. *ACS Sensors*, 5, 2679.
- 19) Verrinder, E., Wester, N., Leppänen, E., et al. (2021). Electrochemical Detection of Morphine in Untreated Human Capillary Whole Blood. *ACS Omega*, 6, 11563.
- 20) Yin, Q., Yang, X., Yang, L., et al. (2023). Cu, I-Doped Carbon Dots as Simulated Nanozymes for the Colorimetric Detection of Morphine in Biological Samples. *Analytical Biochemistry*, 680, 115313.

### 〈기타문헌〉

- 1) Australian Associated Press & Lock. S. (2018.04.27). Victoria Considers a TOTAL Drink-Drive Ban as the State Lowers its Blood Alcohol Limit in a Crackdown on Boozy Motorists. *Daily Mail*.  
<https://www.dailymail.co.uk/news/article-5663151/Victoria-considers-TOTAL-drink-drive-ban-lowers-blood-alcohol-limit-to.html>
- 2) Bishop. (2023). Raman Spectroscopy Identifies Disease Characteristics and In Vitro Structure, *Photonics Media*,  
[https://www.photonics.com/Articles/Raman\\_Spectroscopy\\_Identifies\\_Disease/a68649](https://www.photonics.com/Articles/Raman_Spectroscopy_Identifies_Disease/a68649)
- 3) Coast Guard Academy professors help in battle against drugs, *The day, Military*,  
<https://www.theday.com/military/20190817/coast-guard-academy-professors-help-in-battle-against-drugs/>
- 4) Countrywied Testing社 홈페이지, <https://www.countrywiedtesting.com/collections/saliva-tests>
- 5) DR. Rajapandiyan Panneerselvam group 홈페이지, <https://rajapandiyan.weebly.com/research.html>
- 6) Grand View Research. (2023). Drug Of Abuse Testing Services Market Size, Share & Trends Analysis Report, By Drug Type, By Region, And Segment Forecasts, 2023 – 2030.
- 7) Hemochrom社 홈페이지, <https://www.hemochrom.com/about.php>
- 8) Mazzoni. A. (2019.10.12). Banned for Being Sober: How Roadside Drug Driving Tests are WRONG up to 16% of the Time – and Innocent Drivers Could be Losing their Licences. *Daily Mail*.

- <https://www.dailymail.co.uk/news/article-7454621/Roadside-drug-driving-tests-called-question-return-incorrect-cannabis-readings.html>
- 9) Meizheng社 홈페이지. <https://mzfoodtest.com/>
  - 10) Metrohm社 브로슈어. Application Note AN-RS-033
  - 11) Pharmcheck社 홈페이지. <https://www.pharmchek.com/resources>
  - 12) Rigaku社 홈페이지. PROGENY 1064 nm HANDHELD RAMAN ANALYZER, <https://handhelds.rigaku.com/products/progeny-pharmaceutical-analyzer-1064-nm-raman>
  - 13) SCIEX社 홈페이지, Introduction to LC-MS/MS, <https://sciex.com/kr/education/introduction-to-lc-ms-ms>
  - 14) Thermo Fisher Scientific 홈페이지. TruNarcTM, Handheld Narcotics Analyzer, <https://www.thermofisher.com/order/catalog/product/kr/ko/TRUNARC>
  - 15) Transmed社 홈페이지. <https://transmedco.com/oral-fluid-saliva-drug-tests/>
  - 16) TV 조선. (2023.06.08). 44개 하수처리장서 모두 필로폰 검출...전국에 퍼진 마약. [https://news.tvchosun.com/site/data/html\\_dir/2023/06/08/2023060890178.html](https://news.tvchosun.com/site/data/html_dir/2023/06/08/2023060890178.html)
  - 17) Vrije Universiteit Brussel 홈페이지, <https://we.vub.ac.be/en/research>
  - 18) SBS 본격 연예 한밤. (2019.4.23.). <https://allvod.sbs.co.kr/allvod/vodEndPage.do?mdald=22000331461&btn=free>
  - 19) YTN 뉴스. (2023.04.07). "남이 주는 건 먹지마세요"...학교도 '마약 음료' 비상. YTN, [https://www.ytn.co.kr/\\_ln/0103\\_202304071759204551](https://www.ytn.co.kr/_ln/0103_202304071759204551)
  - 20) 김연진, 권아현. (2023.04.14). [대한민국 10대들의 마약리포트2] 충격적인 그들의 마약암수율. 주간조선, 2754호, <https://weekly.chosun.com/news/articleView.html?idxno=25742>
  - 21) 김상훈. (2022.10.14). '마약과의 전쟁'·4대 검찰청에 특수팀 설치. MBC 뉴스, [https://imnews.imbc.com/replay/2022/nw1200/article/6416990\\_35715.html](https://imnews.imbc.com/replay/2022/nw1200/article/6416990_35715.html)
  - 22) 김영옥. (2023.06.09). 전국 하수처리장, 마약 안 나온 곳이 없다...“매일 1000명 중 1명꼴로 필로폰”. 중앙일보, <https://www.joongang.co.kr/article/25168564#home>
  - 23) 김종용. (2022.07.19). '경고등 켜진' 청소년 마약사범, 울 들어서만 527명... 이미 작년 수치 넘어서. 조선일보, [https://biz.chosun.com/topics/law\\_firm/2022/07/19/ALHZSJ7RM5HQNEOZQZNI717FGY/](https://biz.chosun.com/topics/law_firm/2022/07/19/ALHZSJ7RM5HQNEOZQZNI717FGY/)
  - 24) 박원희. (2022.08.11). 관세청 "마약 대리 운반시 처벌됩니다"...마약 근절 캠페인. 연합뉴스, <https://www.yna.co.kr/view/AKR20220811103000002>
  - 25) 박혜원. (2023.03.22). “보호관찰 출석 날, 마약하는 날” 보호관찰 마약사범 잇따라. 헤럴드경제, <https://biz.heraldcorp.com/view.php?ud=20230322000218>.
  - 26) 박홍주. (2023.04.16). 우리 엄마도 놀랐는데...‘풍당 마약’ 진단키트 9월 상용화 가닥. 매일경제, <https://www.mk.co.kr/news/society/10713561>.
  - 27) 송영두(2023.05.24). 마약검사 의무화에 진단기업 추가 ‘연일 상승’...들여다보니. 팜이데일리, <https://pharm.edaily.co.kr/news/read?newsId=03158646635612200&mediaCodeNo=257>

- 28) 왕태석. (2023.02.07). 딸기 스티커, 팬케이크 시럽이....진화하는 마약 은닉 수법. 한국일보, <https://m.hankookilbo.com/News/Read/A2023020610180003197>
- 29) 이노릭슨社 블로그. <https://blog.naver.com/mrhell/222798503310>
- 30) 장호종. (2023.02.15). 윤석열 정부 마약과의 전쟁, 성공할 수 있을까?. 노동자 연대. <https://wspaper.org/article/29014>



# 융합연구리뷰

Convergence Research Review

# 02

---

## 합성 및 변조 음성 탐지 기술 동향

---

정수환(송실대학교 전자정보공학부 교수)

## 02

정수환(송실대학교)

# 합성 및 변조 음성 탐지 기술 동향

## I. 서론

딥러닝의 등장은 근 수년에서 십 수년 동안 많은 연구 분야의 변명을 이끌었다. 특히나 음성 영역에서는, 최신 인공지능 모델들이 놀라운 성과를 달성하며 실제 발화자(보나 파이드<sup>1)</sup>, Bona fide)의 음성과 가까운 고품질의 음성을 합성할 수 있게 되었다. 이러한 딥러닝의 성공은 현대 문명의 각 분야에서 어려운 문제들을 해결하는 데 도움을 주었지만, 그 파생 기술 중 하나인 딥페이크는 우리 사회와 경제에 심각한 위협이 되고 있다. 국제적인 보안기업 어베스트(Avast)에 따르면, 딥페이크에 따른 위협은 네 가지로 분류가 가능하다: (1) 첫째는 사회적 위협으로, 악의적인 집단이 사회적인 불안을 야기하려 시도할 수 있다. (2) 둘째는 법적 위협으로, 전자 문서와 같은 디지털 파일을 조작하려 할 수 있다. (3) 셋째는 개인적 위협으로, 딥페이크를 통해 온라인으로 개인을 모욕하거나 괴롭힐 수 있으며 (4) 넷째로 온라인 사기나 협박 등에 악용될 수 있다.

이미지나 비디오에서의 딥페이크와는 다르게, 음성 영역에서의 딥페이크는 지난 십 여년 간 사회에서 그리 큰 이목을 끌지 못했다. 그러나 지난 수 년 동안 음성 관련 기술이 매우 정교하게 발전하며, 생체 인식과 같은 음성 관련 응용 프로그램에 대한 보안 우려가 증가하기 시작했다. 음성 합성 기술은 자동 음성 검증 시스템(이하 ASV, Automatic Speech Verification)에 대하여 가짜 신원을 생성하는데 사용될 수 있으며, 이는 기업에 엄청난 경제적 손실을 야기하거나 대중에 대한 신뢰를 약화하는 등 비즈니스의 서로 다른 영역에서 부정적인 영향을 행사하는데 악용될 수 있다. 이러한 우려는 더 이상 숨은 위협이 아니라, 음성을 사용할 수 있는 곳이라면 어디든지 -스마트폰부터 스마트홈에서 사용되는 음성 통제 시스템까지 전 영역에 이르러- 나타날 수 있는 가시적인 위협이 되었다.

딥러닝 영역에서는 이미지, 오디오 및 비디오를 포함한 합성된 미디어를 통칭 “딥페이크”라 한다. 음성 딥페이크는 더 이상 ASV와 같은 시스템에 대한 위협만이 아니다. 딥페이크는 가시적이고 즉각적으로

1) 라틴어에서 보나파이드 (Bona fide)는 “진실된, 진짜의”라는 의미를 가지며, 특히 음성 영역에서 사용될 때에는 “실제 음성”, “실제 발화자에 의해 생성된 오디오 데이터” 등을 가리키기 위해 사용된다. 연구에서는 “bona fide data”, “bona fide trials” 등으로 서술된다.



사람들의 인식과 감정에 영향을 끼쳐, 이제는 개인에서 기업까지 더 넓은 범위에 이르러 사회를 잠식하고 조종할 수 있다. 예를 들어, 심층신경망을 통해 생성되는 음성 샘플들은 사회공학적인 공격이나 잘못된 정보의 확산, 비즈니스에서 사기를 유발하거나 명예를 훼손시키는 등의 범죄를 수행하는데 악용될 수 있다. 사이버 공간의 공격자들이 이제는 프라이버시, 금융, 유저들의 명예 등과 같은 민감한 주제로 눈을 돌리는 만큼, 이러한 위협에 대항할 수 있는 보호 수단의 구축이 필요하다.

근 수년간 합성 음성의 생성과 탐지는 그 사이에서 경쟁적으로 발전해왔다. 공격자의 관점에서, 가짜 음성을 생성하는 데에는 몇 가지 기술을 꼽을 수 있겠지만, 탐지 기술자에 의해 구분되기 힘들만큼 정교한 음성 샘플을 생성하기 위해서는 신경망의 사용이 절대적이다.

방어자의 관점에서는, ASV를 보호하기 위한 방법을 두 가지로 나눌 수 있다.

- (1) 공격자에 대한 화자 인증의 성능을 향상시킨다. 이때 공격자는 인증 시스템에 대해 고급 회피 기술을 사용하지 않고 자신의 목소리 등 생체적 특징을 화자 인증 시스템에 제출할 수 있다.
- (2) 특정한 상황에 적용되거나 일반적으로 통용될 수 있는 스푸핑(spoofing, 공격자가 네트워크, 웹사이트 등의 데이터 위·변조를 통해 정상 시스템인 것처럼 위장하여 일반 사용자를 속이는 해킹 기법)의 방어책인 카운터메저(countermeasure)를 설계한다.

두 방법 모두가 동등하게 중요하지만, 그 중에서도 카운터메저는 기존 ASV 시스템에 통합될 수 있으며, 이는 기관이 유지 보수를 위한 비용을 줄일 수 있게 한다. 여기서의 카운터메저는 일반적으로 스푸핑 탐지를 담당하는 기계학습 모델로, 견고한 카운터메저를 구축하는 것은 2013년 이후 지속적으로 관심을 받는 활발한 연구 분야가 되어 왔다. 특히나 최근 연구에서는 카운터메저를 ASV에 통합될 수 있는 간이 모델로 제공할 뿐만 아니라 ASV와 동등하게 배치하여 함께 최적화하는 등의 관점을 취하고 있다.

2013년부터 2023년까지 위에서 서술한 새로운 위협들에 대응하기 위해 견고한 카운터메저의 구축을 목적으로 하는 여러 대회가 실시되었다. 특히 ASVspoof2021에서 주최측은 ASV가 존재하지 않는 상황에서 스푸핑 된 음성을 탐지하는 것에 초점을 맞춘 새로운 도전 과제인 딥페이크 트랙(이하 DF, Deepfake Track)을 소개했다. 해당 트랙에서는 공격자가 피해자의 음성 데이터를 소셜 미디어 등에서 얻어낼 수 있는 상황을 가정한다. 공격자는 공개된 음성 데이터와 음성 딥페이크 기술을 사용하여 피해자의 목소리와 유사한 스푸핑 음성을 생성하고, 이를 소셜 미디어 및 콜센터, 기타 응용프로그램 등에 게시할 수 있다(Yi et al, 2023). 결과적으로 이는 피해자의 명예를 훼손하고 잘못된 정보를 퍼뜨리며, 이를 통해 피해자를 협박하는 등 심각한 결과를 초래할 수 있다. 따라서 스푸핑 된 음성을 탐지하는 것에 더욱 각별한 주의가 필요한 실정이며, 이를 생성하는 딥페이크 기술 및 딥페이크 탐지를 우회하는 관련 기술을 함께 고려하여야 한다.

본 원고에서는 음성 분야에서의 스푸핑 탐지와 관련된 연구를 요약 및 정리한다. 특히나 딥페이크 합성 음성을 판별하는 상황에서 독립적으로 카운터메저를 사용하는 경우, ASV 시스템이 존재하는 상황에서

카운터메저가 ASV 시스템을 함께 사용하는 탠덤(Tandem) 방어를 수행하는 경우에 초점을 둔다. 현존하는 방어 기술의 한계와 개선 방안을 논의하기 위해 앞서 서술한 카운터메저 시스템에 대한 개별적이고 종합적인 이해가 필요하다. 따라서 2020년부터 2023년까지의 진전된 오디오 딥페이크 카운터메저의 최신 연구들을 요약 및 논의하며, 이는 음성 영역에서 주목받는 주요 학회(ICASSP, Interspeech)에서 논의된 연구를 포함한다. 나아가 본 원고에서는 개별적인 카운터메저를 대상으로 한 적대적 공격과 부분 조작 오디오를 포함한 오디오 딥페이크의 최신 기술에 대해서도 논한다. 딥페이크 음성에 대한 방어 영역에서의 도전 과제와 개선 방안을 논의함으로써 이 문제에 대한 사회의 경각심을 고취시키고, 각 기관에서의 연구 참여와 촉진에 기여하기를 바란다.

## II. 현존하는 음성 합성 기술

### 1. Text-to-Speech

TTS(Text-to-Speech) 기술은 그 시작 이후 놀라운 발전을 이룩했으며, 이는 우리가 기계와 상호작용하는 방법을 변화시키고, 인간과 컴퓨터 사이의 틈을 메울 수 있게 했다. TTS 기술은 언어학, 컴퓨터 과학, 그리고 인공지능의 교차점에 위치해 있다. 이는 기계가 작성된 텍스트를 기반으로 사람의 음성 패턴과 억양을 모사, 음성 언어로 변환할 수 있게 한다. TTS 기술은 시간을 거듭하며 다양한 분야에서 응용되며 크게 발전해왔다. 초기에 TTS 시스템은 음소(언어에서 가장 작은 소리의 단위)를 연결해 음성을 생성하는 단순한 규칙 기반 접근법에 의존했다. 이때에는 TTS가 생성된 음성이 다소 로봇이 말하는 것처럼 들린다는가, 자연스럽게 않다는가 하는 한계를 가지고 있었다. 최근에는 심층 신경망을 사용한 음성 합성의 도입으로 TTS 분야에서 중요한 진전이 이루어졌고, 이는 더욱 이해 가능한 (사람에게 인지 가능한) 결과를 생성해 내도록 했다. 이러한 TTS의 일반적인 메커니즘은 언어학적인 조합이나 음향 모델링을 통해 사람과 비슷하게 들리는 음성을 생성하는 것인데, 이 과정은 다음의 몇 가지 단계로 요약될 수 있다.

먼저 텍스트 분석이다(Tan et al., 2021). 이때 입력 텍스트는 음소, 스트레스 패턴 및 음성의 운율학적 특성(억양, 리듬 및 타이밍 등)과 같은 언어적 특징을 식별하기 위해 분석된다. 다음으로 언어 처리이다. 이 단계에서는 앞서 분석한 텍스트에 언어적 규칙 및 모델이 적용되어 각 단어의 발음과 스트레스를 포착하는 음성표현으로 변환된다. 이후에는 음향 모델링을 거쳐 위에서 변환된 음성 표현을 음향적 특징에 매핑(mapping)하고, 음성 파형을 생성한다. 그 후 실제로 음성을 합성하게 되는데, 위에서 생성된 음향적 특징 및 음성 파형이 연속 합성, 통계적 매개 변수 합성(히든 마코프 모델 등) 및 신경망과 같은 다양한 기술을 통해 실제 음성으로 변환된다. 여기에 마지막으로 프로소디(prosody, 운율 체계)를 생성한다. 이 단계에서 음성에 음의 높낮이, 리듬 및 강조 등이 추가되어 음성이 더 자연스럽게 들린다. 최종적으로 합성된 음성은 스피커나 오디오 재생 장치를 통해 사용자에게 전달된다.

기계 학습, 특히나 심층 학습과 신경망의 발전은 TTS가 “자연스럽고 표현력 있는” 음성을 생성하도록 도왔으며, 이로 인해 청취자가 기계적으로 합성된 음성과 사람이 발화한 음성을 구별하기 어려워졌다. 최근 몇 년 동안, 딥러닝 기술, 특히나 신경망은 TTS 기술 발전에 중추적인 역할을 해왔다. LSTM(Long Short-Term Memory), CNN(Convolutional Neural Networks), Transformer 기반 구조와 같은 딥러닝 모델들은 앞서 언급한 “자연스럽고 표현력 있는” 음성을 생성하는데 놀라운 성능을 보였다. 여기에 구글(Google)의 인공지능 회사 딥 마인드(DeepMind)에 의한 WaveNet 모델의 도입은 TTS 시스템의 발전에 중요한 돌파구를 제시했다. WaveNet은 심층 생성형 모델을 사용, 파형을 직접 합성함으로써 고품질의 음성 샘플을 생성한다. 그 후에 등장한 Tacotron 및 Tacotron2와 같은 변형 모델들은 텍스트로부터 스펙트로그램 프레임(Spectrogram Frame) 및 파형 샘플을 직접 예측해 합성 과정을 크게 개선했다. 현재는 GPT와 같은 Transformer 기반 인공지능 모델도 TTS 연구에 도입되어 모델을 입력 텍스트에 적응 및 조절되도록 하여 종단간 음성 합성(end-to-end)이 가능해졌다.

TTS 기술의 광범위한 적용은 윤리적, 사회적으로 고려할 수 있는 문제들을 동시에 제기하였는데, 이 중 하나는 딥페이크 오디오 생성의 잠재적인 위험이다. 딥페이크 오디오를 생성하는 사람들 중, 악의적인 일부는 TTS를 사용해 특정 개인을 흉내 내거나 오해를 불러일으키는 내용을 생성할 수 있다. 이는 TTS의 책임 있는 사용에 대한 지속적인 논의와 합성 음성을 식별하는 탐지 메커니즘의 개발로 이어졌다. 또한 TTS 기술의 지속적인 개선에 따라 미디어 및 엔터테인먼트 분야 등에서 사람과 기계가 생성한 콘텐츠 간의 경계에 대한 질의가 제기되었으며, 결론적으로는 이로 인해 콘텐츠 저작권 및 무결성 등에 대한 논의가 촉발되었다.

