2023 December Vol. 9 12

융합연구리뷰

Convergence Research Review

FCSC 미래융합전략센터
Future Convergence Strategy Center

CONTENTS

융합연구리뷰 | Convergence Research Review

2023 December | Vol. 9 No. 12

발행일 2023년 12월 11일

발행인 임혜원

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

펴낸곳 공간기획 Tel. 044-863-0978

편집자 주

한국과학기술연구원 미래융합전략센터에서는 매년 과학기술정보통신부와 미래융합협의회 공동으로 '융합연구 Fellowship'공모전을 개최한다. 본 공모전은 신진연구자 및 대학원생들을 대상으로 하며, 융합연구 장려와 연구·정책·기술 아이디어 제시 기회 제공을 목적으로 한다.

본 12월호에서는 '2023년 융합연구 Fellowship'에서 선정된 최우수, 우수 그리고 특별 연구결과물을 소개한다.

금속 닷 데코레이션을 통한 p-type SnO 박막의 표면/계면 제어 방법 연구

몇 년 전까지만 해도 공상과학 영화에서만 볼 수 있었던 투명 디스플레이가 실생활에서 활용될 수 있을 정도로 진화했다. 화면이 투과도를 갖고 있어 화면의 뒷면이 보이는 특징을 갖고 있는 투명 디스플레이를 제작하기 위해서는 반도체와 같은 각종 부품이 투명해야 하는데, 이처럼 투명한 전자 소자를 만들기 위해 p-형 산화 반도체가 필수적이다. 디지털 제품들이 초소형화 및 고기능화 됨에 따라 차세대 반도체를 위한 고성능 박막의 필요성이 증대되었다.

본 호 1부에서는 박막을 낮은 온도에서 원자 수준의 얇은 두께로 균일하게 입히는 공정법인 원자층 증착법으로 형성된 금속 닷을 p-형 반도체 박막의 표면과 계면에 추가하여 전기적 반응 속도가 빠른 차세대 박막 트랜지스터를 개발하는 기술을 소개한다. 또한 금속 닷을 통한 p-형 SnO 트랜지스터 성능 향상과 그 메커니즘을 규명한다.

반도체는 '산업의 쌀'로 불릴 만큼 다양한 산업분야에서 활용되고 있다. 반도체 산업의 주권을 확보하기 위해 주요국 간의 경쟁이 치열한 가운데, 우리나라도 반도체 산업을 전략산업으로 규정하고 투자 규모를 확대하고 지원을 강화하고 있다. 제안된 본 연구가 미래 반도체 기술을 선점하는 데 기여하기를 기대해 본다.

들리는 얼굴: 이미지를 활용한 가상 인물의 목소리 생성 및 변환 인공지능

이제는 사람의 실제 목소리 없이도 애니메이션 또는 게임 속 가상 인물의 특징에 맞는 목소리를 생성할 수 있는 시대가 도래 했다. 이는 인공지능 기술의 발전에 따른 것으로, 인공지능 기술은 애니메이션 또는 게임 등과 같은 가상 세계의 콘텐츠 생성에 드는 인력과 비용을 효과적으로 줄일 수 있는 조력자 역할을 한다.

기존에는 가상 인물의 목소리를 생성하기 위해 실제로 존재하는 사람의 음성에서 음성 정보를 추출하고 변환하였으나 한계가 있었다. 본 호 2부에서는 이 문제를 해결하기 위해 얼굴 외형으로부터 음성과 상관관계가 있는 정보를 추출하여 음성을 생성 및 변화시키는 인공지능 기술에 대해 소개한다.

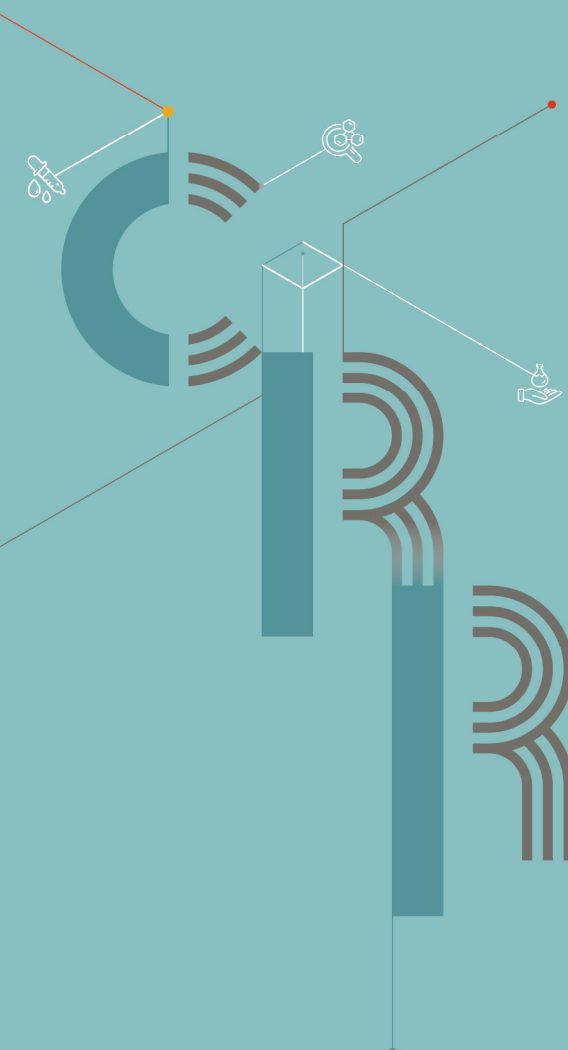
문화체육관광부에 따르면 2024년 세계 애니메이션 시장 규모는 약 73억 달러(약 8조 500억 원)에 달할 것으로 전망된다. 가상 인물이 애니메이션을 비롯하여 게임 등 여러 응용 분야에서 활용될 것으로 전망됨에 따라, 본 연구에서 소개된 가상 인물의 목소리 생성 방법으로 목소리 디자인 비용을 줄이고 생성 기간을 단축함으로써 다양한 콘텐츠 산업에 도움이 될 수 있기를 기대해 본다.