TTS 기술은 각각 분석된 텍스트를 활용하는 방법 및 실제 음성 파형을 합성하는 방법에 따라 아래와 같이 음성 합성 기술을 분류한다.

## 1.1 언어학적 특성을 이용한 어쿠스틱 모델 및 보코더

이 방법은 TTS 구조에서 가장 전통적인 접근법으로, 텍스트 입력으로 프로세스를 시작한다. 여기서의 텍스트 분석은 언어학적 특성, 예를 들면 음소나 프로소디, 타이밍 등을 추출하는 데 사용된다. 이러한 언어학적 특성은 어쿠스틱 모델로 전달되고 이는 음성이 어떻게 들릴지 나타내는 음향 특성으로 변환된다. 마지막으로 보코더(Vocoder)는 이러한 음향 특성을 가져와 원시 오디오 파형을 생성해낸다. 이 방법은 음성 합성 과정을 세밀하게 제어할 수 있고, 언어학적 특성 및 음향 특성을 독립적으로 조작할 수 있다.

## 1.2 음소 기반 특성을 이용한 어쿠스틱 모델 및 보코더

이 접근법은 텍스트를 언어에서 가장 작은 소리의 단위인 음소로 변환하는 것으로 합성 과정을 시작한다. 또한 음소의 순서는 이후 음향 특성을 유도하는 데 사용된다. 이러한 음향 특성은 일반적으로 시간에 따른 음성 신호의 주파수 내용을 포착하는 멜-스펙트로그램(Mel-Spectrogram)의 형태이다. 최종적으로 보코더는 이 스펙트로그램을 원시 파형으로 변환한다. 이 방법은 발음과 음성학적 세부사항을 정확하게 제어할 수 있도록 음소 정보를 활용할 수 있다.

### 1.3 언어학적 특성을 이용한 종단간 TTS

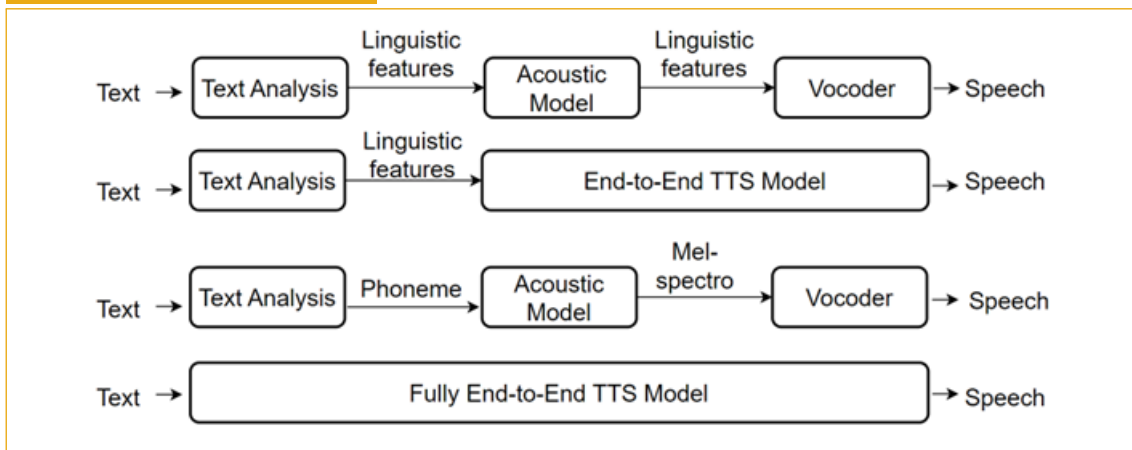
이 접근법은 다른 방법과 같이 입력 텍스트를 처리하여 언어학적 특성을 추출하는 단계를 포함한다. 그러나 앞선 두 방법과는 대조적으로, 어쿠스틱 모델과 보코더가 종단간 모델이라고 하는 단일 시스템으로 통합된다. 이 모델은 언어학적 특성에서 직접 파형을 생성함으로써 음성 합성 과정을 간소화하며 별도의 음향 모델링 및 보코딩 단계를 축소한다.

### 1.4 완전 종단간 TTS

완전 종단간 TTS 모델은 네 접근법 중 가장 간소화된 접근법이다. 이 모델은 텍스트 입력을 직접 받아와 명시적인 언어학적 특성 추출, 음소 변환 및 음향 모델링 없이 파형을 출력한다. 이 접근법은 TTS의 모든 과정을 단일 신경망 구조로 결합시켜 텍스트 입력에서 음성 출력까지의 복잡한 구조적 매핑을 직접 학습할 수 있도록 한다. 완전 종단간 모델은 모델의 단순함과 자연스러운 출력을 생성하는 고성능 학습법으로 알려져 있으며, 이 경우 엔지니어의 수동적인 작업을 줄일 수 있다.

그러나 위에서 서술한 TTS 기술은 각각 그 장점과 트레이드오프(trade-off, 상충 관계)가 있으며, 제어 수준, 컴퓨팅 자원 및 지향하는 합성 음성 품질 등 기업의 요구사항에 따라 선택이 달라질 수 있다.

그림 1. 네 종류의 TTS 시스템 구조



\* 출처: 저자 작성

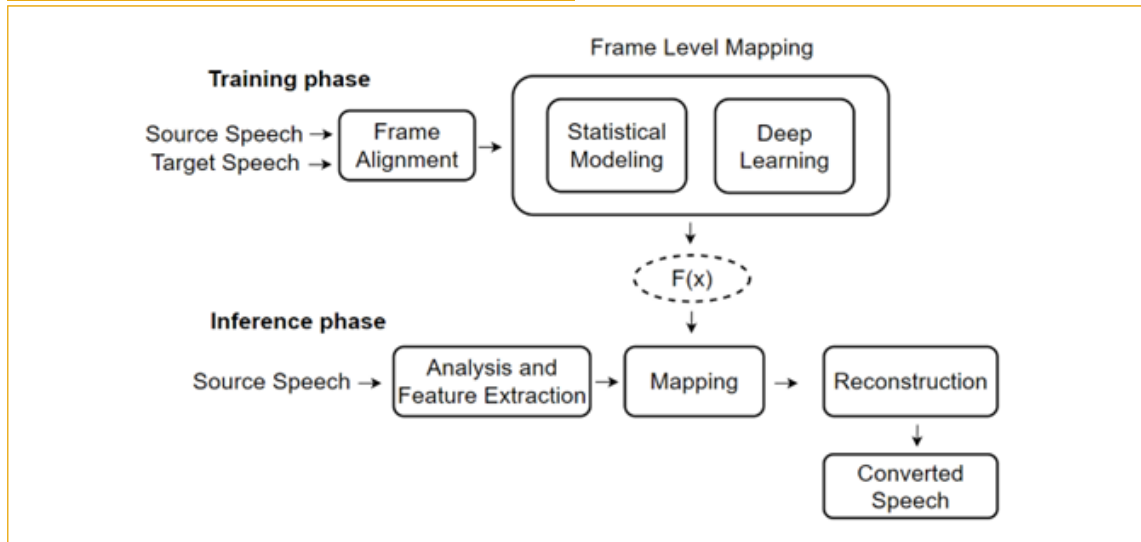
## 2. 보이스 컨버전

보이스 컨버전(Voice Conversion)의 개념은 음성의 지각적 측면을 조작하면서 -단순히 말해 음성의 화자가 누구인지를 조작하면서- 내용의 요해도<sup>2)</sup>를 유지하고자 하는 요구사항을 위해 등장했다. 이 기술은 등장 이래 크게 발전해왔으며, 단순한 음성 고도 변환부터 정교한 목소리 변환을 가능하게 하는 등 다양한

쓰임새로 진화했다. 보이스 컨버전은 엔터테인먼트, AI 음성 비서 서비스 및 보안 응용 애플리케이션 등의 다양한 분야에서 핵심적인 기술로 사용되고 있다. 이러한 보이스 컨버전의 노른자위는 원래 발화자의 목소리 -흔히 소스 음성이라고 한다-의 특정한 특성을 변경해 대상 발화자가 말한 것처럼 들리게 하는 것이다. 소스와 대상 발화자는 목소리의 정체성을 갖는 누구든 될 수 있으며, 여기서 최종 목표는 음성의 언어학적 내용이나 요해도에 영향을 주지 않고 이 변환을 달성하는 것이다.

보이스 컨버전은 특히나 몇몇 분야에서 유용한데, 예를 들어 개인화된 사용자 인터페이스, 즉 음성 어시스턴트나 챗봇의 경우 사용자가 선호하는 발화 특성과 일치하도록 응답하게 하기 위해 보이스 컨버전을 사용할 수 있다. 또한 영화 및 엔터테인먼트 사업에서 더빙 및 음성 연기를 위해 배우나 캐릭터의 목소리를 설득력 있게 제공하는 데 사용될 수 있다. 반면, 보안 응용프로그램에서는 공격자의 목소리를 대상 발화자의 목소리로 변환하여 음성 스푸핑 공격을 방지하기 위해 사용되기도 한다.

그림 2. 보이스 컨버전의 일반적인 워크플로우(workflow)



\* 출처: 저자 작성

## 2.1 프레임 수준 매핑

프레임 수준 매핑은 보이스 컨버전 분야의 핵심적인 방법론 중 하나로 꼽힌다. 이 방식에서는 학습 단계에서 'F'라는 특정 함수를 통해 소스와 대상 음성이 프레임 수준에서 정렬된다. 이러한 정렬 과정은 각 음성의 프레임에서 소스와 대상 화자의 특성 차이를 세밀하게 파악하기 위해 필수적이다. 매핑 함수 'F'는 소스와 대상 음성의 각 프레임 간 관계를 학습하며, 이를 통해 각각의 프레임에 요구되는 변환을 정확하게 포착한다. 해당 과정은 아래의 순서로 진행된다.

2) 요해도는 음성 영역에서 사용되는 개념으로, 발화하는 화자의 이야기를 청자가 이해하는 정도를 의미한다.

□ 훈련 단계

- 소스와 대상 음성을 포함하는 데이터 세트를 수집한다. 이 소스 및 대상 음성 쌍은 매핑 함수를 학습하는데 필수적이다.
- 소스와 대상 음성 모두에서 음향 특징을 추출한다.
- 매핑 함수 'F'를 사용하여 소스와 대상 음성의 프레임을 정렬한다.
- 매핑 함수 'F'를 훈련시켜 소스 음성의 음향 특징을 대상 음성의 음향 특징과 연관시킨다.

□ 추론 단계

- 변환할 새로운 소스 음성에서 음향 특징을 추출한다.
- 훈련된 매핑 함수 'F'를 사용해 소스 특징을 해당 대상 특징에 매핑한다.
- 매핑된 특징으로부터 음성을 재구성해 대상 화자의 특성과 일치하는 음성을 생성한다.

프레임 수준 매핑은 상대적으로 간단한 특징 매핑을 통해 보이스 컨버전을 달성할 수 있으나, 발화의 스타일이나 감정 표현의 변화 등 복잡한 변환을 처리하는 데 다소 어려움이 있다.

## 2.2 콘텐츠-음성 인코더 기반 변환

콘텐츠-음성 인코더 기반 변환은 일반적으로 신경망 기반의 접근법과 연관되어 설명될 수 있으며, 앞선 방식보다 세련된 방법을 취한다. 이 기술은 음성을 변환하기 위해 디코더 외에도 두 가지 핵심 구성 요소인 콘텐츠 인코더와 화자 인코더를 활용한다. 이를 사용한 과정은 다음과 같다:

□ 훈련 단계:

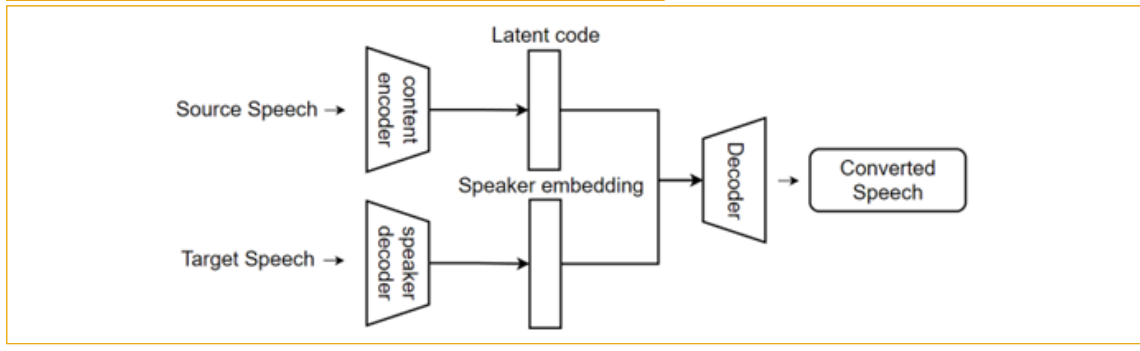
- 여기에서도 앞선 프레임 수준 매핑 기술과 유사하게, 소스와 대상 음성을 포함하는 데이터 세트를 요구한다.
- 콘텐츠 인코더는 언어적인 특징을 포함하는 소스 콘텐츠를 처리하여 언어의 내용을 나타내는 잠재 코드(latent code)를 생성한다.
- 화자 인코더는 대상 음성에 대한 특성을 캡처하여 임베딩(embedding)을 생성한다.
- 위 두 단계에서 생성된 잠재 표현과 화자 임베딩은 디코더를 통과한다.
- 디코더는 대상 화자의 발성 특성과 언어 내용을 결합한 변환된 오디오를 생성한다.

□ 추론 단계

- 새로운 소스 음성의 변환을 위해, 음성이 콘텐츠 인코더를 통과하여 잠재 코드를 생성한다.
- 화자 인코더는 대상 화자의 발성 특성을 캡처한다.
- 1단계에서 생성된 잠재 코드와 화자 임베딩이 디코더에 공급된다.
- 디코더는 언어 내용을 유지하면서 대상 화자에 의해 발화된 것처럼 들리는 변환 음성을 생성한다.

이 기술은 프레임 수준 매핑보다 더 복잡한 변환을 처리할 수 있어, 변환되는 음성이 단순히 화자의 정체성뿐만 아니라 발화 스타일이나 감정 표현 및 감정 변화 등을 포함하는 경우에 더욱 적합하다.

그림 3. 보이스 컨버전을 위한 오토인코더(Autoencoder) 네트워크



\* 출처: 저자 작성

### 3. 데이터셋(Data Set)

합성 음성 및 스푸핑 음성 -이하 딥페이크 오디오라고 하겠다- 탐지에 관한 연구를 지원하기 위해 대학 및 기관을 포함한 여러 기업이 다양한 데이터 세트를 적극적으로 수집 및 엄선하였으며, 이를 연구와 관련된 커뮤니티와 공유하였다. 아래는 이러한 데이터셋과 응용 분야를 종합적으로 정리한 도표이다.

표 1. 실제 및 합성 음성을 제공하는 데이터셋

Name	Generation Technology	Description	Source
VCTK	Recording	110 <i>English speakers with various accents.</i>	The voice bank corpus
VoxCeleb2	Recording	1 million utterances for 6,112 celebrities from YouTube videos.	Deep Speaker Recognition
LJSpeech	Recording	13,100 short audio clips of a single speaker reading passages from 7 non-fiction books.	LJ Speech dataset
FoR	TTS and recording	Real speech and 87,000 synthetic utterances from a total of 33 synthesized voices.	Real-or-Fake (FoR)
M-AILABS	Recording	A thousand hours of audio and the text-files in prepared format.	M-AILABS Speech Dataset
WakeFake	GAN based TTS models	196 hours of generated audio files.	WakeFake dataset
PartialSpoof	VAD, segment replacement	Derived from ASVspoof 2019 bona fide set.	PartialSpoof dataset

Name	Generation Technology	Description	Source
In the Wild	Probably TTS	20.8 hours of bona-fide and 17.2 hours of spoofed audio.	In the Wild dataset
FakeAVCeleb	Voice Cloning	Derived from 500 VoxCeleb2 videos.	FakeAVCeleb Dash-Lab
Attack Agnostic	TTS, VC	31,083 real and 222,035 fake samples covering 27 attack methods.	Piotr Kawa's repo

\* 출처: 저자 작성

표 2. 오디오 딥페이크 탐지를 위한 대표적인 데이터셋의 특성

	ASVspoof 2021 DF	ADD 2022		ADD 2023				In-the- Wild	Wave- Fake	FoR
		LF	PF	FG-D	FG-D	LR	AR			
Year	2021	2022	2022	2022	2023	2023	2023	2022	2021	2019
Language	English	Chinese	Chinese	Chinese	Chinese	Chinese	Chinese	English	English	English
Goal	Detection	Detection	Detection	Game fake	Game fake	Forensics	Attribution	Detection	Detection	Detection
Fake Types	VC, TTS	TTS, VC	Partially fake	TTS, VC	TTS, VC	Partially fake	TTS, VC	TTS	TTS	TTS
Condition	Clean, Noisy	Noisy	Clean	Clean, Noisy	Clean, Noisy	Noisy	Clean	Noisy	Clean	Clean
#Hours	325.8	222	201.8	396	394.7	131.2	194.5	38	196	150.3

\* 출처: 저자 작성

\*\* 약어 정리

- DF: 딥페이크 트랙
- LF(Low-quality Fake Audio Detection): 저품질 페이크 오디오 탐지
- PF(Partially Fake Audio Detection): 부분 조작 오디오 탐지
- FG-D(Fake Game Detection): 페이크 게임 탐지
- AR(Algorithm Detection): 알고리즘 탐지

## 4. 관련 대회

ASVspoof(Automatic Speaker Verification and Spoofing Countermeasure)의 기원은 2015년 열린 Interspeech의 특별 세션으로, 스푸핑 탐지에 있어서는 첫 시도라고 할 수 있다(Wu et al., 2015). 최근에는 ASV 없이 카운터메저만 존재하는 시나리오와 노이즈·채널 효과 또는 품질이 낮은 오디오 등 음성 파일이 실행되는 실제적인 환경을 고려하는 등 일반적인 오디오 딥페이크 탐지에 초점을 두고 있다.

SASV (Spoofing-aware Audio Spoofing Verification)은 ISCA Interspeech 2022의 하위 특별 세션으로 처음 개최되었다(Jung et al., 2022a, Jung et al., 2022b). 이는 ASV와 카운터메저의 융합에



초점을 두어 ASVspoof를 보완하고자 하였다.

ADD(Audio Deepfake Detection)은 ICASSP(International Conference on Acoustics, Speech and Signal Processing, 국제 음향 음성 신호처리 학회)의 그랜드 챌린지로, 심층 합성 및 조작된 오디오, 이를 탐지하기 위한 연구를 촉진하기 위한 시도이다(Yi et al., 2023). 여기서는 특히 부분 조작 오디오, 저품질의 발화, 소음 환경 및 음성 합성 알고리즘 분류와 같은 현실적인 시나리오를 고려한다.

그림 4. 합성 음성을 탐지하기 위한 각종 대회

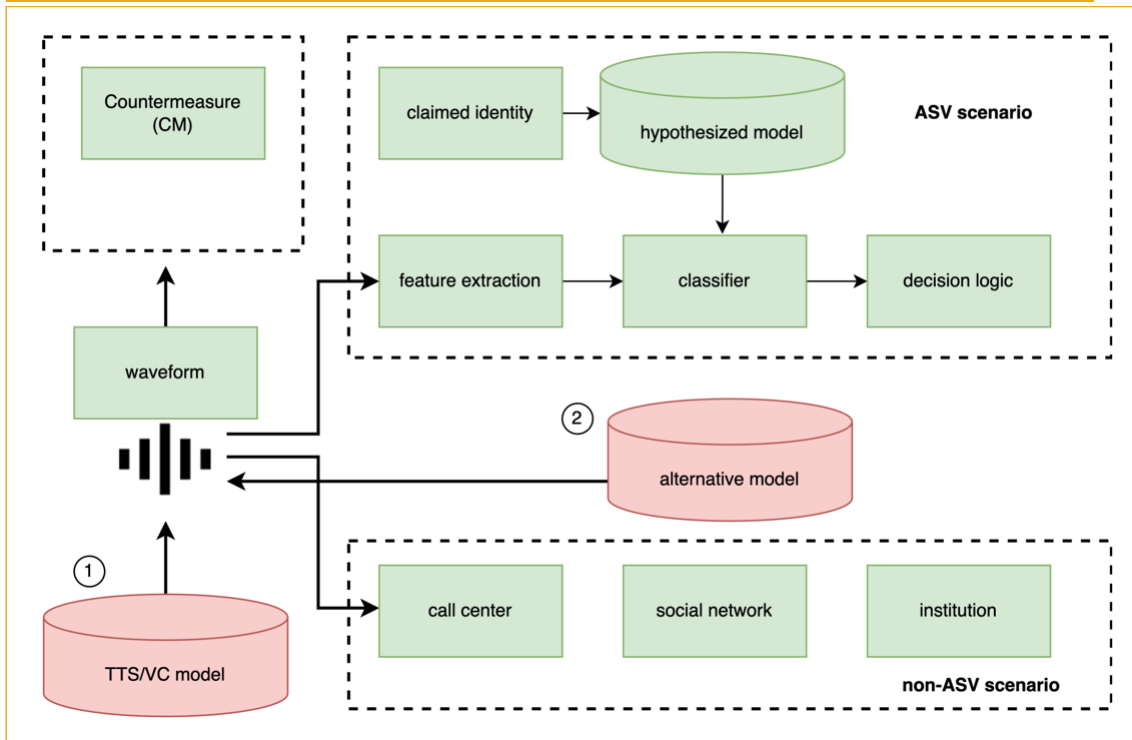


\* 출처: 저자 작성

## 5. 위협 모델

본 장에서는 ASV 시스템 유무와 관계 없이 최신 오디오 딥페이크에 관련된 공격 및 방어를 설명한다. 나아가, 보이스 컨버전 및 TTS 기술로 생성된 합성 음성을 고려하고 ASV가 존재하거나 존재하지 않는 상황에서 발생할 수 있는 현대의 공격들을 논의한다. 이러한 공격은 적대적 왜곡, 보이스 컨버전, 부분 합성 오디오 등을 채택하여 자신들의 공격을 은닉할 수 있다.

그림 5. ASV와 여러 다른 애플리케이션에 대한 스푸핑 악용 공격 시나리오. 카운터메저는 스푸핑/딥페이크 공격 방어에 사용



\* 출처: 저자 작성

## III. 현존하는 합성 음성 탐지 기술(카운터메저)

카운터메저(또는 스푸핑 방어책)는 일반적으로 '프론트 엔드'와 '백 엔드'로 구성된다. 프론트 엔드에서는 음성 파형의 표현을 추출하는 역할을 하며, 백 엔드에서는 이를 분류하는 작업, 즉 스푸핑을 실제로 탐지하는 역할을 담당한다. 다음은 카운터메저의 일반적인 구조를 서술한다.

## 1. 일반적인 카운터예제의 구조

### 1.1. 디지털 신호 처리(DSP, Digital Signal Processing) 기반 프론트 엔드

음성 신호 처리를 요하는 많은 작업들이 멜-주파수 켈스트럼 계수(이하 MFCC, Mel-Frequency Cepstrum Coefficients)를 광범위하게 사용해왔다. MFCC와 비슷한 대안책으로는 선형 주파수 켈스트럼 계수(이하 LFCC, Linear Frequency Cepstrum Coefficients)와 역 MFCC(이하 IMFCC, Inverse MFCC)가 있다. 이러한 변형들은 고유한 주파수 척도를 사용하며 입력 파형 내에서 다른 주파수 대역에 중점을 둔다.

추가로 흔히 사용되는 프론트 엔드 기법으로는 단시간 푸리에 변환(이하 STFT, Short-Time Fourier Transform) 대신 상수 Q 변환(이하 CQT, Constant Q Transform)을 활용하는 상수 Q 켈스트럼 계수(이하 CQCC, Constant Q Cepstral Coefficient)가 있다. CQCC와 LFCC는 모두 ASVspoof 챌린지에서 응용 사례가 있으며, 이는 오픈 소스로 제공된다.

켈스트럼 계수와는 대조적으로, 일부 프론트 엔드 방식은 스펙트로그램만을 추출하기로 하는데, 예를 들면 선형 필터 뱅크(이하 LFB, Linear Filter Bank) 특징은 켈스트럼 분석을 수행하지 않고 스펙트로그램을 압축하는 기법으로, 멜 스펙트로그램 추출과 유사한 기법이다.

### 1.2. 심층 신경망(DNN, Deep Neural Network) 기반 프론트 엔드

DSP 기반 프론트 엔드와 달리, DNN 기반 프론트 엔드는 매개변수를 데이터로부터 직접 학습한다. 이렇듯 ‘학습 가능한’ 프론트 엔드는 여러 이점을 제공하는데, 첫째로, DNN이 더 긴 파형에서 특성을 추출하도록 한다. 둘째로, 데이터 기반 DNN은 비선형 변환 가능성 덕에 DSP에서 발견되는 선형 연산 능력을 초월한다. 셋째로, 다양한 백 엔드와 통합하기 쉽다.

#### 1.2.1 SincNet

학습 가능한 프론트 엔드에서 가장 주목할 만한 예는 단연 SincNet이다. SincNet은 합성곱 신경망(이하 CNN, Convolutional Neural Network)의 일종으로 입력 계층 내에 학습 가능한 다수개의 윈도우드-싱크(Windowed-sinc) 필터를 갖춘 모델이다. 각 싱크 필터(차단 주파수 이상의 주파수를 모두 차단하는 신호 필터)는 조절 가능한 절단된 주파수를 갖는 시간 불변 대역 통과 필터로 작동한다. 훈련된 입력 계층은 입력 신호로부터 효과적으로 다양한 주파수 구성 요소를 추출한다. 이 방식에서 학습 가능한 싱크 필터는 전통적인 합성곱 계층과 비교해 더 적은 매개변수를 필요로 하며, 더욱 해석이 가능하다는 이점을 가진다. 어쿠스틱한 특징들을 프레임 수준에서 계산하기 위해 CNN은 지역적 시간 풀링(Local Temporal Pooling)을 통해 이 특징들을 압축한다. 추가적으로, 전역적 시간 풀링(Global Temporal Pooling) 계층은 모든 프레임에 걸쳐 특징들을 풀링함으로써 단일 벡터를 얻기 위해 사용될 수 있다.

### 1.2.2. Wav2vec

SincNet 외에도 음파를 임베딩으로 변환하는 또 다른 주목할만한 접근법에는 Wav2vec가 있다. 이는 목적에 맞게 학습하는 자기 지도 학습(이하 SSL, Self-Supervised Learning)의 성공적인 사례이다. Wav2vec은 오디오 데이터로부터 임베딩을 생성하기 위해 두 단계의 프로세스를 거친다.

사전학습 단계에서는 라벨링이 되지 않은 대규모의 음성 말뭉치(Audio Corpus)를 활용하여 신경망을 훈련시킨다. 이 단계에서 모델은 오디오 입력의 일부를 가려진 상태로 예측하는 방법을 학습하게 된다. 기본적으로 이 과정은 모델이 명시적인 주석 없이도 원시 오디오 데이터에서 유용한 특징을 학습할 수 있게 한다.

다음으로 미세 조정 단계에서는 사전 훈련 이후 모델을 특정한 하위 작업에 대해 모델을 미세하게 조정하는 작업을 수행한다. 여기서 하위 작업은 자동 음성 인식(이하 ASR, Automatic Speech Recognition), 화자 확인(이하 SV, Speaker Verification)과 같은 음성 관련 작업들을 의미할 수 있다. 예를 들어, ASR에 대한 미세 조정을 수행하면 모델은 음성으로 내뱉어진 언어를 전사하는 작업을 위해 임베딩을 더욱 정제하게 된다.

결론적으로, 미세하게 조정된 Wav2vec 모델로부터 얻은 임베딩은 입력 오디오의 특성을 풍부하게 가지며, 문맥을 고려할 수 있는 표현일 것이다. 이러한 임베딩은 다양한 하위 분야에서 활용될 수 있으며, 그 성능을 크게 향상시킬 수 있다. 이러한 SSL 접근법은 특히 파동 형식의 데이터에서는 의미 있는 정보를 더욱 효과적으로 포착할 수 있음이 증명되어, 현재 최첨단의 음성 모델들에 널리 사용되는 추세이다.

아래 <표 3>은 Wang et al.(2022) 논문으로부터 각색되었으며, 이 중 SSL-AASIST(Tak et al., 2022, Tak., 2023)는 훈련 가능한 Wav2vec를 이용해 프론트 엔드 특징을 추출한 초기 연구 중 하나이다.

표 3. 다양한 프론트 및 백엔드를 취하는 PAD 모델(현재는 Wav2vec 모델이 가장 널리 쓰임)

Front end	Back end		Database	EER(%)
	Model	Type		
Spec.	CNN	Fixed input size	LA2019	7.66
CQT+VAE	CGCRNN	Length-agnostic	LA2019	-
Spec.	SENet34	Fixed input size	LA2019	11.75
LFB + ResNet	MLP	Utt. level	LA2019	1.81
Spec., CQT	VGG	Seg. level	LA2019	10.52
Wave	SincNet + MLP	Seg. level	LA2019	20.11
Wave	RawNet2	Fixed input size	LA2019	5.64
Wav2Vec	lightDART	Fixed input size	DF2021	7.86
Wav2Vec	FeedForward (FF) + Atn.Pool	Fixed input size	DF2021	4.98
Wav2Vec	biLSTM	Fixed input size	DF2021	4.75
Wav2Vec	ViT-based + FF	Fixed input size	DF2021	3.18
Wav2Vec	AASIST	Fixed input size	DF2021	2.84
Wav2Vec	Conformer	Fixed input size	DF2021	2.58
Wav2Vec	SCL conf-3	Fixed input size	DF2021	2.17

\* 출처: Wang et al.(2022) 내용을 저자가 각색

### 1.3 DNN 기반 백 엔드: 발화 수준

백엔드 시스템의 주요 구성 요소는 프론트 엔드 시스템의 발화 수준 임베딩을 조작된 음성 및 보나 파이드 음성으로 변환하는 신경망이다.

### 1.4 DNN 기반 백 엔드: 세그먼트 수준

이 기술은 부분 조작 오디오를 탐지하기 위해 다중 시간 해상도를 구현하는 효과적인 수단으로 Partially Spoof에서 채택되었다.

SENet, LCNN과 같은 일부 CNN 기반 백 엔드는 고정 크기의 입력을 필요로 할 수 있다. 이때 입력을 잘라내거나 패딩할 수 있으나, 서로 다른 입력 형태를 가진 여러 배치(batches, 일괄 처리)를 처리하기는 어려울 수 있다. 이러한 경우에는 적응형 풀링(Adaptive Pooling)을 사용하면 합성곱 연산의 출력을 제어하고 이를 완전 연결 레이어(Fully Connected Layer)에서 사용할 수 있다.

## 2. 카운터메저의 성능 향상을 위한 주요 접근 방식

### 2.1 모델 융합(Model Fusion)

모델 융합은 서로 다른 신경망을 특정한 방식으로 결합하는 기술이다. 이 신경망은 다음과 같은 구조를 가질 수 있는데, 첫째로, 입력 수준에서 특징들을 연결하여 초기 단계에서부터 결합될 수 있다. 둘째로 중간 단계에서 하나의 모델의 출력을 다른 모델의 입력으로 사용할 수도 있으며, 이를 임베딩 융합이라 한다. 마지막으로, 모델이 결정을 내리는 시점에서 통합될 수 있는데, 이를 점수 수준 융합이라 부르며, 여기서의 “점수”는 모델이 입력 음성을 진짜 사람의 음성이라고 판단할 확률을 의미한다. 이 경우는 앙상블 기법의 모델 학습 방법과 매우 유사하다고 할 수 있다. 오디오 딥페이크 탐지의 성능을 높이기 위한 방법 중 모델 융합 방식은 CM의 성능을 향상시키는 가장 유망한 방안으로 평가되기도 한다.

ASVspoof 2019에서 한 연구는 음성 합성 및 보이스 컨버전 공격을 포함하는 논리적 접근 시나리오(이하 LA, Logical Access, 공격자가 음성 합성에 대한 인식을 가지고 이를 공격에 포함하는 경우를 가정한 대화의 시나리오 중 하나)에 대한 융합이 효과적임을 보였다(Nautsch et al., 2021). 또한 저자들은 이산 푸리에 변환(Discrete Fourier Transform)과 이산 코사인 변환(Discrete Cosine Transform)을 사용, 스펙트럼 표현을 추출해 내는 여러 시스템과 그들의 융합을 설명했다. 이 연구에서는 점수 수준의 융합이 다양한 접근법으로 설계되었는데, (1) 서버 시스템의 점수가 동일한 수준의 서버 시스템에서 나온 점수의 표준 편차에 따라 정규화되어 동등한 가중치로 통합되거나, (2) 또는 점수를 로지스틱 회귀에 결합할 수 있다. (3) 기타 방법으로는 융합 전에 단일 계층 신경망을 통해 점수를 유도할 수 있으며 (4) 점수들을 단순히 동등한 가중치를 주어 평균을 내는 방법도 있다.

ASVspoof와 같은 챌린지에서는 일반적으로 융합 시스템이 단일 시스템보다 성능이 뛰어난 것으로

관찰된다. ASVspoofer 2021에서는 LA(Logical Access)와 DF(DeepFake) 트랙에서 가장 높은 점수를 받은 상위 5개의 모델이 모두 점수 평균화 기반의 융합 방법을 채택했으며, 이는 짧은 기간의 스펙트럼 특성, 또는 원시 파형을 입력으로 동작한다. 한 연구에 따르면 일반적으로 사용되는 아키텍처는 시간 지연 신경망(TDNN, Time Delay Neural Network)과 같은 CNN 기반 모델, 가우시안 혼합 모델(GMM, Gaussian Mixture Model), 그리고 Rawnet2와 같은 잔차 신경망(Resnet, Residual Neural Network) 계열 모델이 가장 인기 있다(Tan et al., 2021). 장단기 메모리(LSTM, Long Short Term Memory) 기반 아키텍처는 CNN 기반 아키텍처와 비교했을 때 연산 속도 측면에서 불리함이 있어 거의 사용되지 않는다. 따라서 임베딩 수준 및 점수 수준 융합에서 CNN 모델이 더욱 유연한 것도 자명한 사실이다. 현재까지는 구현의 용이성 때문에 점수 수준의 융합을 사용하는 시스템이 임베딩 수준의 융합보다 우세하게 사용된다. 그러나 가까운 미래에는 더 많은 임베딩 기반 접근법이 제안될 것으로 예상된다. 이 이유는 ASVspoofer와 SASV(Spoofing-Aware Speaker Verification) 챌린지에서 엿볼 수 있는데, ASVspoofer에서는 일반적으로 융합이 카운터메저 시스템 사이에서만 수행된다. 그러나 SASV 챌린지에서는 점수 수준 융합이 세 가지 결과(대상 음성, 대상 음성이 아닌 경우, 스푸핑 음성인 경우)를 가지는 것이 의미가 없다. 즉, 바이너리 분류 점수를 보정하는 기존의 접근법은 더 이상 적절하지 않다. 그러므로 SASV 챌린지가 SPF-EER의 스푸핑 탐지, SV-EER의 화자 인증 두 가지 작업에서 시스템을 공동으로 최적화하기 위해 ASV와 카운터메저 사이의 융합을 받아들일 것으로 예상된다. ASVspoofer는 고정된 ASV와 함께 일부 변경 사항이 존재할 수 있으며, DF 트랙에서 독립적이거나 융합된 카운터메저의 개발을 촉진할 수 있다.

## 2.2 데이터 증강(Data Augmentation)

데이터 증강(이하 DA, Data Augmentation)은 신경망이 더 견고한 특징 임베딩을 학습하게 하는 방법 중 하나로, 훈련 데이터의 접근이 제한될 때 데이터의 변화 등을 극복하기 위해 사용될 수 있다. 증강 기법은 일반적으로 미디어 코덱과 채널 다양성, 부가적인 소음, 주파수 마스킹, 그 외 기타로 네 가지 주요 카테고리로 분류될 수 있다. 이 중 가장 흔하게 사용되는 데이터 증강 기법은 다양한 유형의 미디어 코덱을 사용하는 방법이다. 또 VoIP(Voice Over Internet Protocol, 인터넷을 이용한 음성 전송)·전화 시스템을 통해 재전송을 하는 방법도 있다.

한 연구에서 크로스 데이터세트 실험은 채널의 변화로 인한 현저한 성능 저하를 보였다(Chen et al., 2021, Zang et al., 2021). 이러한 변화는 재전송 및 압축 알고리즘에 대한 오디오 효과와 같은 채널 효과에 의해 포착되었다. 이들의 연구는 ASVspoofer 2015, 2019 및 VCC 2020을 포함한 여러 데이터세트를 테스트하여 보나 파이드로 분류되는 점수가 낮아지는 것을 보고 EER(Enhanced Entity Relationship, 향상된 객체 관계)이 저하되었다는 것을 인식하는 것에서 시작했으며, 관찰한 사실을 기반으로 ECAPA-TDNN(Emphasized Channel Attention and Propagation and Aggregation Time Delay Neural Networks) 아키텍처를 도입, ASVspoofer 2021의 새로운 도전 과제인 DF를 해결하기 위해 채널의 견고성을 증강시키는 전략으로 모델을 훈련시켰다.

또 다른 연구자들은 다양한 데이터 증강 기법의 융합을 수행하였다. 그들은 ASVspoof 챌린지의 평가 프로토콜에서 몇 가지 “알려진” 데이터 증강 기법 고려, 이를 유용한 몇 가지 “알려지지 않은” 기법들과 결합했다. 예를 들어, mp2, ogg, ttg, wma 등의 코덱은 ASVspoof 2021의 딥페이크 트랙에서 도메인 내 훈련 데이터를 풍부하게 만들기 위한 “알려지지 않은” 데이터 증강 기법 중 일부이다(Das, 2021). LA 트랙에서는 샘플들이 G.711(alaw) 및 G.722 코덱을 사용하여 PSTN(Public Switched Telephone Network, 통신사업자가 전화 서비스를 제공하는 전화망) 또는 VoIP 네트워크를 통해 전송되었다. 이러한 채널 및 압축 알고리즘을 통한 데이터 증강 기법에 대한 연구가 해당 챌린지에서 카운터메저를 융합하기 위한 방법 중 최고로 뽑혔다.

ADD 챌린지(실제와 합성된 오디오에서 생성된 조작된 발화를 감지하기 위한 목적으로 등장한 최초의 대회) 2022에서는 또 다른 잘 알려진 기법인 mixup이 활용되었다. 연구팀은 주파수 특징 마스킹(이하 FFM, Frequency Feature Masking) 증강 기법을 제안, 5개의 스펙트로그램 기반 DNN에 mixup 증강을 적용했다. 이 방법은 특히나 소음이 많은 환경, 저품질 오디오 샘플에서 유용한 결과를 도출해냈다.

### 2.3 시스템 통합

앞서 서술한 바와 같이, 일반적으로 카운터메저는 ASV 시스템과 결합되어 스푸핑된 음성을 탐지, 거부하여 ASV 하위 시스템에서 처리되지 않도록 하여 개별적인 시스템인 카운터메저와 ASV가 하나로 통합될 수 있으며, 이는 특정한 대상 발화자만을 수용하는 견고한 ASV 시스템으로 기능한다.

위는 SASV 챌린지의 목표이며, SASV는 단일 통합 솔루션 외에도 공동으로 최적화되거나 융합된 카운터메저-ASV 솔루션의 연구를 촉진한다. 카운터메저와 ASV 하위 시스템의 결합은 스푸핑 공격에 대한 강건성을 가져 보안적 성능을 개선할 여지가 있지만(스푸핑 시도인 경우 거부, 진짜 대상이 시스템을 사용할 때의 시도를 거부하는 등 의도하지 않은 사이드 이펙트(side effect)로 되려 사용성을 저하할 가능성도 있다. 따라서 입력된 음성의 진짜 여부를 평가하기 위해서는 통합된 접근법이 바람직하다. 이는 곧 ASV 시스템에 대해서 스푸핑 및 카운터메저의 영향을 평가해야 하고, 카운터메저와 ASV를 독립적으로만 평가해서는 안된다는 것을 의미한다. 즉, 여기서는 궁극적으로 “신뢰할 수 있는(reliable) ASV”를 생성하는 것이 목표이다.