공통 데이터 모델(CDM)을 활용한 약물 치료 패턴 연구 동향

약물에 대한 부작용을 인지하지 못하거나 과다 복용할 경우에는 생명을 위협할 수 있는 심각한 상황에 처하게 되기 때문에 올바른 처방이 매우 중요하다. 최근 의료계에서는 의료 빅데이터를 활용하여 환자에게 처방된 약물의 패턴을 분석할 수 있게 되었다. 이를 통해 약물의 효과와 안전성 그리고 어떤 약물로 인해 치료 목표에 얼마나 도달했는지, 약물을 왜 중간에 변경했는지에 대한 분석이 가능해졌다.

본 호 3부에서는 2010년에 미국 정부의 지원으로 결성된 OMOP(Observational Medical Outcomes Partnership)의 공통 데이터 모델(CDM, Common Data Model)을 활용한 문헌고찰 방법으로 약물 치료 패턴을 분석하는 법을 소개한다. 아울러, 본 분석 방법을 이용하여 약물 치료 패턴 연구 현황을 분석한다.

부적절한 약물 처방 및 치료는 사회·경제적으로 큰 손실을 야기할 수 있다. 약물의 처방과 치료 패턴을 분석하는 것은 의료 현장에 적합한 치료법을 선택하고 진료 지침을 제공하는데 매우 유용하다. 본 연구에서 제안된 연구 방법이 국민의 안전을 확보하고 의료비를 절감하는데 기여하며, 의료 데이터에 대한 체계적인 분석으로 현대의학에 대한 신뢰도가 더욱 향상될 수 있기를 기대해 본다.



융합연구리뷰

Convergence Research Review

01

금속 닷 데코레이션을 통한 p-type SnO 박막의 표면/계면 제어 방법 연구

백인환(인하대학교 화학공학과 교수)

01

금속 닷 데코레이션을 통한 p-type SnO 박막의 표면/계면 제어 방법 연구

백인환(인하대학교)

I. 서론

1. 연구 개요 및 해결 과제

1.1 연구 개요

금속 닷(dot, 10nm 이하의 물질)의 선택적/부분적 증착 방법을 활용하여 p-type 반도체 박막의 계면과 표면을 제어함으로써 소자 점멸비(on-off ratio) 및 홀 이동도(hole mobility)가 큰 박막 트랜지스터를 개발하는 기술에 관한 연구이다.

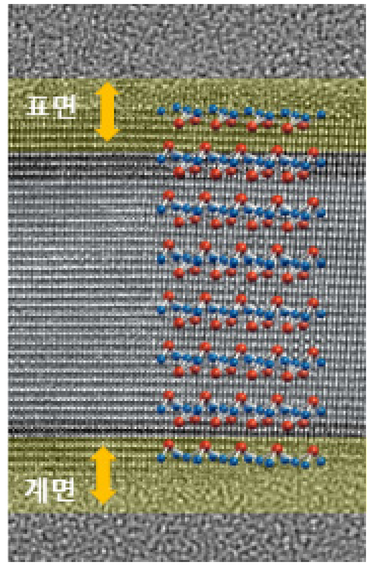
1.2 해결 과제

최근 인공지능 및 사물인터넷(IoT, Internet of Things) 기술의 발달과 클라우드 서비스가 확대됨에 따라 대용량 데이터들이 초고속으로 처리될 필요성이 증대되고 있다. 이에 따라 고성능 디지털 회로와 메모리 반도체에 대한 수요가 지수 함수적으로 증가하는 추세다. 반도체 소자 수요자들의 고성능 및 고집적도의 요구에 맞추어 공정이 미세화 및 입체화됨에 따라 트랜지스터(transistor) 채널의 폭과 길이는 줄어들고 있다. 수평 방향의 소자 크기인 채널 폭과 길이뿐만 아니라 수직 방향으로의 길이, 즉 반도체 박막의 두께 또한 제어되어야 하는 단계에 진입했다. 특히, 차세대 3D-디램과 V-낸드(Vertical NAND)플래시 메모리의 셀 트랜지스터용 반도체 박막 및 모놀리틱(Monolithic) 3-D(M3D) 구조에 적용되는 박막은 세대가 거듭할수록 두께의 제한이 불가피할 것으로 예상된다. 집적화가 심화될수록 요구 박막의 두께가 얇아질 것이며 동시에 물리적 초박막의 상태에서도 고성능을 발휘할 수 있는 재료의 필요성이 대두될 것이다.

초박막은 일반적으로 반도체 벌크(bulk, 수 μm(마이크로미터) 이상의 물질) 물질의 내재적인 특성과는 다른 성능(ex. 전자 이동도, 전하 농도, 결정화도, 결함 농도)을 보인다. 소자에 적용되는 박막이 얇아질수록

상대적으로 벌크가 차지하는 비율이 줄어들며 반도체 계면과 표면으로부터 유발되는 전기적 특성 변화가 전체 박막의 특성에 지대한 영향을 끼치게 될 것이다. 특히, 10 nm 이하 박막의 경우 계면/표면에서의 전하 산란 및 접합된 유전체와의 상호작용을 통해 변조된 전기적 특성이 소자의 전체 특성을 결정지을 수 있기 때문에 우수한 전기적 특성을 확보하기 위해서는 계면/표면에 대한 이해가 필수적일 것이다(그림1).