이를 위해 최소 텐덤 탐지 비용 함수(이하 min tDCF, minimum tandem Detection Cost Function)와 같은 평가 지표를 개발, 통합된 카운터메저와 ASV 하위 시스템을 평가할 수 있다(Kinnunen et al., 2020).

## IV. 다양한 기술적 이슈

### 1. 소음 및 채널 다양성

위의 내용에 더해, 강력한 카운터메저를 설계하기 위한 다양한 이슈가 존재한다. 먼저, 탐지 시스템이 배경 소음이나 채널의 특성과 같은 다양한 리코딩 조건(예를 들어, VoIP와 같은 채널을 통한 전송을 위한 특정 인코딩 체계)에 직면했을 때 그 성능을 얼마나 유지할 수 있는가에 관한 것이다. 최근 여러 연구가 카운터메저에서 다양한 데이터셋을 비교해 현대의 스푸핑 탐지기가 예기치 않은 조건에 놓였을 때 성능 저하를 겪을 수 있다는 것을 밝혔다. 이러한 크로스 데이터셋 효과에 기여하는 다양한 요소 중 특히 채널 효과가 카운터메저에 두드러진 성능 저하를 일으키는 역할을 한다는 것이 포착되었다. 이러한 효과는 탐지 시스템이 다양한 환경 변화에 걸친 시스템의 일관되고 신뢰성 있는 결과를 도출하게 하기 위해서는 무시되어선 안된다. 따라서 “증강” 데이터를 동반하는 기법은 모델이 특정 증강 기법에 과적합되지 않으면서 시나리오에 일반화할 수 있게끔 사용에 주의를 요한다.

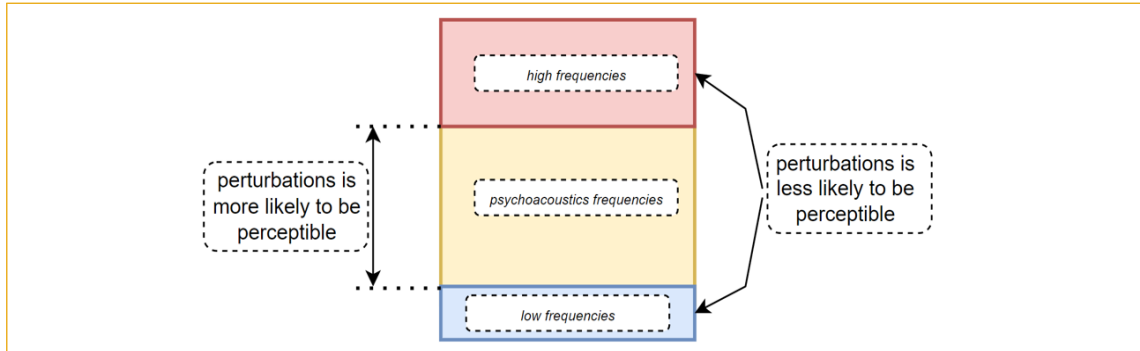
한편 크로스 데이터셋 연구는 견고한 카운터메저를 개발하는데 결정적인 역할을 한다. 실제 시나리오에서 시스템은 훈련 데이터 접근의 제약, 서로 다른 리코딩 채널에 의해 도입된 변동성 등으로 취약할 수 있다. 앞서 언급한 한 연구에서 논의한 바와 같이 “채널 효과”는 녹음 및 전송 단계에서 도입된 오디오 변경을 포함, 방음, 장치 주파수 응답, 통신 압축 알고리즘과 같은 요소 또한 포괄한다(Chen et al., 2021, Zhang et al., 2021). 음성 영역은 특히나 음성이 디지털 신호로 변환, 전송되는 과정에서 다양한 포맷이 존재하므로, 이러한 효과로 인해 카운터메저가 훈련 데이터에 존재하는 제한된 채널 변화에만 과적합되거나, 이로 인해 예기치 않은 채널을 직면하는 경우 시스템이 실패할 가능성이 있다.

### 2. 적대적 공격

적대적 공격(Adversarial Attack)은 학습 모델의 통계적인 가정에 위반을 일으키는 부가적인 데이터를 고의로 삽입하는 공격 기법이다. 딥러닝 모델은 대체로 많은 수의 하이퍼파라미터를 가지고 있기 때문에 이러한 공격에 취약하다. 이미지 도메인과 마찬가지로, 파형 샘플에는 인간의 인식에 영향을 주지 않고 변경할 수 있는 많은 아티팩트(artifact)가 포함되어 있어, 공격자가 음성 데이터를 조작하여 카운터메저와 같은 시스템을 속이려 할 수 있다. 과거에는 화자 인식 시스템(이하 SRS, Speaker Recognition System)에 대한 적대적 공격 주제에 관한 많은 연구가 제안되었다. 심리 음향 모델을 기반으로 인간이 인식할 수 있는 임계치보다 낮은 신호가 조작될 수 있으며, 이렇게 추가되는 변동은 그 인지도를 최소화하는 방향으로 생성된다. 청각 마스킹을 사용하면 단일 소리를 인식하는 데 있어 다른 소리의 존재에 영향을 받게 되어, 결론적으로 SRS를 손상시킬 수도 있다.



그림 6. 적대적 공격의 변조 범위

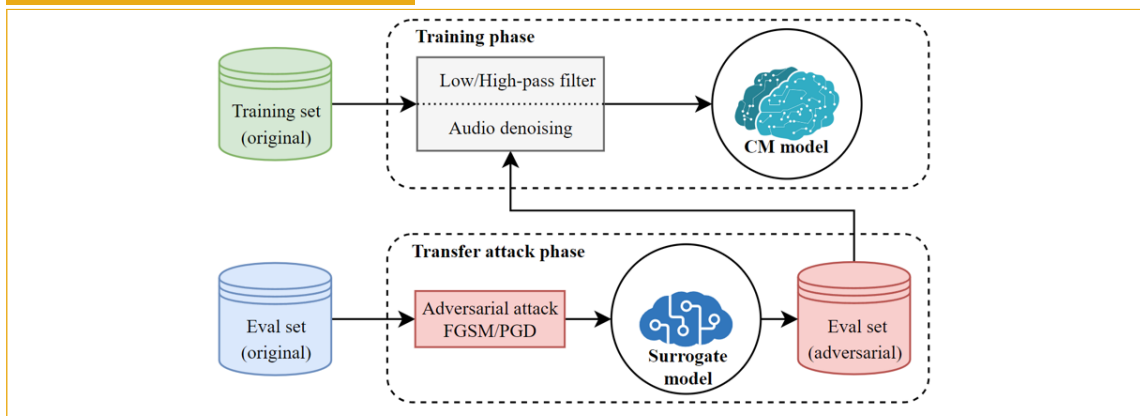


\* 출처: 저자 작성

적대적 공격은 대리 모델(Surrogate Model)을 사용하여 다양한 신경망 아키텍처 간에 전송될 수 있다. 일반적으로 적대적 공격은 블랙박스 또는 화이트박스 환경에서 수행되며, 이러한 차이점은 공격자가 대상 시스템에 대한 지식, 그러니까 모델의 아키텍처나 가중치, 활성화 및 손실 함수 등을 이해하고 있는지에 따라 다르다. 최근 몇 년 동안 이미지 도메인에서 효과적이었던 기술들이 ASV-CM 시스템을 공격하는 데 사용되었으며, 주로 FGSM(Fast Gradient Sign Method), BIM(Basic Iterative Method) 및 MI-FGSM 등이 있다. PGD(Projected Gradient Descent)는 적대적 예제를 생성하기 위한 인기있는 또 다른 접근법이며, JSMA(Jacobian-based Saliency Map) 및 C&W(Carlini & Wagner) 등도 존재한다.

한편, 이 적대적 공격에 대한 방어(Defense)는 두 가지의 주요 카테고리로 분류된다. 수동적(Passive) 방어와 능동적(Active) 방어가 그것이다. 수동적 방어의 경우 공간 스무딩(Spatial Smoothing)을 사용하여 잡음처럼 보이는 적대적인 변조의 영향을 억제하는 것이 가능하다. 그 외에도 입력 샘플을 정화(Purify)하기 위해 캐스케이드 TERA 모델이 개선되었는데, 이는 적대적 변조를 제거하고 보나 파이드 샘플의 손실 없는 재구성을 허용한다.

그림 7. 적대적 공격 샘플 생성 프로세스



\* 출처: 저자 작성

능동적 방어의 경우, 적대적 훈련(Adversarial Training)을 가장 첫째로 꼽을 수 있다. 적대적 샘플을 감지하기 위해 신경망을 사용할 수 있는데, 특히 보코더를 사용하여 오디오 샘플을 재합성하는 기술과 결합할 때 유용하다. 이 접근법은 보코더 처리 전후 임베딩 값을 비교함으로써 구현될 수 있는데, 이것이 성립되는 이유는 보코더가 “진짜” 데이터의 분포를 모델링하여 “진짜” 오디오 파형의 경우 왜곡이 덜 발생한다는 것을 근거로 한다. 비슷한 접근법으로는 “디노이징(Denoising)” 기술이 있으며 이 경우 추가된 적대적 변조가 노이즈로 간주된다.

이러한 적대적 공격 및 방어는 그간 이미지 영역에서 활발히 연구되어 왔으며, 이는 적대적 오디오 샘플의 생성 및 탐지에 빠르게 적용될 수 있음을 의미한다. 한편, FGSM, PGD와 같은 일반적으로 적용 가능한 적대적 공격의 출현 덕에, 적대적 훈련, 탐지 네트워크, 공간 스무딩과 같은 인기 있는 방어 대책이 뒤따라 개발되었다. 여기에 더해 음성 영역에서는 기존 생산적 적대 신경망(GAN, Generative Adversarial Networks)에 의존하는 적대적 공격의 성능을 향상시키기 위해 음성 변환 및 후처리와 같은 향상된 기술들도 등장했다.

또 다른 논의로, 능동적 방어의 탄력 ASV-CM 시스템 외에 전용 네트워크를 도입해야 하므로 많은 오버헤드(overhead, 어떤 처리를 하기 위해 들어가는 간접적인 처리 시간)가 발생할 수 있는 반면, 수동적 방어의 공간 스무딩 또는 정화, 디노이징 같은 기술로 적대적 노이즈를 중화하는 데에 중점을 둔다. 여기에 더해 앙상블 학습 등 강화된 블랙박스 공격은 보편적인 방어를 발견할 것으로 예상된다. 이렇듯, 이미지 분야에서 활발하게 논의된 적대적 공격과 그 진보들이 음성 영역에서 채택되는 것은 머지않아 보인다.

### 3. 부분 조작 오디오

부분 조작 오디오(Partially Fake Audio)는 “진짜” 음성과 “가짜” 음성을 혼합하여 생성된 파형 샘플을 의미한다. 일반적으로, 문장의 의미를 극적으로 변경하면서 동시에 탐지를 피하기 위해서는 합성된 음성의 길이, 그리고 이것의 주입이 최소화되어야 한다. 이러한 조건들은 언어의 특성 때문에 비교적 손쉽게 달성될 수 있는데, 보통 문장의 의미는 지리적 위치나 사람의 이름 같은 몇 단어만 수정해도 크게 바뀔 수 있기 때문이다. 예를 들어, “철수는 영희를 좋아해”와 “철수는 영희를 싫어해”의 두 문장은 짧은 두 음절의 차이로 정반대의 의미를 지닌다. 이 위협은 ADD 2022(ADD에는 PF(Partial Fake) 트랙이 따로 존재한다)에서 특히 강조되었다. 본 장에서는 이 새로운 위협에 대한 최근 연구를 논의할 것이다. 아직 관련 연구 및 사회에서의 인식은 제한적이나, 부분 조작 오디오의 기술적 측면, 그를 방어하기 위한 통찰력을 제공한다.

부분 조작 오디오를 연구하기 위한 초기 시도 중 하나는 2021년에 제안되어, 이는 PartialSpoof라는 새로운 데이터베이스를 소개했다. 연구자들은 현재까지의 카운터메저가 음성의 일부만 가짜인 부분 조작 오디오를 탐지하고, 가짜 세그먼트를 찾을 가능성에 대한 능력을 의심했다. 따라서 연구팀은 이를 시험하기 위해 다음의 단계를 따라 PartialSpoof 데이터베이스를 생성했다.

1. VAD(Voice Activity Detection) 알고리즘을 사용하여 음성 활동을 찾는다.

2. 진짜 세그먼트를 가짜 세그먼트와 무작위로 대체한다.
3. 파형을 정규화하고, 아티팩트를 피하기 위한 최적의 융합 지점을 찾는다. 여기서 아티팩트란 음성 부분 조작 시 융합지점에서의 인위적인 조작에 의해 발생하는 부자연스러움을 의미한다.
4. 발화는 가짜로, 세그먼트는 원본 오디오에 따라 가짜 또는 진짜로 표기한다.
5. ASVspoof 2019와 유사한 샘플을 생성하기 위해 2-4의 단계를 반복한다.

그 후 연구팀은 크로스 데이터셋 평가를 수행, 부분 조작 데이터베이스에서 훈련하는 것이 더 견고한 모델을 생성할 수 있음을 입증하였다. 부분 조작 오디오 샘플을 감지하는 적절한 방법은 발화 수준의 훈련 대신 세그먼트 수준 훈련이라는 점도 흥미롭다. 연구자들은 이후 이 아이디어를 개선, 발화 수준과 세그먼트 수준 탐지를 위해 단일 카운터메저를 훈련시킬 수 있는지에 대한 가능성에 의문을 제기하며 멀티태스크 학습(Multi-task Learning)을 제안했다. 이는 발화 수준 탐지와 세그먼트 수준 탐지가 서로 직접적으로 연관이 있기 때문이었는데, 그 실험 결과는 가히 우려스러웠다. 평가 세트에서 세그먼트 수준의 탐지가 향상되지 않았기 때문이다.

또 다른 연구에서는 AISHELL 코퍼스에서 파생된 HAD(Half-truth Audio Detection)이라는 데이터셋이 개발되었다. 이 데이터셋이 구축된 기술적 세부사항은 이전 연구와는 다른데, 그 과정은 다음과 같다.

1. 실제 오디오에서 대체할 키워드를 선택한다.
2. 해당하는 합성 클립을 선택한다.
3. 볼륨을 정규화하고 진짜 클립을 합성 클립으로 교체한다.

예를 들어, 교체에 사용되는 키워드는 구체적인 고유명사가 등장하는 엔터티(장소, 이름, 날짜 및 시간)이거나 특정 단어의 반의어가 될 수 있다. 결론적으로 연구자들을 세그먼트 수준의 실험이 조작된 특정 부분을 찾아내는 것이 어렵다는 것을 보였다.

또 다른 접근으로는, 진짜 음성과 가짜 음성을 연결하여 부분적으로 조작된 음성을 생성하는 시도가 있다. 이 접근에서는 다양한 비율로 연결된 가짜 샘플이 포함된 데이터 세트는 CQCC-GMM 모델에 대해 평가되었으며, 이는 제로-임계값 접근법으로 평균 기반 접근법보다 더 나은 성능을 보였다.

ADD 2022 챌린지를 위한 여러 연구 중 하나는 셀프 어텐션 메커니즘(Self-attention Mechanism)을 이용한 추출 기반 질문-답변 전략 도입, 안티 스푸핑 모델에 부분 조작 오디오의 시작과 끝을 찾도록 하는 작업을 수행했다. 이 전략은 모델의 구조 및 특징에 구애받지 않을 것으로 예상되며(모델-특징 중립적), ADD 2022 챌린지에서 EER이 7.9%로 두 번째로 높은 순위를 차지하였다. 또 다른 팀은 오디오를 세그먼트 단위로 분할하고, 각 세그먼트에 점수를 매기는 세그먼트 검출 전략에 대해 논의하였다. 이 기술은 챌린지에서 EER 40.5%의 성적을 받았다.

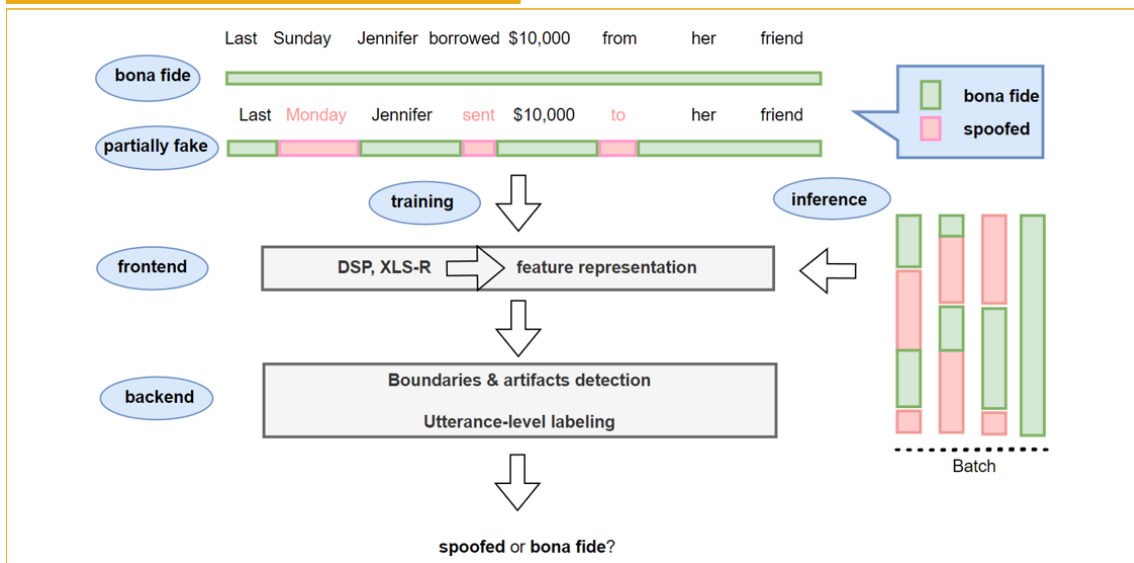
ADD 2022에서 1위를 차지한 것은 비지도 사전 학습(Unsupervised Pre-training) 접근법이었다. 해당

탐은 0.5%의 확률로 다른 발화에서 대상 발화로 클립(0.5초에서 1초까지 균일하게)을 무작위로 삽입했다. 그 후 부분적으로 조작된 오디오를 판별하기 위해, 조작된 클립의 삽입으로 인한 시간 도메인에서의 불연속성을 찾으려했다. 이는 EER 4.8%의 성적을 얻었다.

Vicomtech의 접근법은 ADD 2022에서 네 번째로 랭크되었는데, 여기서는 부분 조작 오디오로 데이터 증강을 수행했다. 이 기술은 세그먼트(진짜, 혹은 조작된 음성)를 사용하여 원래 발화에 무작위한 위치에 겹쳐 넣음으로써 데이터 증강을 도모했다.

이러한 부분 조작 오디오는 일반적인 가짜 오디오 탐지에 비해 상당히 복잡한 문제이다. 현재까지는 세그먼트 수준에서 점수를 매기는 것이 가장 일반적인 방법이며, 발화 내에서 대부분의 진짜 세그먼트에 숨겨진 가짜 세그먼트를 지역화하는 등의 셀프 어텐션 기술이 뒤따라 등장하고 있다. 부분 조작 오디오 샘플을 생성하는 데에는 파형 정규화, 음성 검출(VAD, Voice Activity Detection) 및 융합과 같은 많은 작업을 요구되며, 이에 따라 카운터메저의 성능을 향상시키는 작업 자체에도 여러 장애물이 존재한다. 예를 들어, 가짜 세그먼트의 지속 시간이 매우 짧아 식별하기 어려울 수 있다. 또는 부분 조작 오디오 탐지를 위한 데이터셋의 수가 매우 제한적이기 때문에, 가짜 세그먼트를 단순히 무작위로 배치한 데이터 등 데이터가 실제 사용 사례를 반영하지 않을 수 있다. 이는 모델 연구에 대한 실효성 및 신뢰성을 해칠 수 있다. 다른 문제로는 세그먼트 수준의 점수를 매기기 위한 표준적인 지표가 없다는 문제도 있다. 그러나 향후 연구에서는 부분 조작 오디오 생성 과정의 한계를 역이용하여 이러한 조작을 탐지할 수 있으리라 예상되는데, 이 이유는 다음과 같이 서술할 수 있다. 우선 부분 조작 세그먼트는 주로 음성 활동이 가능한 곳에 존재하며, 자동 파형 정규화 등으로 인해 부분적으로 조작되는 부분에서는 음성 파형의 불일치가 분명하게 나타날 수 있다. 이는 카운터메저 진짜와 가짜 음성의 주파수 스펙트럼의 사이에서 급격한 변화를 학습하게 한다. 게다가 조작

그림 8. 부분 조작 오디오 탐지의 일반적인 프로세스



\* 출처: 저자 작성

탐지의 궁극적인 목표는 입력 발화를 받아들일 것인지, 거부할 것인지 결정하는 것이므로, 가짜 세그먼트를 찾아내는 것은 실질적인 사용성에 있어서는 부차적인 과업이 될 수 있다. 서술한 바와 같이, 부분적으로 조작된 오디오 탐지에 대한 연구는 초기 단계이며, 가까운 미래에 지속적으로 성장할 전망이다. 이에 대한 방어 또한 인간의 음성과 소리를 활용하는 여러 산업의 핵심 요소가 될 수 있다.

표 4. 부분 조작 오디오 탐지를 위한 일반적인 기술

Method	ASV/CM System	Corpus	Performance	Metric
Antonym named entity editing Utterance and segmentation-level classification	GMM or LCNN	HAD	6.99(LCNN,CQCC) 72.54(LCNN,CQCC)	EER (%) Precision (%)
Frame Selection Threshold Selection	GMM	ASVspooft 2015 ASVspooft 2017 ASVspooft 2019	10.55 (CQCC)	EER (%)
Multi-Task Learning	SELCNN and Bi-LSTM	PartialSpooft	SegBW 6.07 (Utterance-level) 16.60 (Segmental-level)	EER (%)
Cross-database Analysis Utterance- and Segmental-level detection	LCNN and Bi-LSTM	PartialSpooft	16.21 (Segmental-level)	EER (%)
Self-Attention Fake span discovery	SENet	AISHELL-3 (derived) and other sources	11.1 (LFCC)	EER (%)
Segmenting detection mean predicted socre	Transformer-based encoder	AISHELL-3, 4	40.5 (Spectrogram)	EER (%)
Discontinuity spotting find spectrum blurs or sudden changes in spectrum	Pre-trained XLS-R [105]	CommonLanguage [106]	4.8 (LFCC feature)	EER (%)
Data Augmentation adaption data classifier finetuning with W2V2 (Wav2vec)	Pre-trained wav2vec2-based (XLS) and classifier (cosine layer)	AISHELL-3	16.59 (XLS-128: CNN encoder)	EER (%)

\* 출처: 저자 작성

## V. 향후 연구 이슈

강력한 카운터메저를 구축하기 위한 여러 노력에도 불구하고 여전히 해결해야 할 문제들이 존재한다. 여기서는 이제까지 제안된 방어 시스템의 성능을 저해하는 기술적 도전 과제를 간략히 검토하고 논의한다. 또한 향후 유망하게 연구될 수 있는 요소, 기술에 대해 소개한다.

## 1. 데이터 및 개념의 변화

이전까지의 카운터메저에서는 정적이고 명확한 발화로 시스템을 훈련시키다 보니, 성능적인 문제가 존재했다. 모델의 훈련을 위해서는 정교한 코퍼스를 구축하는 것이 중요하고, 여러 종류의 소음을 포함한 실제 데이터를 사용해 모델의 견고성을 평가해야 한다. 이전의 연구들은 인위적인 소음, 채널 변동, 여러 인코딩 체계 도입 등을 통해 이를 부분적으로 달성했다.

그러나 일부 연구팀에서는 훈련 데이터가 실제 환경의 데이터를 반영하지 않는다고 주장한다. 문제의 해결책 중 하나는 다양한 모델의 융합을 통해 극단적인 케이스에 대한 견고성을 증진, 카운터메저의 일반화를 강화하는 것이다. 예를 들어 에너지 기반 VAD는 침묵하는 기간을 제거함으로써 모델이 훈련 데이터로부터 더 의미 있는 특징을 학습하게 할 수 있다.

## 2. 부분 조작 오디오

현재까지의 연구 결과, 완전히 조작된 가짜 발화를 탐지하기 위한 대부분의 방어 메커니즘은 부분 조작 오디오를 탐지하는 능력이 없음이 밝혀졌다. 오디오 샘플에서 합성된 부분을 식별하는 것도 마찬가지로 어렵다. 이러한 부분 딥페이크 오디오를 생성하는 데에는 큰 노력을 요하지는 않지만 (예를 들어, 공격자는 문장에서 단어 하나만을 잘라내어 교체함으로써 전체 의미를 변화시킬 수 있다.) 그것들을 탐지하는 것은 도전적이고 유의미한 과제일 것이다.

## 3. 적대적 공격

적대적 공격은 이미지 영역에서는 이미 많은 발전을 이룩한 분야이나, 오디오 딥페이크 분야에서는 아직 많은 연구가 존재하지 않는다. 적대적 오디오 샘플을 생성하는 기술적인 세부사항(예를 들어 멜 스펙트로그램을 변조)은 이미지를 변조하는 것과 기본적으로 유사하지만, 발화의 생동감이나 자연스러움을 유지하는 것과 같은 여러 제약 조건이 적용될 수 있다. 그러나 이것들은 공격자가 FGSM, JSMA 및 그들의 변형을 사용해 여러 DNN 기반 모델을 무력화하는 샘플을 성공적으로 생성하는 것을 막지는 못한다. 이에 따라 오디오 딥페이크 분야에서의 적대적 공격에 대한 철저한 연구가 필요하다.

## 4. 결합 및 분리된 ASV-CM 시스템

공격 탐지(PAD, Presentation Attack Detection) 시스템을 개발하는 것의 궁극적인 목표는 가짜 발화로 인한 사기로부터 ASV 시스템을 보호하는 것이다. 지금까지 수행된 대부분의 연구는 독립적인 카운터메저의 평가에 중점을 두었으며, 이로 인한 사용성 문제가 발생했다. 따라서 PAD와 ASV를 공동으로 평가하는 것에 더욱 주의를 기울여야 한다. SASV에 따르면 ASV 시스템은 본질적으로 대상이 아닌 화자가

발화하는 소리 및 변환되거나 합성된 가짜 입력을 거부하도록 한다. 그러나 ASV 시스템이 스푸핑 시도를 마주했을 때 어떻게 조정되어야 하는지에 관한 주의가 없으며 카운터메저와 함께 작동하는 방법에 대해서도 다루지 않는다. 두 시스템을 함께 최적화하는 것 또한 탐구해야 할 문제 중 하나이다.

## 5. 보이스 컨버전

컨버전 된 음성을 탐지하는 것 또한 높은 장벽이 될 수 있다. 현재 보이스 컨버전 기술은 실시간에 가까운 성능으로 음성을 합성할 수 있는 능력이 있어 많은 측면에서 리스크가 존재한다. 현재까지 보이스 컨버전과 연관된 사건들은 제한적이지만(Sisman et al., 2020), 각 합성 기술의 특징을 따져보았을 때 공격자들이 취할 수 있는 음성 합성의 방법 중 악용될 가능성이 가장 높후하다고 판단된다.

### 5.1 공격의 특징화(Attack Attribution)

최신 연구에서는 합성 음성을 이용한 위협 환경을 더 잘 이해하기 위해 합성된 음성을 생성한 모델에 연관 짓는 것이 중요하다는 것을 발견했다. 즉, 존재하는 딥페이크 음성이 보이스 컨버전으로 생성된 음성인지 TTS로 생성된 음성인지, 원본 시스템을 찾아 연관시키는 것이다. 또한 단순 음성 신호만으로는 공격자의 신원을 특정 짓기에 충분하지 않으므로, Wav2vec 혹은 Deep-fusion과 같이, 해당 음성이 어떤 특징, 또는 융합 방식을 사용하여 생성되었는지 특징화하는 것도 합성 음성을 이해하는 데 많은 도움이 될 수 있다.

## VI. 정책 및 기업 동향

### 1. 합성·변조 음성 생성 및 탐지 기술에 대한 정책 동향

앞에서 서술한 바와 같이, 딥페이크 음성 분야는 합성도, 이에 대한 탐지도 주목을 받은 지 얼마 되지 않았다. 이로 인해 국내에서는 아직 딥페이크 합성이나 탐지에 대한 구체적인 규제나 정책이 마련되지 않은 실정이다. 그러나 인공지능 자체의 활용과 인공지능 윤리에 관한 정책은 몇몇 존재한다. 예컨대 2022년 과학기술정책연구원에서는 ‘인공지능 기술 활용 강국을 향한 과학기술정책 제고 전략’(오승환, 2022)을 발표했고 한국형사·법무정책연구원에서는 2017년 인공지능 기술을 이용한 ‘법과학을 적용한 형사사법의 선진화 방안’을 발표했다(윤지영, 2017). 또한 인공지능의 과도한 발전으로 이에 따른 윤리적 문제가 부상, 이를 다루기 위한 2017년 한국행정연구원의 ‘인공지능 기술 확산에 따른 위험 거버넌스 연구’(류현숙, 2017), 2018년 정보통신정책연구원의 ‘인공지능시대 법제대응과 사회적 수용성’(조성은, 2018) 등의 정책이 제시되었다. 국가 차원에서도 이러한 문제에 대한 관심과 대응이 존재한다. 현재 발표된 의안으로는 2023년 8월 발표된 ‘인공지능 책임 및 규제법안’ 및 2023년 2월 발표된 ‘인공지능책임법안’ 등이 제안되었으며,

현재까지는 계류상태이다.

그러나 외국의 사례는 조금 다르다. 중국과 미국의 경우 이미 딥페이크 악용 사례가 빈번히 발생, 이에 따른 위협을 인지하고 있다. 특히 바이든 정부는 AI와 관련된 공식 웹사이트(<https://ai.gov/actions/>)를 개설하고 'AI는 바이든 대통령의 최우선 과제'라는 슬로건을 내걸며, AI와 관련된 여러 행보를 보였다. 최근에는 2023년 10월 30일 '안전, 보안, 신뢰 가능한 인공지능 개발 및 사용에 관한 행정명령'을 발령, AI 개발에 대한 다양한 위협과 보호 방안을 포함하는 엄격한 지침을 다루었다. 중국에서는 2022년 11월 '인터넷 정보서비스 딥페이크 관리 규정'이라는 정책을 발표하였는데, 이는 정확히 '딥페이크'라는 용어를 사용하여 이와 관련한 관리 문제를 다룬다.

따라서, 외국에서의 사례와 같이 우리나라도 후발주자가 되지 않기 위해 빠르고 체계적으로 'AI 창작물과 그 윤리', 더 나아가 '딥페이크 합성 및 탐지'와 같은 화제에 대응해야 한다. 이를 위해 국내외의 사례를 철저히 분석, 다양한 이해당사자들과의 소통을 통해 실질적이고 효과적인 대책 수립이 필요하다. 이를 통해 우리나라도 딥페이크와 같은 현재의 문제에 적극적으로 대응할 수 있으며, 안전하고 신뢰할 수 있는 디지털 사회를 구축할 수 있을 것이다.

## 2. 합성·변조 음성 생성 및 탐지 기술에 대한 기업 동향

여기서는 합성·변조 음성 생성 및 탐지 관련 기술에 관여하는 국내의 기업들의 동향을 살펴본다. 그에 앞서, 간단히 음성 합성 기술의 유래와 최신 쓰임을 살펴보려 한다. 합성 및 변조 음성 생성 기술은 본래 텍스트 형식의 정보 전달이 어려운 상황에서 사용자들, 개중에서도 시각장애인을 보조하기 위한 기술로써 각광 받았으나, 현재에는 미디어 산업에 호황에 따라 급격한 발전을 이룩하고 있다. 예를 들어 코로나-19 이후, '오디오북' 시장은 지속적으로 성장세를 보이고 있으며, 2030년까지 약 350억 달러의 시장을 형성할 것이라고 한다(매일신문, 2023.10.18). 이 외에도, AI 음성 합성 분야는 사물인터넷(IoT, Internet of Things)이나 챗봇, AI 비서 또는 장애인 보조기기에 더해 아이돌 문화가 만연한 한국에서는 AI를 이용한 가수나 아이돌 그룹 등 다양한 분야에서 사용되고 있다.

한국에서도 여러 기업들이 합성 및 변조 음성 생성 기술을 더욱 정교하게 구현하기 위해 힘쓰고 있다. 그중 KT는, 음성 데이터 확보가 용이한 통신 3사 중에서도 음성 관련 AI 개발에 가장 주력하고 있다. KT는 AI 보이스 스튜디오를 출시한 후, 이를 온라인 독서플랫폼인 '밀리의 서재'에 적용하기도 했다(KERC, 2022). 특히나 합성 음성에 '감정 더빙'을 더해 음성 합성 기술을 고도화하였으며(뉴스토마토, 2022.10.05), 멀티링구어 서비스를 각국에 제공하고 있다. 그 외에도 SK텔레콤은 현재 독거노인 가구 증가에 따른 사회적 문제 해결을 위해 AI 음성 합성 기술을 이용해 대상 가구의 안부를 확인하는 '누구 비즈콜' 서비스를 출시하였다. 또 LG유플러스는 소상공인을 위한 'AI 가게 매니저' 등의 서비스를 출시할 예정이다(서울신문, 2023.11.01).

통신 3사뿐 아니라, 게임 분야에서도 음성 합성 AI에 대한 연구가 활발하다. 흔히 국내 대형 게임사라



불리는 3N(엔씨소프트·넷마블·넥슨)은 게임 분야에서 AI를 활용하기 위한 R&D에 힘을 쏟고 있으며, 그중 엔씨소프트는 스피치 랩을 따로 두어 음성 관련 분야의 연구개발을 아끼지 않고 있다. 특히나 디지털 휴먼을 실제로 구현하기 위해 노래하는 AI를 개발, 음성 분야에서는 유력한 학회로 꼽히는 ICASSP에도 매년 논문을 내는 등 활발한 활동을 이어가고 있다. 넷마블에서도 마젤란실, 콜럼버스실로 구성된 AI 센터를 신설, 음성 합성 기술 등을 연구하며 게임을 위한 AI 연구에 박차를 가하고 있다.

뿐만 아니라, 이제는 NES(Natural End-to-end Speech Synthesis)로 잘 알려진 네이버의 음성 인공지능 클로바, 딥브레인 AI에서 출시한 AI studios 서비스도 잘 알려져 있다.

위처럼 정교한 음성 합성을 위한 연구가 심화, 합성되는 음성의 품질이 향상될수록 이를 악용하는 사례가 증가할 것은 자명하다. 실제로 최근 딥페이크 음성을 악용한 보이스피싱이나 사기 위협이 증가하고 있는 추세인데, 이에 대응하기 위한 국내의 노력은 터무니없이 부족하다.

먼저 국내에서 AI 딥페이크 탐지 기술을 선도하는 기업은 딥브레인 AI라고 할 수 있다. 딥브레인은 종합 탐지, 특정 인물 탐지 및 음성 탐지 모델을 개발, 최근 공개하였으며, 이는 CNN 및 트랜스포머 모델을 기반으로 설계되었다(인공지능신문, 2023.08.03). 그러나 딥브레인을 제외하고는 딥페이크 탐지 분야에서 두각을 보이는 국내 기업은 전무한 실정이다. 현재까지는 소수의 국내 대학 연구팀에서 딥페이크 미디어 악용의 심각성을 깨닫고, 선진 연구를 수행하고 있다. 앞서 기술했듯 딥페이크 미디어의 위험성이 국내에서는 최근 대두되었으므로, 국내의 많은 기업들도 이러한 현상에 발맞추어 딥페이크 탐지 분야로 눈을 돌릴 것으로 예견된다.

반면 해외에서는 리얼리티 디펜더(Reality Defender), 센티널(Sentinel), 인텔(Intel)의 Real-Time Deepfake Detector 및 마이크로소프트(Microsoft)의 Video Authenticator tool 등 딥페이크를 탐지하려는 노력이 활발하게 이루어지고 있다. 그러나 대부분은 합성물의 프레임 간의 미세한 차이 등을 기반으로 딥페이크를 탐지한다. 인텔의 탐지기의 경우, 흥미롭게도 사람의 얼굴에서 픽셀 단위의 '혈류'를 캡처, 이 신호를 시공간 지도로 변환해 탐지한다. 그러나 음성 측면에서 딥페이크 탐지에 대한 노력은 여전히 불충분하다. 동영상에서 음성을 기반으로 딥페이크를 탐지하기 위해 입 모양의 변화와 실제 발화되는 음성의 차이를 분석하여 딥페이크를 탐지하려는 시도도 존재하지만(Agarwal et al., 2020) 이 또한 연구자료로, 아직 딥페이크 음성 탐지, 혹은 음성을 이용한 딥페이크 탐지 등은 연구 및 개발 단계에 있는 것으로 추측된다.

## Ⅶ. 결론

딥러닝 시대의 도래와 함께, 인공지능의 응용 분야와 뒤따르는 위협들이 확산되었다. 특히나 음성 영역에서는 TTS나 보이스 컨버전 같은 음성 합성 기술이 눈부신 발전을 이룩하였고, 이러한 합성 음성 등을 악용한 범죄 또한 가짜 뉴스에서부터 보이스 피싱까지 확장되었다. 이렇듯 음성 합성 기술을 남용 및

오용하는 사례가 수면 위로 떠오르면서, 오디오 딥페이크를 방어하기 위한 대응책, 즉 카운터메저 시스템 또한 그 자체로 유망한 연구주제가 되었다. 해당 과업을 수행하기 위해 ASVspoof, SASV 및 ADD 등 본 원고에서 서술한 많은 챌린지들이 관련 트랙을 소개하였으며, 음성 영역 스푸핑 탐지에 대한 활발한 연구를 추구하고 있다. 각 챌린지들은 고유의 과업을 가지고 있기는 하지만, 일반적으로는 카운터메저를 자동 화자 인증 시스템에 탑재하거나, 혹은 개별 카운터메저를 사용해 합성 음성을 탐지하는 것을 목적으로 한다.

현존하는 카운터메저는 그 기반 구조에 따라 DSP 기반 프론트 엔드, DNN 기반의 프론트 및 백 엔드, 그리고 백 엔드를 다시 두 카테고리로 나누어 발화 수준과 세그먼트 수준으로 구분할 수 있다. 자세한 기술적 세부사항은 위에서 묘사한 바와 같으며, 이렇게 구성된 카운터메저의 성능을 향상시키기 위한 방안으로는 모델 융합 및 데이터 증강 등의 방법이 있다. 그 외 주목할 만한 이슈로는 음성 데이터의 소음 및 채널 다양성, 카운터메저에 대한 적대적 공격, 부분 조작 오디오 등을 꼽을 수 있다.

본 원고에서 서술한 바와 같이, 딥페이크 오디오를 탐지하기 위한 카운터메저 기술은 활발히 연구 중이다. 여러 연구팀이 보다 강건한 카운터메저 시스템을 구축하기 위해 다방면으로 노력하고 있으나, 음성 탐지 영역에서의 AI는 여전히 성숙하지 못하다. 이는 일반적으로 사용되는 데이터의 차이라고 여겨지는데, 음성 데이터는 이미지 데이터와 연관성은 있으나 실상 서로 아주 다른 성향을 갖기 때문이다. 이미지는 픽셀로 이루어진 행렬 데이터에 불과하지만, 음성은 실제 음성 신호(raw-waveform)에서 스펙트로그램, 그리고 LFCC나 MFCC 같은 고수준 특성까지 사용될 수 있어 사용 가능한 데이터의 특징이 다채롭고, 정형화되어 있지 않다. 이에 더해, 어떤 수준의 데이터(발화 수준 및 세그먼트 수준 등)를 사용할 것인지, 어떤 인코딩을 써서 녹음하고, 어떤 채널을 통해 전송될 것인지 등 모델을 설계하기 전에 주의를 기울일 부분이 많다.