그림 1. 박막의 표면과 계면



* 출처: 저자 작성

차세대 트랜지스터 소자의 반도체 박막은 단차 피복성(step coverage) 및 균일도(uniformity)가 우수한 원자층 증착법(ALD, Atomic Layer Deposition)을 통해 형성되어야 한다. 원자층 증착법은 실제 반도체 산업의 핵심 반도체 공정으로써 널리 활용되고 있다. 원자층 증착법은 두 종류 이상의 화학 물질을 교반하여 주입함으로써 옹스트롬(Ångström, 길이의 단위로 10^{-10} 미터 또는 0.1nm를 나타냄) 단위로 박막을 증착할 수 있는 기술이다. 이 때, 각각의 화학 물질들은 열 분해되지 않는 이상 자기 제한적(self-limiting)인 특성을 유지할 수 있기 때문에 리간드(ligand) 교환(착화합물의 중심 원자에 배위되어 있는 원자 또는 원자단이 화학적 특성이 같은 다른 배위자와 교환하는 반응) 메커니즘을 통해 상대 화학 물질과만 반응하는 특징이 있다. 이러한 자기 제한적 특성으로 인해 단차 피복성이 큰 3차원 기판에도 균일한 박막을 형성할 수 있는 것이다.

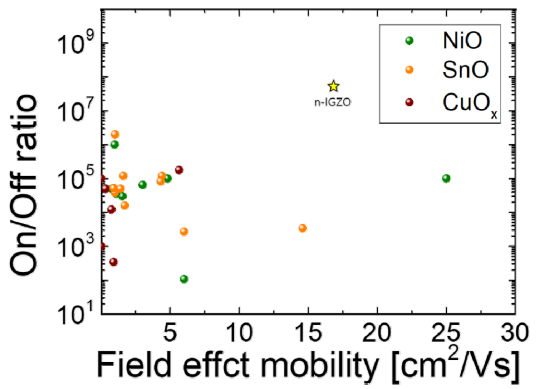
일반적인 원자층 증착 공정에서, 기판 표면의 작용기와 화학 물질의 초기 상호작용에 따라 흡착 방식이 결정된다. 흡착 화학 물질의 밀도나 방향은 결정학적 핵(nucleus)의 형태를 결정하는 몇 가지 요소 중 하나이다. 아울러 발생된 핵에 따라 전체 박막 벌크부의 결정학적 배향, 결정립 크기, 거칠기가 영향을 받는다고 알려져 있다. 즉, 원자층 증착법에서 초기 화학 물질의 흡착 방식이 공정의 최종 형성물인 박막의 형태를 정할 수 있다는 것을 의미한다. 상기 제시된 결정학적 배향, 결정립 크기, 거칠기 등의 박막 특성 관련 인자들은 모두 박막의 전기적인 특성에 지대한 영향을 미친다. 따라서, 초기 계면 상태 제어를 통한 ALD 박막 성장 거동 변화의 연구를 통해 둘의 상관관계를 밝혀내는 것이 매우 중요하다고 볼 수 있다. 이때, 초기 계면 상태 제어를 위해 본 연구에서는 금속 닷의 선택적/부분적 증착 방법을 활용하고 해당 기술을 통해 변화된 박막의 전기적 특성을 관찰하였다.

새로운 고성능 초박막 개발의 중요성이 대두된 이후 벌써 몇 년의 시간이 흘렀다. 따라서 차세대 반도체 박막으로 활용될 수 있는 신물질들은 이미 대부분 검토되었으며 실제 합성 결과물이 다수 문헌으로 보고된 단계다. 그러나 신물질 연구에 비해 기 개발된 물질들의 계면/표면 제어를 통해 박막의 특성을 원하는 방향으로 조절할 수 있는 연구는 아직 부족한 상태라는 점에 본 연구는 주목하였다.

지금까지 보고된 p-type 산화물 반도체 박막은 n-type 산화물 반도체 박막에 비해 보고된 물질의 종류가

적고 전기적 특성 또한 낮은데, 이는 정공(hole) 생성과 관련된 결함 준위 형성이 내재적으로 어려우며 정공(hole)의 이동 경로인 가전자대(valence band)가 이방성(anisotropy, 방향에 따라 물리적으로 다른 성질을 보임)을 갖기 때문이다. 〈그림 2〉를 참고하면, IGZO(InGaZnO, In, Ga, Zn의 산화물)로 대표되는 n-type 반도체 박막 트랜지스터 성능과 지금까지 보고된 p-type 박막 트랜지스터들의 성능을 비교해볼 수 있다. 트랜지스터로 문제없이 동작할 수 있도록 100,000 이상의 점멸비를 가지는 p-type 박막 트랜지스터(TFT, Thin Film Transistor)들 중 최대 이동도는 현재까지 보고된 문헌에 따르면 약 6 cm^2/Vs 수준으로, 전자 이동도 30 cm^2/Vs 이상의 n-type 반도체에 비해 매우 부족한 수준이다.