이에 따라, 딥페이크 오디오를 탐지할 수 있는 강력한 카운터메저를 설계 및 구축하는 한편, 이것이 기존 ASV 시스템과 함께 사용되는 상황에서도 그 강건성을 시험해야 한다. 결론적으로 ASV가 존재하는 상황 및 존재하지 않는 상황에서 카운터메저는 일관되고 신뢰성 있는 결과를 도출해야 하며, 이를 위해서는 각각의 시스템을 개별적으로 일반화하는 것이 아닌 각 시스템끼리의 상호 작용을 이해, 함께 동작할 수 있도록 시스템을 설계 및 최적화하여야 할 것이다. 우리나라에서도 딥페이크 오디오를 악용한 보이스 피싱, 그에 따른 금전적인 피해가 심화되는 만큼 강력한 카운터메저 개발을 위한 연구 지원 및 원천 기술 확보를 위해 더욱 노력해야 할 것이다.

#### 저자소개 정수환(Souhwan Jung)

##### • 학력

(美) 워싱턴대학교(Univ. of Washington) 전기공학 박사  
서울대학교 전자공학 석사  
서울대학교 전자공학 학사

##### • 경력

現) 송실대학교 전자정보공학부 교수  
前) 한국정보보호학회 회장

## ●● 참고문헌 ●●

### 〈국내문헌〉

- 1) 류현숙. (2017). 인공지능 기술 확산에 따른 위험 거버넌스 연구. 한국행정연구원.  
[https://www.nkis.re.kr/subject\\_view1.do?otpld=KIPA00052307&otpSeq=0&popup=P](https://www.nkis.re.kr/subject_view1.do?otpld=KIPA00052307&otpSeq=0&popup=P).
- 2) 오승환. (2022). 인공지능 기술 활용 강국을 향한 과학기술정책 제고 전략. 과학기술정책연구원.  
[https://www.nkis.re.kr/subject\\_view1.do?otpld=OTP\\_0000000000006674&otpSeq=0&popup=P](https://www.nkis.re.kr/subject_view1.do?otpld=OTP_0000000000006674&otpSeq=0&popup=P).
- 3) 윤지영. (2017). 법과학을 적용한 형사사법의 선진화 방안 (Ⅷ) : 인공지능 기술. 한국형사·법무정책연구원.  
[https://www.nkis.re.kr/subject\\_view1.do?otpld=KIC00052966&otpSeq=0&popup=P](https://www.nkis.re.kr/subject_view1.do?otpld=KIC00052966&otpSeq=0&popup=P).
- 4) 조성은. (2018). 인공지능시대 법제대응과 사회적 수용성. 정보통신정책연구원.  
[https://www.nkis.re.kr/subject\\_view1.do?otpld=OTP\\_0000000000001589&otpSeq=0&popup=P](https://www.nkis.re.kr/subject_view1.do?otpld=OTP_0000000000001589&otpSeq=0&popup=P).

### 〈국외문헌〉

- 1) Agarwal, S., Farid, H., Fried, O., & Agrawala, M.. (2020). Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops(CVPRW), Seattle, WA, USA: IEEE, pp. 2814-2822. doi: 10.1109/CVPRW50498.2020.00338.
- 2) Chen, X., Zhang, Y., Zhu, G., Duan, Z. (2021). UR Channel-Robust Synthetic Speech Detection System for ASVspoof 2021. Proc. 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge, 75-82, doi: 10.21437/ASVSPOOF.
- 3) Das, R. K. (2021). Known-Unknown Data Augmentation Strategies for Detection of Logical Access, Physical Access and Speech Deepfake Attacks: ASVspoof 2021. Proc. 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge, pp.29-36.
- 4) Doan, T., Nguyen-Vu, L., Hong, K., Jung, S. Synthetic Speech Detection Using Vocoder Signature Extraction, ICASSP 2024.
- 5) Jung, J. W., Tak, H., Shim, H. J., Heo, H. S., Lee, B. J., Chung, S. W., Kang, H. G., Yu, H. J., Evans, N. and Kinnunen, T.(2022a). Sasv Challenge 2022: A Spoofing Aware Speaker Verification Challenge Evaluation Plan. arXiv preprint arXiv:2201.10283.
- 6) Jung, J. W., Tak, H., Shim, H. J., Heo, H. S., Lee, B. J., Chung, S. W., Yu, H. J., Evans, N., Kinnunen, T. (2022b). SASV 2022: The First Spoofing-Aware Speaker Verification Challenge. arXiv preprint arXiv:2203.14732.
- 7) Kinnunen T, Delgado H, Evans N, Lee KA, Vestman V, Nautsch A, Todisco M, Wang X, Sahidullah M, Yamagishi J, Reynolds DA. (2020). Tandem assessment of spoofing countermeasures and automatic speaker verification: Fundamentals. IEEE/ACM Transactions on Audio, Speech, and Language

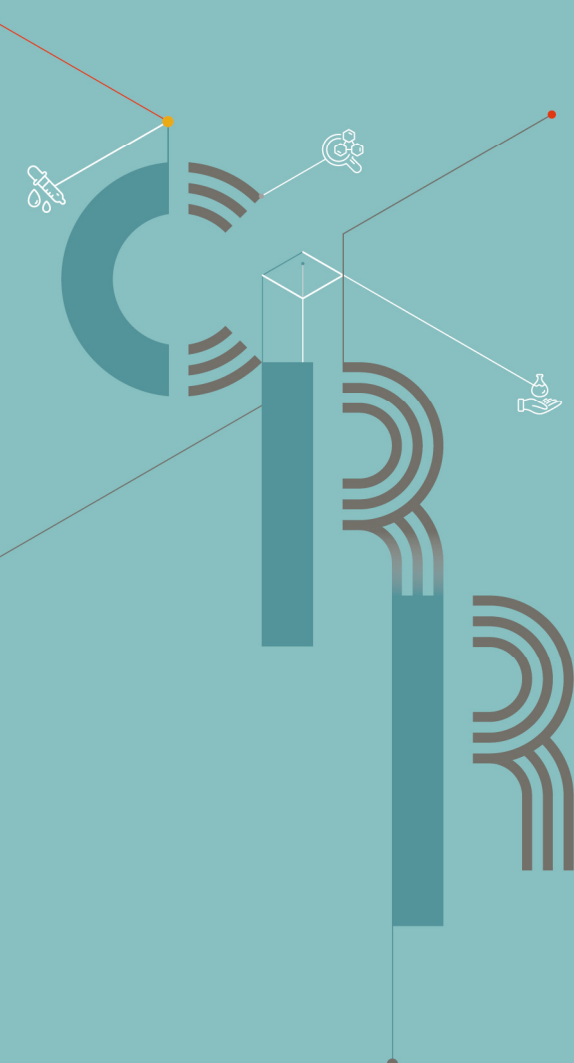
Processing.

- 8) Nautsch, A., Wang, X., Evans, N., Kinnunen, T. H., Vestman, V., Todisco, M., Delgado, H., Sahidullah, M., Yamagishi, J. and Lee, K. A. (2021). ASVspooF 2019: Spoofing Countermeasures for the Detection of Synthesized, Converted and Replayed Speech. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(2), pp.252–265.
- 9) Sisman, B., Yamagishi, J., King, S., Li, H. (2020). An Overview of Voice Conversion and its Challenges: From Statistical Modeling to Deep Learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:132–57.
- 10) Tak, H., Todisco, M., Wang, X., Jung, J.-w., Yamagishi, J., Evans, N. (2022) Automatic Speaker Verification Spoofing and Deepfake Detection Using Wav2vec 2.0 and Data Augmentation. *Proc. The Speaker and Language Recognition Workshop (Odyssey 2022)*, 112–119, doi: 10.21437/Odyssey.2022-16.
- 11) Tak, H. (2023). End-to-End Modeling for Speech Spoofing and Deepfake Detection(Doctoral dissertation, University of Southampton, UK).
- 12) Tan, X., Qin, T., Soong, F., Liu, T. Y. (2021). A Survey on Neural Speech Synthesis. *arXiv preprint arXiv:2106.15561*.
- 13) Wang, X., Yamagishi, J. (2022). A Practical Guide to Logical Access Voice Presentation Attack Detection. *Frontiers in Fake Media Generation and Detection*. pp. 169–214. Singapore: Springer Nature Singapore.
- 14) Wu, Z., Kinnunen, T., Evans, N., Yamagishi, J., Hanilçi, C., Sahidullah, M. and Sizov, A., (2015). ASVspooF 2015: the First Automatic Speaker Verification Spoofing and Countermeasures Challenge. In *Sixteenth annual conference of the international speech communication association*.
- 15) Yi, J., Wang, C., Tao, J., Zhang, X., Zhang, C. Y., Zhao, Y. (2023). Audio Deepfake Detection: A Survey. *arXiv preprint arXiv:2308.14970*.
- 16) Yi, J., Fu, R., Tao, J., Nie, S., Ma, H., Wang, C., Wang, T., Tian, Z., Bai, Y., Fan, C. and Liang, S. (2022) Add 2022: the First Audio Deep Synthesis Detection Challenge. In *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 9216–9220). IEEE.
- 17) Zhang, Y., Zhu, G., Jiang, F. and Duan, Z. (2021). An Empirical Study on Channel Effects for Synthetic Voice Spoofing Countermeasure Systems. In *Interspeech*.

## 〈기타 문헌〉

- 1) ABurt, T., & Horvitz, E. (2020.09.01). New Steps to Combat Disinformation. Microsoft. <https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-news-guard-video-authenticator/>.
- 2) KEREC. (2022.07.20). KT Releases Multilingual AI Voice Synthesis Service Capable of Emotional Expression.

- <https://k-erc.eu/kt-releases-multilingual-ai-voice-synthesis-service-capable-of-emotional-expression/>.
- 3) The White House. Administration Actions on AI. <https://ai.gov/actions/>.
  - 4) The White House. (2023.10.31). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.
  - 5) SK텔레콤 보도자료. (2022.11.03). SKT, '누구 비즈콜'로 서울 시민의 안부 챙긴다. <https://news.sktelecom.com/182708>.
  - 6) 김민석. (2023.11.01). LGU+ 구축·구독·소상공인용 '3대' AICC 서비스 낙점. 서울신문. <https://www.seoul.co.kr/news/newsView.php?id=20231101500077>.
  - 7) 넷마블. 마젤란실. <https://netmarble.engineering/tag/마젤란실/>.
  - 8) 네이버. 클로바. <https://clova.ai/voice>.
  - 9) 딥브레인AI. AI Avatar Video Generator | AI STUDIOS. <https://www.deepbrain.io/aistudios>.
  - 10) 리얼리티 디펜더. Reality Defender — Enterprise-Grade Deepfake Detection. <https://realitydefender.com>.
  - 11) 박진중. (2023.10.18). 셀바스, 'AI 오디오북 Editor' 선보여...글로벌 시장 정조준. 매일신문, <https://news.imaeil.com/page/view/2023101817283713378>.
  - 12) 센티넬. Sentinel - Defending Against Deepfakes and Information Warfare. <https://thesentinel.ai/>.
  - 13) 엔씨소프트(NCSOFT). NCSOFT Speech AI. <https://nc-ai.github.io/speech/>.
  - 14) 인텔. (2022.11.14). Intel Introduces Real-Time Deepfake Detector. <https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html#gs.2i4bt3>.
  - 15) 의안정보시스템. (2023.08.08). 입법부. 인공지능 책임 및 규제법안. [https://likms.assembly.go.kr/bill/billDetail.do?billId=PRC\\_W2E3F0D8D0C1A0B9I3J1I2I1G4F0G1](https://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_W2E3F0D8D0C1A0B9I3J1I2I1G4F0G1).
  - 16) 의안정보시스템. (2023.02.28). 입법부. 인공지능책임법안. [https://likms.assembly.go.kr/bill/billDetail.do?billId=PRC\\_B2B3H0G2F2B0A1Z4Y0N1M4I2I6G2N7](https://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_B2B3H0G2F2B0A1Z4Y0N1M4I2I6G2N7).
  - 17) 중국중앙인민정부. (2022). 인터넷 정보서비스 딥페이크 관리 규정(互联网信息服务深度合成管理规定). [https://www.gov.cn/zhengce/zhengceku/2022-12/12/content\\_5731431.htm](https://www.gov.cn/zhengce/zhengceku/2022-12/12/content_5731431.htm).
  - 18) 최강민. (23.08.03). 딥브레인AI, 딥페이크 탐지 솔루션 고도화...“조작된 영상·이미지, 음성까지 탐지 범위 확대. 인공지능신문. <https://www.aitimes.kr/news/articleView.html?idxno=28663>.
  - 19) 홍연. (2022.10.05). 고도화되는 AI 음성합성기술...업계, 서비스 확대 고심. 뉴스토마토, <https://www.newstomato.com/ReadNews.aspx?no=1151197>.



# 융합연구리뷰

Convergence Research Review

# 03

---

## 국가R&D 현황 분석

---

# 03

## 국가R&D 현황 분석

융합연구리뷰 11월호에서 다룬 2개의 주제(마약 신속·간편 검출 기술 및 변조 음성 탐지 기술)에 대한 각각의 국가 R&D 현황을 살펴보기 위해 국가연구개발 과제 분석을 수행하였다. 연구비를 기준으로 연구비 규모별 과제수, 연구수행 주체, 연구수준, 연구분야(국가과학기술표준분류, 미래유망신기술분류) 등 여러 측면에서의 분석 결과를 제시한다.

### I. 마약 신속·간편 검출 기술

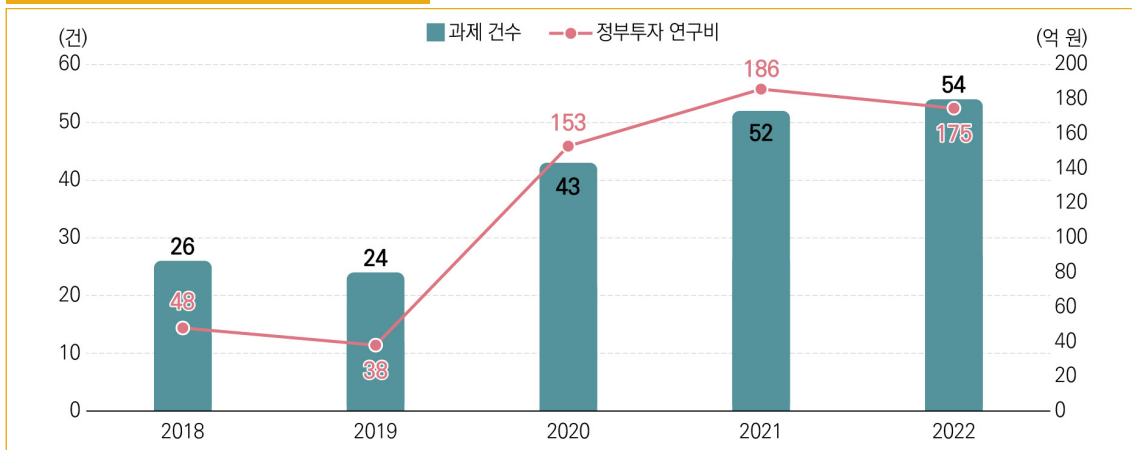
■ **(총괄)** 최근 5년간('18~'22) 총 199건의 과제에 대해 602억 원의 연구비가 투자됨

※ 국가과학기술지식정보서비스(NTIS) 플랫폼을 기반으로 관련 국가 연구개발 과제 분석 수행 : 핵심 키워드인 '마약 검출', '마약 검사', '마약 진단'으로 검색 실시

\* 마약 AND (검사 OR 검출 OR 진단)

\* 신속, 간편을 포함한 결과는 10건 미만으로 해당 데이터만으로는 분석이 어려워 마약 검출·검사·진단으로 분석 수행

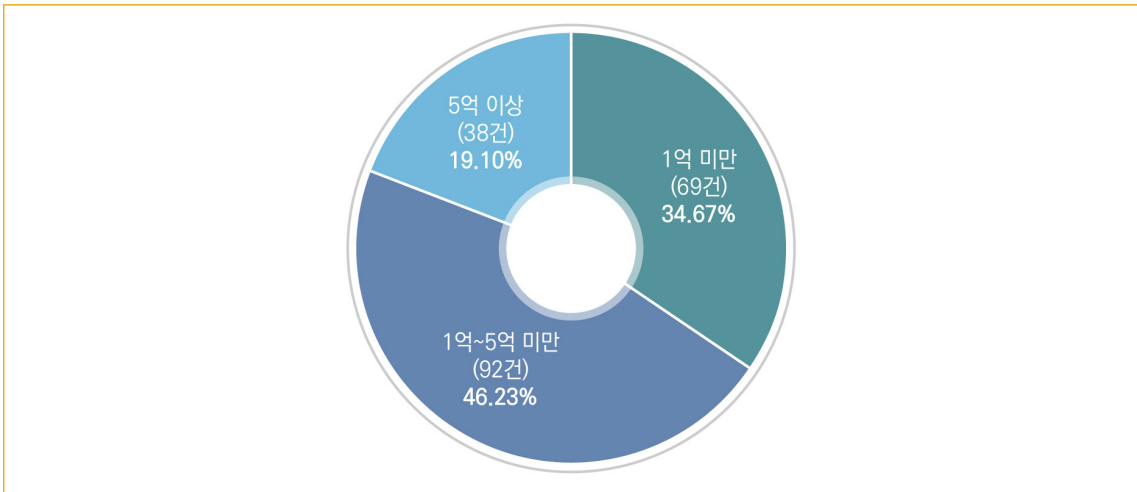
그림 1. 연도별 연구과제 건수 및 연구비





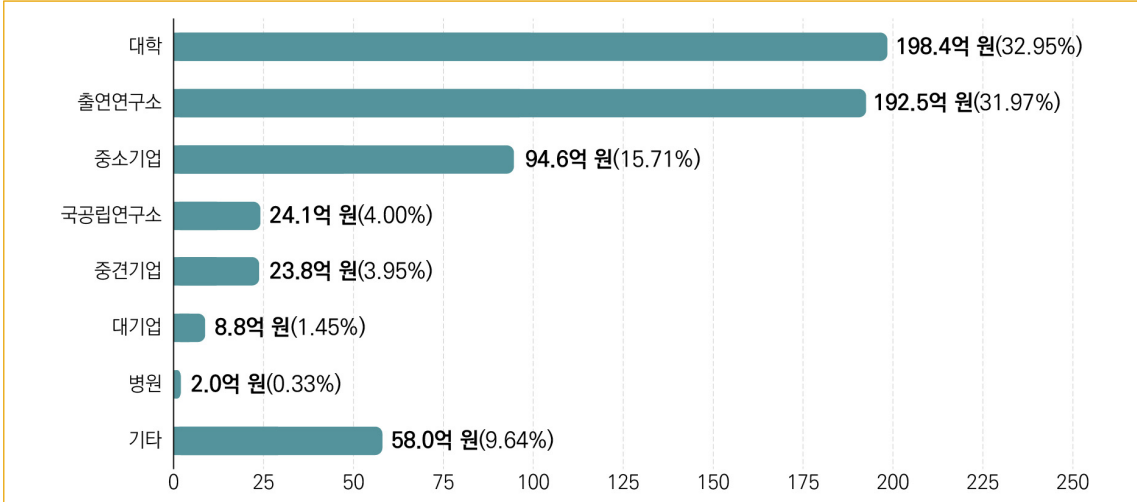
- **(연구비 규모별 과제 수)** 연구비가 1억 원 이상 5억 원 미만인 과제가 46.23%(92건)으로 가장 큰 비중을 차지하고, 1억 원 미만의 과제(34.67%, 69건), 5억 원 이상의 과제(19.1%, 38건) 순으로 큰 비중을 차지하는 것으로 나타남

그림 2. 연구비 규모별 과제 수 및 비중



- **(연구수행주체)** 대학과 출연연구소가 지원받는 연구비의 비중이 각각 32.95%(198.4억 원), 31.97%(192.5억 원)로 비슷하고, 중소기업(15.71%, 94.6억 원), 국공립연구소(4.00%, 24.1억 원), 중견기업(3.95%, 23.8억 원), 대기업(1.45%, 8.8억 원), 병원(0.33%, 2.0억 원) 순으로 큰 것으로 확인됨

그림 3. 연구수행주체별 연구비 규모 및 비중



- (연구수준) 마약 검출 연구는 응용연구 단계의 연구와 도입기 연구 위주로 연구비가 투자되고 있음
  - (연구개발단계 분석 결과) 응용연구에 투자되는 연구비 비중이 44.39%(267.3억 원)로 기초연구(31.85%, 191.8억 원), 개발연구(22.04%, 132.7억 원)에 비해 큰 것으로 나타남
  - (연구개발성격 분석 결과) 시작품개발의 연구비 비중이 40.05%(241.2억 원)인 반면, 아이디어개발과 제품 또는 공정개발의 연구비 비중은 각각 5.04%(30.3억 원)와 3.42%(20.6억 원)에 불과한 것으로 확인됨
  - (기술수명주기 분석 결과) 도입기에 투자되는 연구비 비중이 44.26%(266.5억 원)로 가장 크고, 그 다음으로 성장기에 투자되는 연구비 비중이 15.27%(91.9억 원)로 큰 것으로 드러남

그림 4. 연구개발단계별 연구비 규모 및 비중

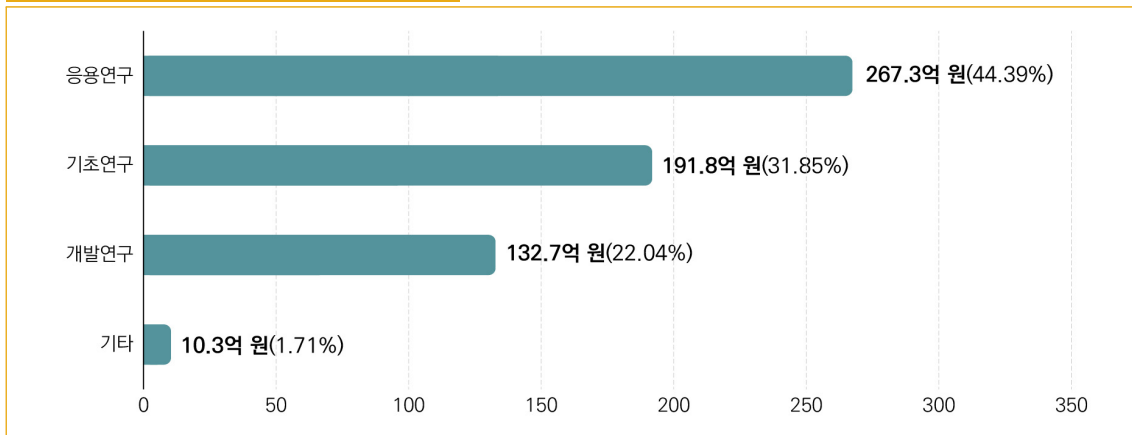


그림 5. 연구개발성격별 연구비 규모 및 비중

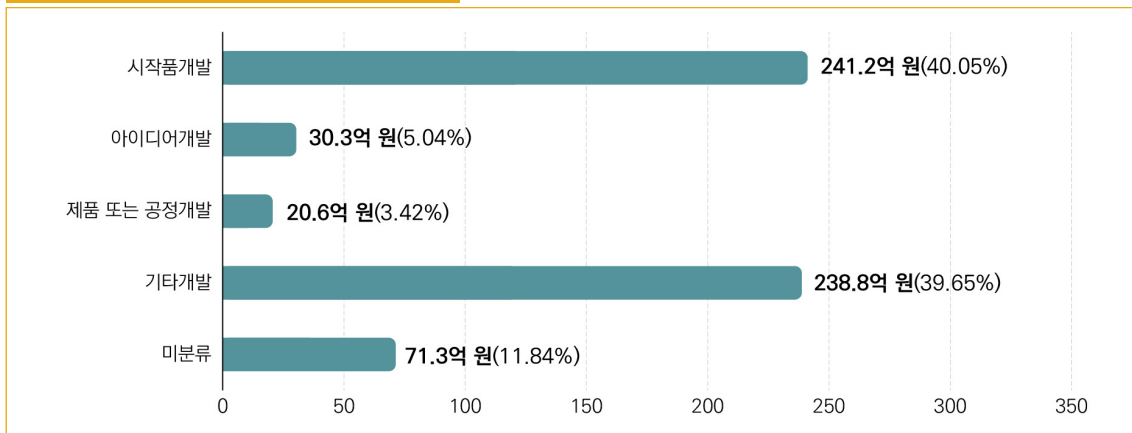
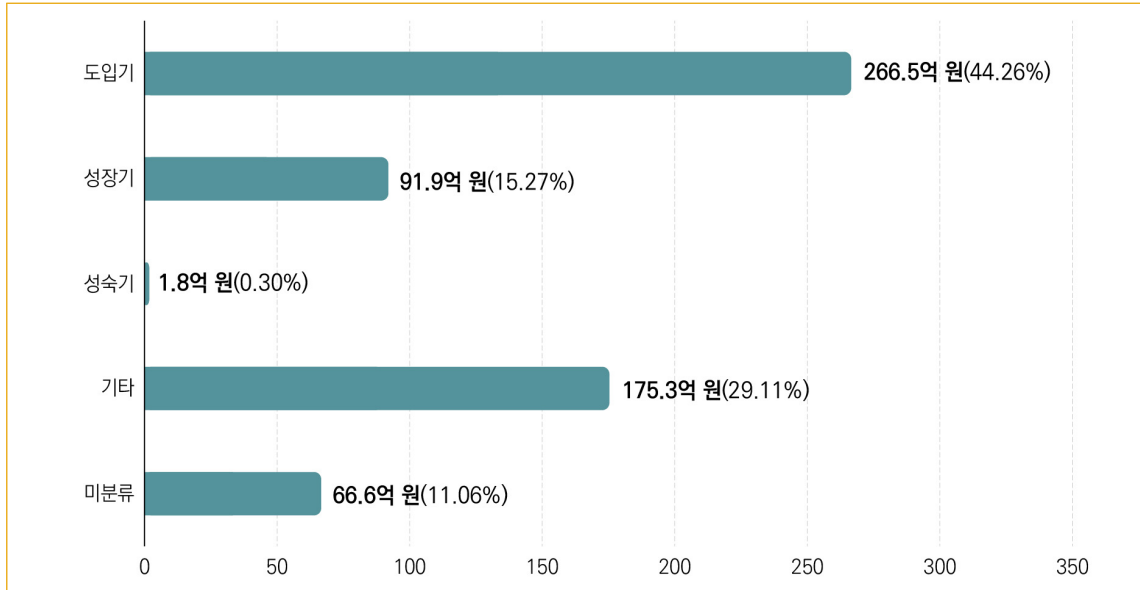


그림 6. 기술수명주기별 연구비 규모 및 비중



- **(연구분야)** 국가과학기술표준분류와 미래유망신기술분류(6T) 분석 결과, 건설/교통 분야와 생명공학 분야 중심으로 마약 검출 연구에 대한 투자가 이루어짐
  - **(국가과학기술표준분류 분석 결과)** 건설/교통 분야에 대한 연구비 비중이 28.56%(172억 원)로 가장 큰 것으로 나타났는데, 이는 화물 내 마약 검출 기술 개발의 필요성 증대로 관련 연구가 다수 이루어지고 있기 때문인 것으로 판단됨
    - ※ 연구책임자가 최대 3개까지 지정한 국가과학기술표준분류의 대분류에 대한 각 가중치를 고려한 결과임
  - 국가과학기술표준분류의 분야 중 2개 분야와 3개 분야에 해당하는 과제의 연구비 비중은 각각 22.89%(137.8억 원), 18.12%(109.1억 원)로 마약 검출 관련 전체 연구비의 41.01%(246.9억 원)이 융합과제에 투자된 것으로 드러남
    - ※ 융합과제란 연구책임자가 지정한 국가과학기술표준분류의 대분류가 두 개 이상의 분류에 해당하는 과제를 의미함
  - **(미래유망신기술분류(6T) 결과)** 생명공학(BT) 분야 연구에 대한 연구비 비중이 29.57%(178.1억 원)로 가장 큰 것으로 나타났고, 나노기술(NT) 분야 연구에 대한 연구비 비중(21.75%, 131억 원)이 그 다음으로 큰 것으로 드러남

그림 7. 국가과학기술표준분류별 연구비 규모 및 비중

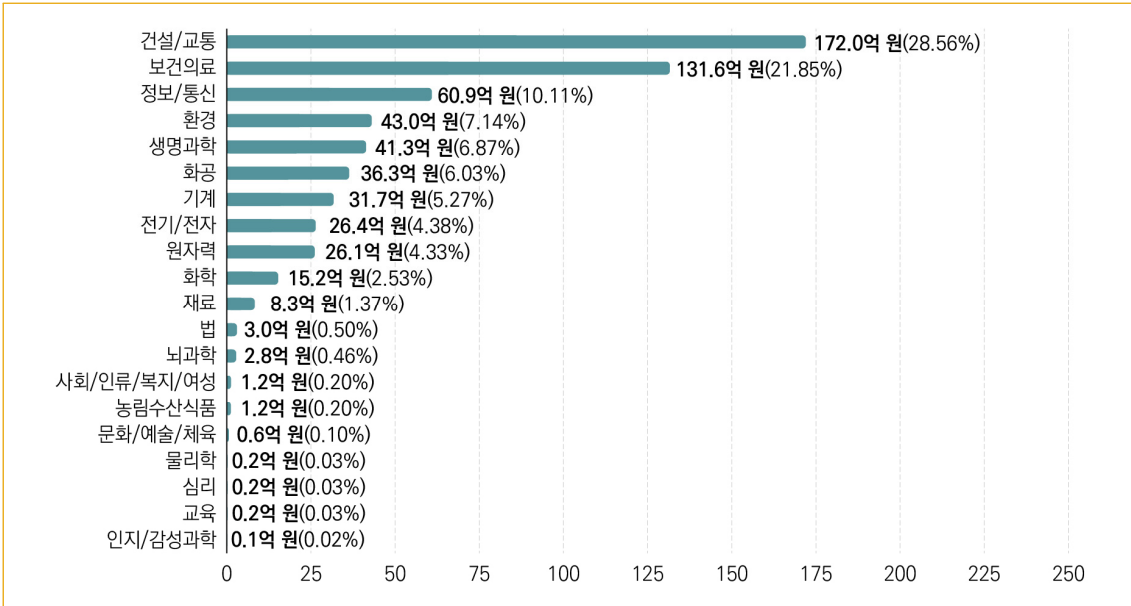


그림 8. 융합 R&D 과제 연구비 규모 및 비중

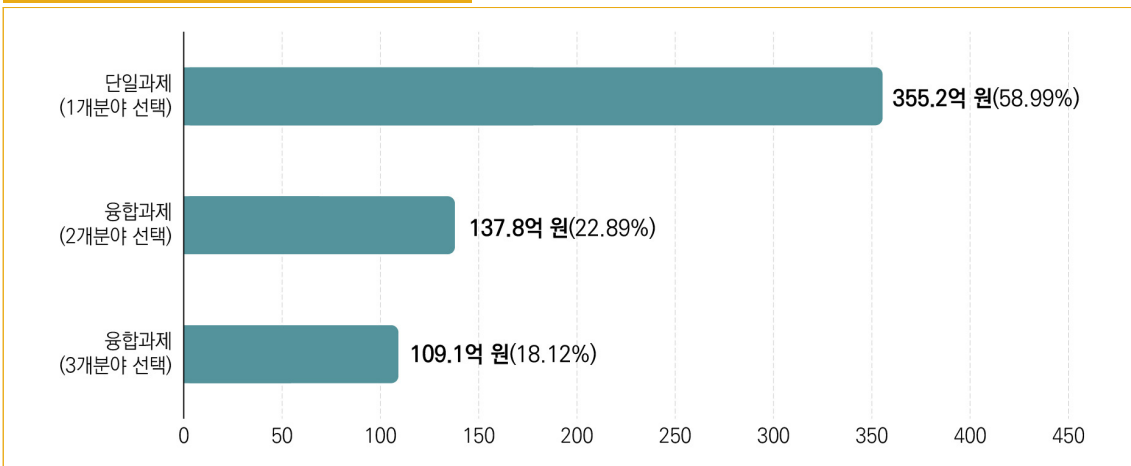
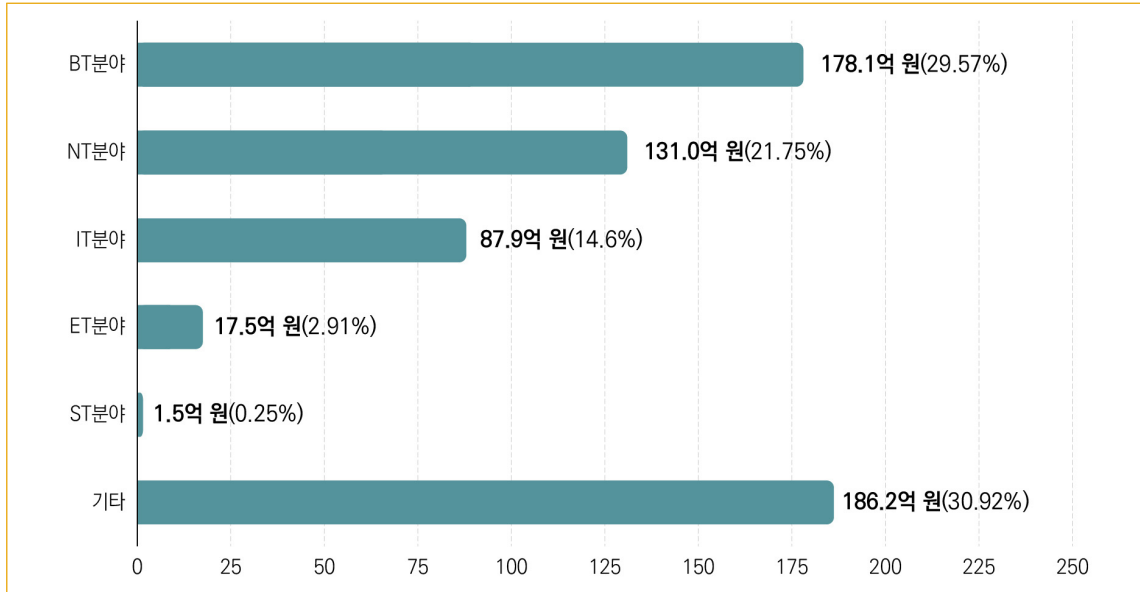


그림 9. 미래유망 신기술분류(6T)별 연구비 규모 및 비중



■ (주요 과제) 원고의 주요 내용 및 키워드 등을 기준으로 선정함

과제명 (사업명, 부처명)	수행기관, 총 연구기간, 연구비 규모	과제 주요 내용
<b>마약류 검출을 위한 나노소재 합성 및 기능화 기술 개발</b> (약물이용 범죄 사전예방을 위한 휴대용 신속탐지 기술 개발, 과학기술정보통신부 및 경찰청)	한국재료연구원, 2020-2022년, 5.5억 원('22)	<ul style="list-style-type: none"> <li>전문 수사관용 고감도 마약류 진단 장비 제작을 위한 다공성 나노소재 기반 마약류 포집소재 개발</li> <li>고감도 마약류 진단 장비 제작을 위한 센서 핵심소재 기능화 기술 개발</li> </ul>
<b>MEMS 기반 GC 요소기술 및 칼럼형 광이온화 센서기술 개발</b> (나노·소재기술개발, 과학기술정보통신부)	홍익대학교, 2017-2022년, 2.7억 원('20)	<ul style="list-style-type: none"> <li>기상 마약류 유해물질 실시간 현장 검출을 위한 흡착/분리 역할을 하는 전처리 농축기와 MEMS 기반 고감도 칼럼형 센서 원천기술 개발</li> </ul>
<b>유연재질 미세유체장치 기반의 땀수집-표면증강라만 센서 연구개발</b> (개인기초연구, 과학기술정보통신부)	육군사관학교, 2022-2024년, 0.43억 원('22)	<ul style="list-style-type: none"> <li>피부 위에 부착하는 유연한 소재로 만든 웨어러블 미세유체 기기와 그 안에 포함된 표면증강라만(SERS) 기반의 땀 속 특정 바이오마커를 검출할 수 있는 바이오센서 메커니즘의 연구개발</li> </ul>

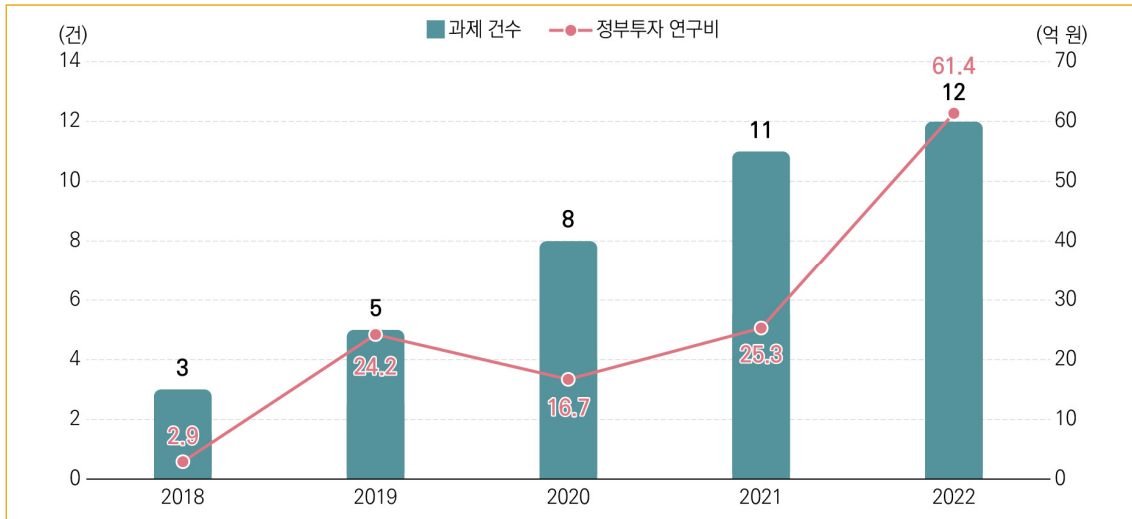
## II. 변조 음성 탐지 기술

■ **(총괄)** 최근 5년간('18~'22) 총 39건의 과제에 대해 130.5억 원의 연구비가 투자됨

※ 국가과학기술지식정보서비스(NTIS) 플랫폼을 기반으로 관련 국가 연구개발 과제 분석 수행 : 원고의 핵심 키워드를 고려하여 검색 실시

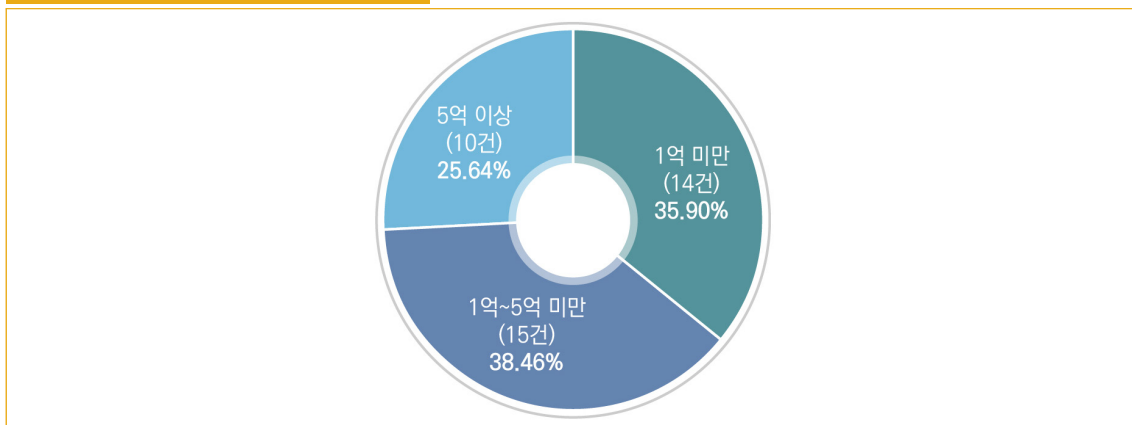
\* 음성 AND (합성 OR 변조 OR 위조 OR 스푸핑) AND (탐지 OR 검출)