그림 2. p-SnO, p-CuOx, p-NiO 박막 트랜지스터의 성능 비교표



* 출처: 최민기 외(2023)

따라서, p-type 반도성 박막을 차세대 소자에 차용하되 한 단계의 공정 과정만을 계면/표면에 추가하여 n-type의 성능을 따라잡을 만큼 전기적 성능을 향상시킬 수 있는 계면/표면 제어 기술 개발이 필요하다고 여겨진다. 본 연구에서는 p-type SnO(산화주석) 박막 상부에 금속 닷을 증착함으로써 계면/표면 제어 공정을 수행하였다. 금속 닷을 통한 계면/표면 제어 기술은 메모리 반도체의 셀 트랜지스터, M3D구조의 디지털 회로뿐만 아니라 디스플레이용 박막 트랜지스터(TFT), 박막형 포토 센서, 및 가스 센서의 반도체 박막에도 적용될 수 있는 범용적인 기술로 활용될 수 있다. 원자층 증착법으로 형성된 금속 닷을 이용, 계면과 표면에 데코레이션하여 전기적 특성을 효과적으로 제어하고자 본 연구를 제안하였다.

2. p-type 산화물 박막 트랜지스터

IGZO(InGaZnO)로 대표되는 n형 산화물 반도체가 2010년대부터 디스플레이 패널(panel)에 적용되어 활발하게 양산되고 있는 중이며, 또한 반도체 DRAM 트랜지스터로 검토되고 있는 반면, p-type 산화물 반도체는 산업계에 발도 붙이지 못한 상태이다. p-type 산화물 반도체는 재료의 내재적 한계로 인해 높은 성능을 갖는 소자 제작이 매우 어렵다는 특징이 있다. p형 산화물 반도체가 n형 산화물 반도체와 견줄만한 성능을 가지게 된다면, 고성능 박막 CMOS(Complementary Metal-Oxide-Semiconductor, 금속 산화막 반도체로 반도체 소자의 일종) 논리 회로를 구현하는 것이 가능해질 것이다. 결과적으로 고성능 CMOS 박막 기술을 활용해 디스플레이 패널 설계가 용이해지고 전력 소비를 효과적으로 줄일 수 있을 것이다. 또한 산화물 CMOS 시스템은 메모리·비메모리 반도체 소자의 집적도 한계를 직접 돌파하지 않고도 해결할 수 있는 우회 기술로도 응용될 수 있다. 해당 우회 기술은 수평 방향의 공정 미세화(scaling down, 나노미터(nm) 단위로 반도체 칩 회로 선폭을 줄여 공정을 미세화하는 작업) 패러다임에서 벗어나 소자를 수직 방향으로 쌓아 올리는 방식을 사용한다. 이는 모놀리식 3D(Monolithic 3D) integration 이라 불리며 현재 산업계 및 학계에서 많은 주목을 받고 있다.

3. 원자층 증착법

원자층 증착(ALD, Atomic Layer Deposition)은 박막 두께와 조성을 원자 수준에서 제어할 수 있는 매우 정교한 기술이다. 특히 현대 반도체 산업에서 DRAM 캐피시터(capacitor), 낸드 플래시(NAND flash)의 셀 스트링(cell string), 및 로직 게이트 옥사이드(logic gate oxide) 등의 핵심 전자 소자를 형성하기 위한 미세 공정에 필수적으로 쓰이고 있다. 원자층 증착의 핵심 원리는 자기 제한적(self-limiting) 반응에 있다. 이 공정은 같은 종류의 케미컬이 아무리 오랫동안 기판 표면에 노출되더라도 한 원자층의 화학 물질이 화학 흡착(chemisorption, 화학적 결합에 의한 흡착으로 전자의 이동을 수반) 방식으로 흡착되고 난 이후에는 더 이상의 반응 또는 흡착이 일어나지 않는 것이다. 이 공정을 피딩(feeding) 또는 펄스(pulse)로 명명한다. 한 케미컬의 피딩(feeding)이 끝나면 질소나 아르곤 등의 불활성 가스를 사용하여 퍼지(purge)를 진행한다. 퍼지(purge)는 화학 흡착된(chemisorbed) 화학 물질을 제외한 나머지 물리 흡착된(physisorbed) 화학 물질들을 모두 챔버(chamber) 밖으로 배출시키는 시퀀스(sequence)이다. 첫 번째 퍼지(purge)가 끝난 이후 타겟의 표면은 단 한 층의 화학 흡착된(chemisorbed) 화학 물질로 포화된 상태가 된다. 이후 두 번째 화학 물질이 주입되면 단 한 층의 화학 흡착된(chemisorbed) 화학 물질과 표면 반응을 시작한다. 먼저 주입된 화학 물질이 단 한 층으로 이미 한정되어있으므로 두 번째 화학 물질을 아무리 많이 주입하더라도 반응량은 증가하지 않을 것이다. 추가적으로 두 번째 화학 물질 또한 열분해 되지 않는 이상 같은 화학 물질끼리 흡착되지 않는 특성을 보인다. 즉, 원자층 증착법에 사용되는 모든 화학 물질은 자기 제한적 반응을 갖는다. 두 번째 화학 물질이 첫 번째로 흡착된 화학 물질과 모두 반응하고 나면 두 번째 화학 물질로 표면이