그림 1. 연도별 연구과제 건수 및 연구비



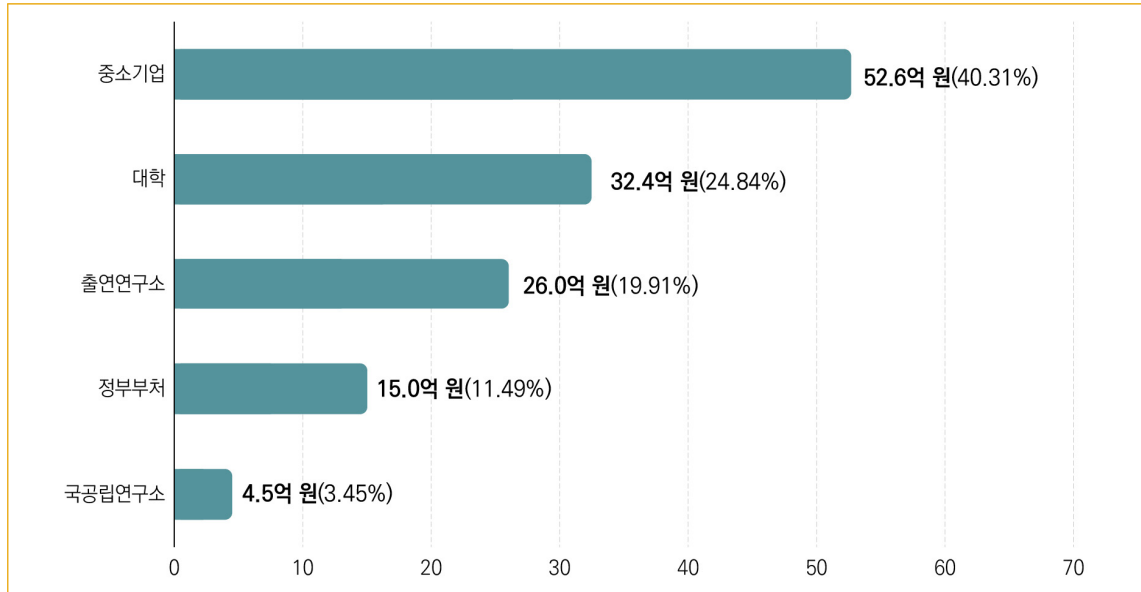
■ **(연구비 규모별 과제 수)** 연구비가 1억 원 이상 5억 원 미만인 과제의 비중이 38.46%(15건)로 가장 큰 비중을 차지하고, 1억 원 미만의 과제와 5억 원 이상의 과제는 각각 35.90%(14건), 25.64%(10건)를 차지하는 것으로 확인됨

그림 2. 연구비 규모별 과제 수 및 비중



- **(연구수행주체)** 연구수행주체별로 살펴보면, 중소기업(40.31%, 52.6억 원), 대학(24.84%, 32.4억 원), 출연연구소(19.91%, 26억 원), 정부부처(11.49%, 15억 원) 순으로 연구비 비중이 큰 것으로 드러남

그림 3. 연구수행주체별 연구비 규모 및 비중



- **(연구수준)** 연구수준 분석 결과, 음성 변조 탐지 기술 관련 연구는 도입기이며 개발연구 단계인 것으로 나타남
  - **(연구개발단계 분석 결과)** 개발연구의 연구비가 음성 변조 탐지 기술 관련 전체 연구비의 44.33%(57.9억 원)를 차지하고, 응용연구는 39.06%(51억 원), 기초연구는 13.16%(17.2억 원)를 차지하는 것으로 확인됨
  - **(연구개발성격 분석 결과)** 시작품개발 관련 연구의 연구비 비중이 25.8%(33.7억 원)로 제품 또는 공정개발 관련 연구비 비중인 20.63%(26.9억 원)보다 큰 것으로 나타나 추후 기술의 상용화 및 제품화 단계로 이어질 가능성이 높음
  - **(기술수명주기 분석 결과)** 도입기에 대한 연구비의 비중이 31.78%(41.5억 원)로, 성장기(30.01%, 39.2억 원), 성숙기(1.26%, 1.6억 원)에 비해 큰 것으로 나타남

그림 4. 연구개발단계별 연구비 규모 및 비중

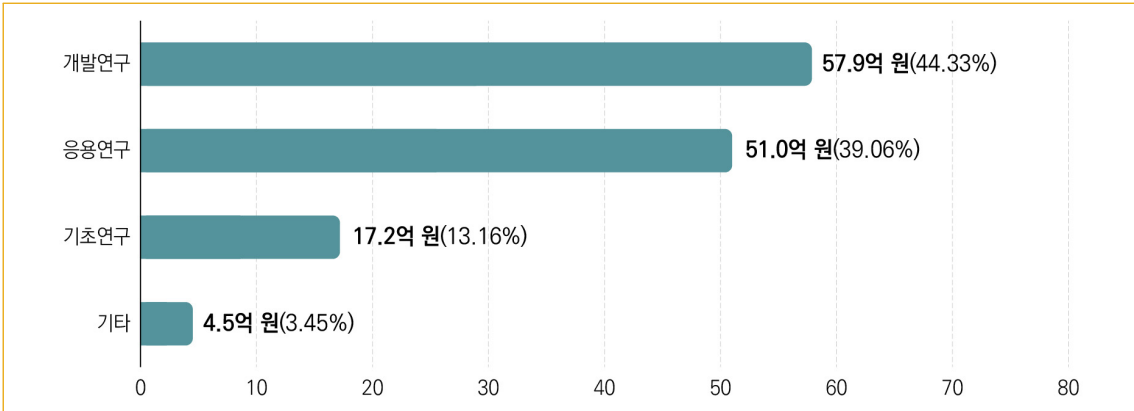


그림 5. 연구개발성격별 연구비 규모 및 비중

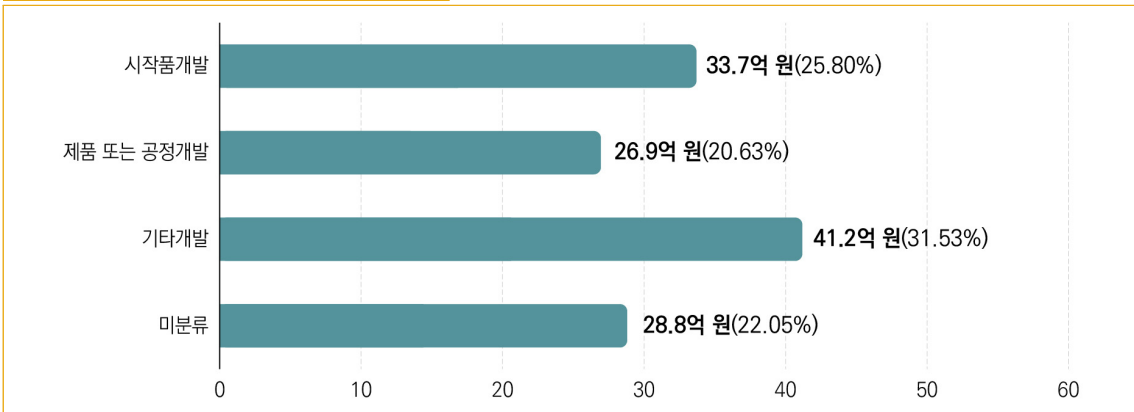
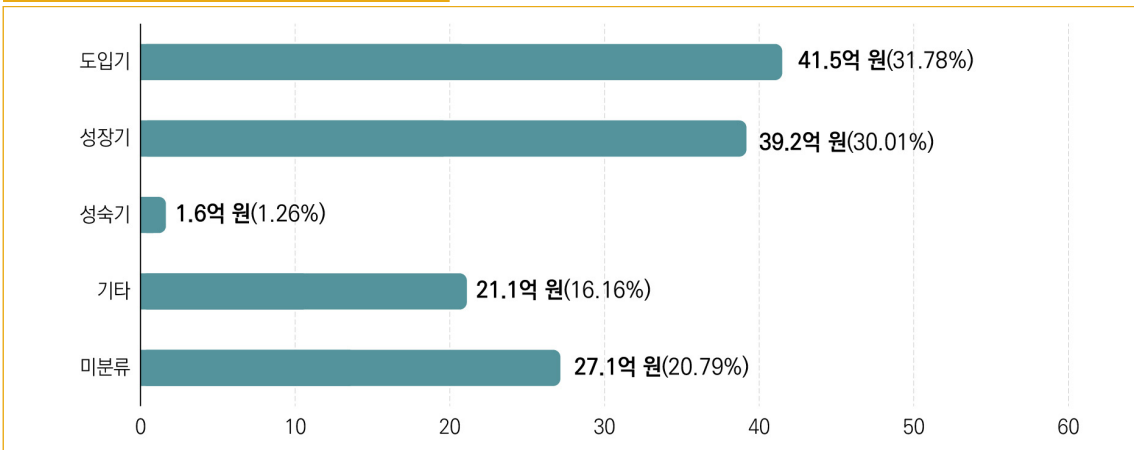


그림 6. 기술수명주기별 연구비 규모 및 비중





- **(연구분야)** 국가과학기술표준분류와 미래유망신기술분류(6T) 분석 결과, 음성 변조 탐지 기술에 대한 연구비 투자는 정보/통신 및 IT 분야를 위주로 이루어짐
  - **(국가과학기술표준분류 분석 결과)** 국가과학기술표준분류 중 정보/통신 분야에 대한 연구비 비중은 96.10%(125.5억 원)로 음성 변조 탐지 기술에 대한 전체 연구비의 대부분을 차지하는 것으로 확인됨
    - ※ 연구책임자가 최대 3개까지 지정한 국가과학기술표준분류의 대분류에 대한 각 가중치를 고려한 결과임
  - 음성 변조 탐지 기술 관련 과제 중 ‘융합과제’에 해당하는 비중은 5.34%(2개 분야 선택 비율: 4.70%, 3개 분야 선택 비율: 0.64%)이며, 6.9억 원의 연구비가 투자됨
    - ※ 융합과제란 연구책임자가 지정한 국가과학기술표준분류의 대분류가 두 개 이상의 분류에 해당하는 과제를 의미함
  - **(미래유망신기술분류(6T) 결과)** 미래유망신기술분류 기준으로 정보통신기술(IT) 분야에 대한 연구비가 99.54% (130억 원)을 대부분을 차지하는 것으로 나타남

그림 7. 국가과학기술표준분류별 연구비 규모 및 비중

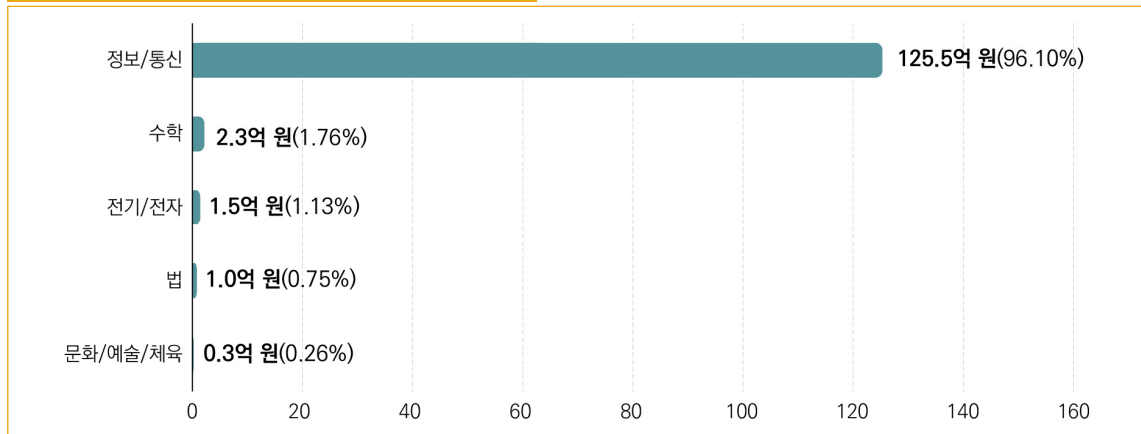


그림 8. 융합 R&D 과제 연구비 규모 및 비중

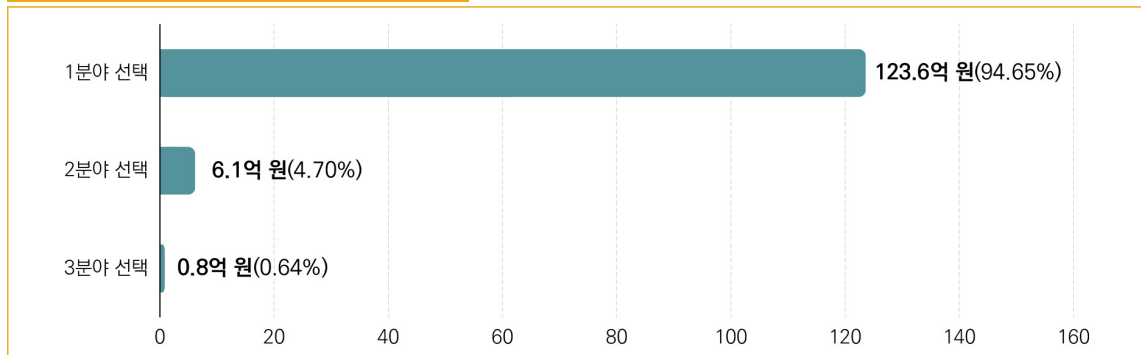
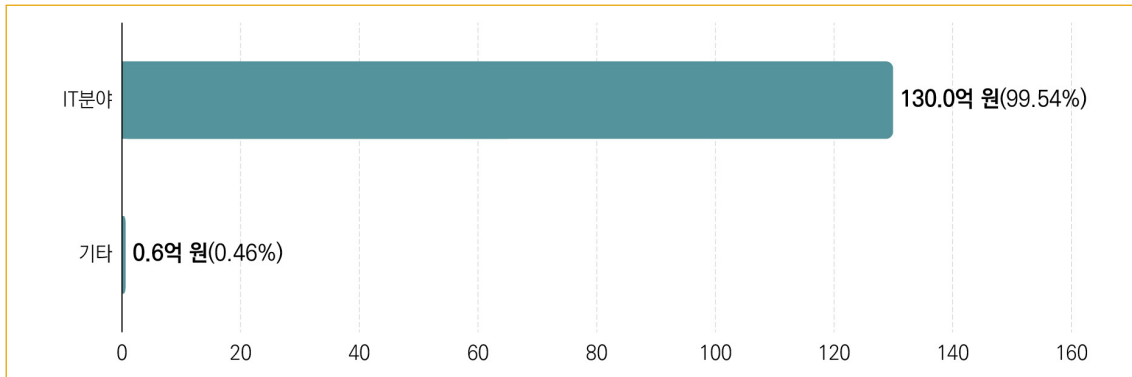
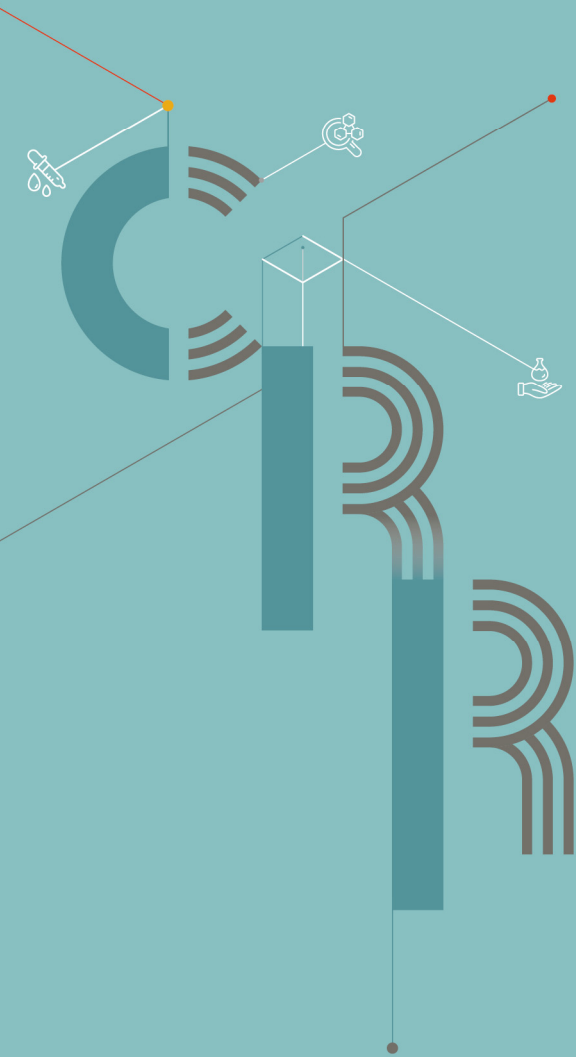


그림 9. 미래유망 신기술분류(6T)별 연구비 규모 및 비중

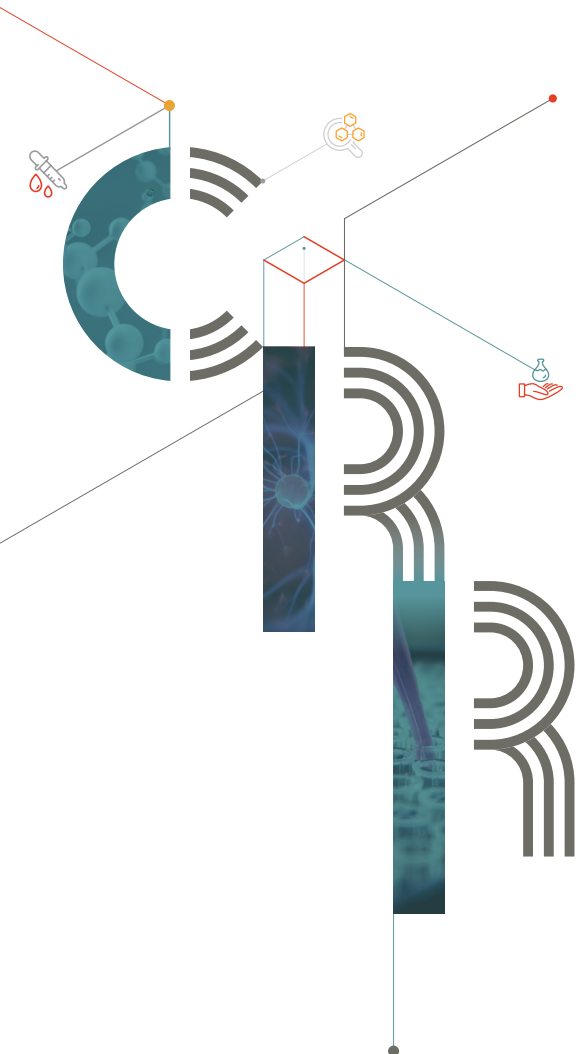


■ (주요 과제) 원고의 주요 내용 및 키워드 등을 기준으로 선정함

과제명 (사업명, 부처명)	수행기관, 총 연구기간, 연구비 규모	과제 주요 내용
인공지능 기반의 딥페이크 멀티미디어 탐지 연구 (개인기초연구, 과학기술정보통신부)	성균관대학교, 2020~2022년, 5.41억 원(*22)	• 인공지능기술을 이용하여 악의적인 목적으로 생성한 딥페이크 사진, 영상, 음성을 탐지하는 딥페이크 멀티미디어 탐지 모델 개발 및 데이터셋 구축
화자인증을 위한 음성 위조 검출 시스템 개발 (개인기초연구, 과학기술정보통신부)	서울시립대학교, 2017~2020년, 3.0억 원(*20)	• 사칭으로 인한 피해를 예방할 수 있는 화자인증 시스템 연구 및 개발
딥러닝 기반 한국어 스마트폰 녹음파일 위변조 검출 시스템 개발 (개인기초연구, 과학기술정보통신부)	송실대학교, 2021~2023년, 1.32억 원(*23)	• 딥러닝을 활용한 한국어 스마트폰 녹음일의 위변조 검출 시스템 개발



이 보고서는 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 사업임  
(NRF-2023M3C1A6043400)



# 융합연구리뷰

Convergence Research Review



02792 서울특별시 성북구 화랑로 14길 5 TEL. 02.958.4973

ISSN. 2465-8456