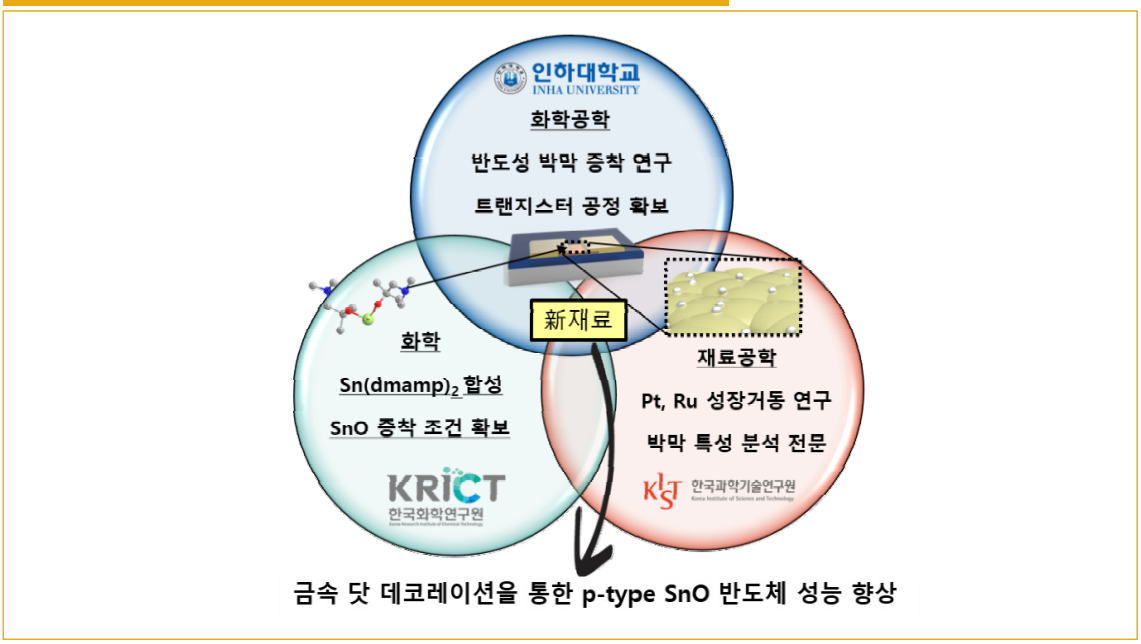
모두 포화될 것이다. 이 때, 두 번째 퍼지(purge) 공정을 진행하면 물리 흡착된(physisorbed) 두 번째 화학 물질이 모두 챔버(chamber) 밖으로 배출되고 화학 흡착된(chemisorbed) 화학 물질만 남게 된다. 이 과정이 원자층 증착의 한 사이클인 것이다. 따라서 원자층 증착법에서 박막 증착 속도를 이야기할 때 growth rate(성장률)이라는 표현보다 growth per cycle(주기당 성장)이라는 표현이 더욱 많이 통용된다. 원자층 증착은 위에 제시된 반도체 산업 뿐 아니라 배터리, 태양전지, 연료전지 소재의 개질 및 개선을 위해서도 응용되고 있다. 파우더 시편에도 원자층 단위로 증착이 가능하기 때문에 촉매 분야에서도 최근 많은 주목을 받고 있다. 원자층 증착법의 연구 개발 및 적용 분야는 계속해서 확장 중이며 기존 재료의 튜닝 및 최적화에도 쉽게 적용될 수 있는 무궁무진한 가능성을 보인다.

II. 연구 방법론

1. 융합연구 방법

〈그림 3〉은 금속 닷 데코레이션을 통해 성능이 개선된 SnO 반도체 박막이라는 신(新)소재를 합성하기 위한 융합연구 모식도를 나타낸다. 각 기관의 화학적, 화학공학적, 재료공학적 지식을 충분히 활용할 수 있게 역할을 분담하였다.

그림 3. 화학, 화학공학, 재료공학의 전문성을 살려 진행한 융합연구 모식도

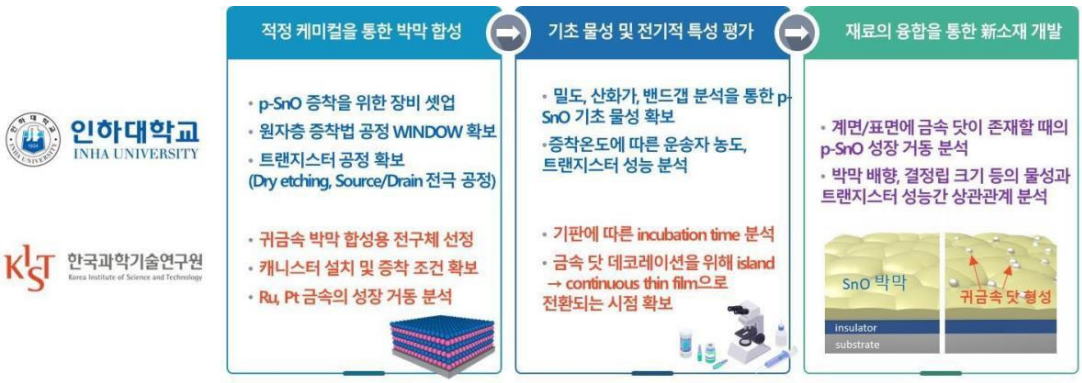


본 연구의 p-type SnO의 원자층 증착을 위해 한국화학연구원에서 개발한 비(非) 상용 주석(Sn) 전구체를 활용 하였다. 해당 Sn 전구체는 기존 상용 전구체와는 달리 Sn의 산화가(oxidation state)가 +2가로 존재한다. 이러한 특징으로 인해 +4가보다 준 안정적이라 볼 수 있는 +2가의 산화가를 갖는 산화제1주석(SnO) 박막을 제조할 수 있게 된다. SnO 재료는 n형 산화제2주석(SnO_2) 재료와는 달리 다수 캐리어가 정공(hole)인 p-type 반도체이다. +4가의 상용 전구체를 사용한다면 모두 SnO_2 박막이 합성된다는 것을 선행 연구 결과를 통해 알 수 있었기 때문에 본 연구에서는 성공적인 p-type 박막 합성을 위해 한국화학연구원에서 공급받은 전구체만을 사용하였다.

공급받은 +2가 전구체는 $Sn(dmamp)_2$ (bis(1-dimethylamino-2-methyl-2-propoxide)Sn)로 물과 반응하여 SnO 박막을 형성한다. SnO 박막 합성을 위한 원자층 증착 공정은 한국화학연구원 및 인하대학교 화학공학과에서 함께 확보하였다. 인하대학교 화학공학과에서는 박막 공정에 더하여 반도체 공정을 진행하여 박막 트랜지스터 소자를 제조하였다. 트랜지스터 제조 조건이 확보된 이후 금속 닷 데코레이션 연구를 한국과학기술연구원과 협업하여 진행하였다. 금속 닷 데코레이션은 유전막과 반도체 막 사이에 닷이 삽입되어 계면이 제어되는 경우와 반도체 막 표면에 증착되어 표면 특성이 제어되는 두 가지 경우로 나눌 수 있다. 본 연구에서는 두 경우에 대해 모두 연구를 진행하였으며, 금속 닷이 박막의 물성 및 박막 트랜지스터의 전기적 성능에 미치는 영향에 대해 분석하였다.

SnO 계면과 표면 제어를 위해 금속 닷을 형성하는 공정을 진행한 프로세스는 〈그림 4〉에 더욱 자세히 나타나 있다. 인하대학교 화학공학과의 원자층 증착 공정 기술과 한국과학기술연구원 전자재료연구센터의 금속재료 기술을 융합하기 전에 각 담당한 재료에 대한 물성 분석을 자세하게 진행하였다. SnO 박막의 경우 밀도, 산화가, 밴드갭(band gap, 하나의 전자가 그 결합된 상태로부터 벗어나는데 필요한 최소량의 에너지) 등의 기초 물성과 금속 닷이 없을 때의 전기적 성능에 대해 분석되었다. 금속 닷의 경우 면저항 측정을 통해 박막이 연속적으로 형성되는 시점을 판단할 수 있었다. 금속이 SnO 표면이나 계면에 형성될 때 연속적인 박막을 이루면 '닷' 형태가 아니라 금속 '박막'이 형성되는 것이므로 데코레이션으로 볼 수 없을 것이다. 즉, 데코레이션을 위해서는 연속적으로 금속이 형성되지 않는 사이클과 기판에서 증착이 일어나기 시작하는 인큐베이션 사이클(incubation cycle) 확보가 요구되며 실제 루테늄(Ru), 백금(Pt) 금속 닷의 사이클 범위를 한국과학기술연구원에서 확보하였다.

그림 4. SnO의 계면/표면을 제어하기 위해 진행된 연구 과정

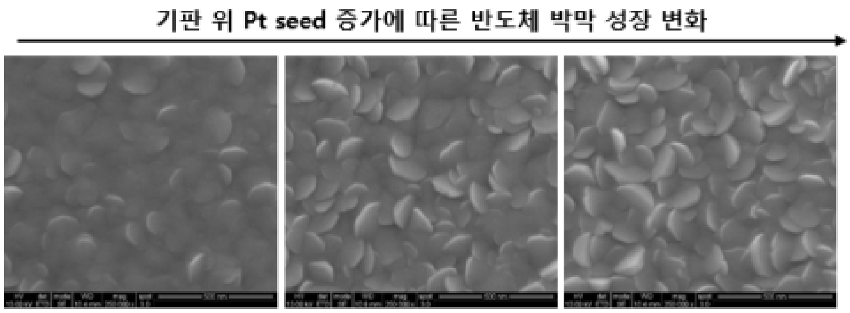


2. 성능 향상 가설

표면 작용기에 따른 금속 닷의 성장 거동을 파악하고 유전체의 표면(최종적으로 유전체와 반도체의 계면)에 형성된 금속 닷이 반도체 박막의 결정학적 배향 및 결정립 크기 등을 제어하는 메커니즘에 대한 연구를 수행하였다. 금속 물질은 표면에너지가 크기 때문에 원자층 증착법으로 성장시킬 때 본질적으로 아일랜드 성장(island growth, 기판 위에 한 덩어리의 성장이 동시다발 적으로 여러 군데에서 일어나고, 이것들이 커지면서 서로 합쳐져 박막이 형성되는 형태)을 하게 된다. 연속적인 박막을 형성하지 않고 초기 아일랜드(island) 단계에서 공정을 멈추어 부분적으로만 금속 닷을 형성함으로써 소스와 드레인(drain)이 전기적으로 연결되지 않고 후속 반도체 박막의 성장 거동 및 반도체의 전기적 특성에만 영향을 끼칠 수 있도록 하는 것이 중요하다.

금속 닷이 형성된 유전체는 표면에너지가 국부적으로 달라진다. 특히, 백금속 금속 닷을 사용한 경우에는 표면에너지뿐만 아니라 촉매효과가 더해져서 원자층 증착법을 통한 반도체 박막의 성장 거동이 크게 변화할 것으로 예상된다. 그 예로, 선행연구인 〈그림5〉는 p-type SnO의 기판 상 백금(Pt) 금속 닷 양에 따른 성장 배향 변화를 보여준다. 성장 축이 바뀌며 성장 속도 변화 또한 달라지는 것이 관찰되었다. 백금 씨앗(Pt seed) 증가에 따라 수직 배향성이 커지는데 이 때 p-type 박막의 물성이 변화하여 상온에서의 이산화질소 가스 센싱 감도가 더욱 높아지는 결과를 얻은 바 있다. 금속 닷은 이러한 배향 변화뿐만 아니라 결정립 크기, 성장 속도 변화를 야기하는 것을 관찰할 수 있었다.

그림 5. p-type SnO의 배향 변화

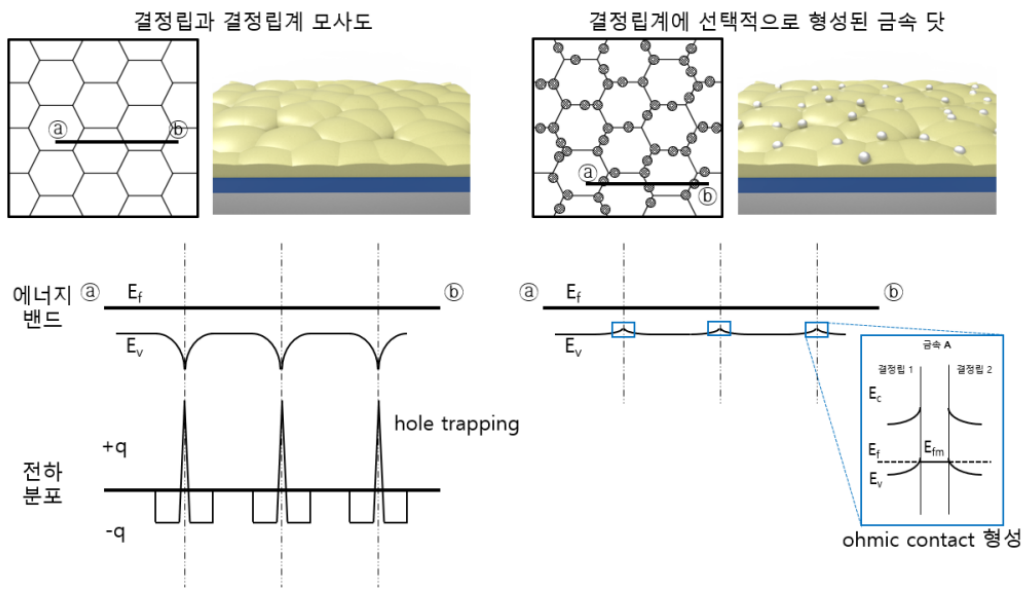


* 출처: 저자 작성

본 연구에서 금속 닷을 활용하여 얻고자 하는 것은 결정립계에 선택적으로 합성된 금속 닷을 통해 결정립계의 운송자 산란을 억제하여 이동도를 개선하는 것이다. 예상되는 '소자 성능 개선 가설'을 〈그림 6〉에 나타내었다. 〈그림 6〉 왼쪽 그림의 a지점과 b지점 사이에는 3개의 결정립계가 위치하며 각 결정립계에 정공이 트랩(trap) 되어있는 상태가 모사되어 있다. 결정립계에 트랩 된 정공으로 인해 결정립계에는 +q의 매우 높은 전하분포를 갖게 되며 결정립계 근처에는 공간 전하로 인한 −q 전하를 갖는 공핍층(depletion layer)이 형성된다. 이러한 공간 전하는 〈그림 6〉에 모사된 바와 같이, 에너지밴드를 휘어지게 만들 것이다. 결정질 박막의 p-type 반도체의 이동도를 개선하려면 해당 결정립계에서의 운송자 산란을 줄여야 한다. 본 연구는 확보된 금속 닷 결정립계 선택적 증착 기술로 운송자 산란을 감소시키는 것을 목표로 한다.

만약 p-type 반도체보다 일함수가 큰 백금(Pt) 또는 이리듐(Ir)의 귀금속을 결정립계에 증착할 경우 해당 M-S 접촉은 저항(ohmic) 접촉을 형성하게 될 것이다. 즉, 〈그림 6〉의 오른쪽처럼 결정립계에 선택적으로 일함수가 높은 금속 닷이 위치하면 정공 트랩 및 공핍층으로 인한 에너지 장벽 형성을 방지하게 되며 저항(ohmic) 접촉을 유도할 수 있다. 결과적으로 결정립에서의 운송자 산란으로 인한 정공의 이동도 열화(degradation)를 억제함으로써 이동도를 증가시킬 수 있을 것으로 예상된다. 결정립계는 에너지적으로 불안정하기 때문에 결정립 표면보다 더욱 먼저 선택적으로 금속 닷이 형성될 것이라는 사실은 쉽게 예측할 수 있다. 이때, 과 증착된 메탈로 인해 소스-드레인(source-drain)이 전기적으로 연결되지 않도록 금속 닷 원자층 증착 공정의 사이클 수를 면밀히 제어하여 반도체 박막 특성을 유지하는 것이 중요하다.

그림 6. 결정립계에 선택적으로 증착된 금속 닷을 통한 운송자 산란 억제 메커니즘



* 출처: 저자 작성

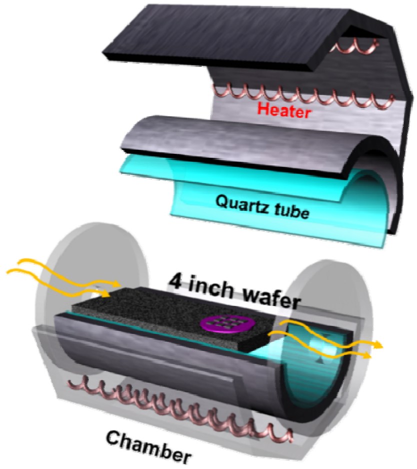
위와는 반대로 p-type SnO 반도체보다 일함수가 작은 금속 닷(Ti, Al 등)이 결정립계에 선택적으로 형성될 경우 공핍층이 더욱 커질 것을 예상할 수 있다. 이때 공핍층의 에너지 장벽으로 인한 정공 이동도 감소 폭은 더욱 커지게 되는데, 결과적으로 트랜지스터 적용이 불리해질 것이다. 하지만, 이를 가스 센서에 적용할 경우, 표면의 정공 축적층은 크기가 상대적으로 감소하기 때문에 오히려 더 높은 감도를 가지면서도 검출한계가 매우 작은 고성능의 가스 센서를 제작할 수 있을 것으로 예상된다. 즉 소자의 종류에 따라 적절한 일함수의 금속 닷을 선택함으로써 해당 소자의 성능을 극대화할 수 있는 기술이라고 볼 수 있다. 하지만 본 연구에서는 일함수가 높은 귀금속 닷을 적용하여 트랜지스터 소자의 전기적 특성을 향상하는 것을 우선적인 목표로 한다.

2. 실험 방법

SnO 박막 증착을 위해 실험실에서 직접 제작한 원자층 증착 장비를 사용하였다(〈그림 7〉 왼쪽 참고). 해당 장비는 상온부터 300℃까지 온도를 제어할 수 있어서 한국화학연구원에서 합성한 2가 Sn 전구체의 원자층 증착 가능 윈도우를 파악하기에 적절한 장비이다. 또한 퍼니스 형태(furnace type)로 제작해 챔버 내 온도가 매우 균일하게 제어된다는 장점을 갖는다(〈그림 7〉 오른쪽 참고). Sn 전구체가 들어있는 캐니스터를 장착한 뒤 충분한 증기압이 챔버 내로 공급될 수 있도록 52℃로 캐니스터(canister)를 가열하였다. 이 때 카운터

리액턴트(counter reactant)로는 물이 사용되었는데, 물의 높은 증기압을 낮추기 위해 쿨링 자켓을 사용하여 7℃로 냉각하였다. 확보된 SnO 원자층 증착의 피딩/퍼지(feeding/purge) 레시피는 Sn 전구체(2초) - 퍼지 (10초) - 물 (5초) - 퍼지 (40초)이다. 주로 사용된 SnO 박막은 210℃에서 증착되었다. Ru와 Pt 금속닷 또한 원자층 증착법으로 합성되었지만, 오염을 방지하기 위해 상기 장비와는 다른 장비를 활용하였다. 각각의 증착 온도는 SnO 박막이 손상되지 않는 온도인 150℃내지 200℃를 선정하였으며, 실제 30 사이클(cycles) 이내로 증착하여 아일랜드(island) 형태의 성장(growth)만 일어나도록 유도하였다.

그림 7. (좌) Lab-made 원자층 증착 장비, (우) Furnace type의 원자층 증착장비 챔버 내부 및 Flow 모식도



* 출처: 저자 작성

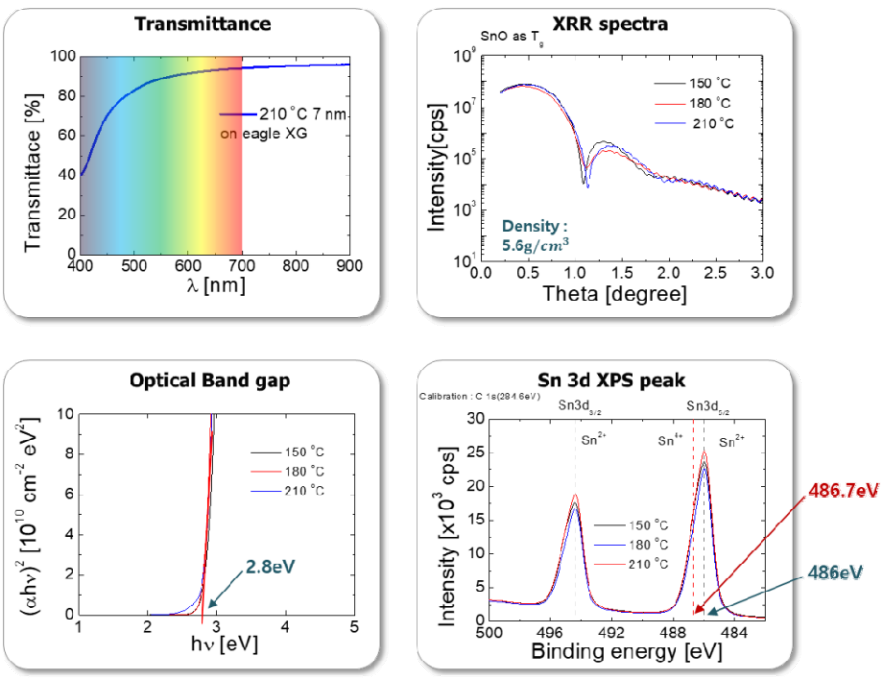
트랜지스터 소자의 게이트(gate)는 heavily doped Si(강하게 도핑된 실리콘)을 사용하였으며 유전막으로 활용하기 위해 SiO_2를 건식 산화(dry oxidation) 공정으로 100 nm 두께로 합성하였다. 해당 기판에 증착된 SnO 박막의 패터닝을 위해 GXR601의 양성 감광제(positive photo-resist)와 건식 식각(dry etching) 공정이 사용되었다. 소스/드레인 전극의 패터닝은 리프트 오프(lift-off, 희생 재료를 사용하여 표면에 미세 구조를 만드는 방법) 공정으로 진행했기 때문에 음성 감광제(negative photo-resist)인 AZ5214를 사용했고, 전극 물질로는 Ni/Au bilayer를 도입하였다. SnO 박막의 두께는 8 nm였으며 트랜지스터의 폭 및 길이(width/length)는 각각 300/50 ㎛이다.

III. 연구 내용

1. 기확보된 연구 결과

SnO 박막의 기초 물성에 대해 자외선-가시광선(UV-Vis), X-선 반사율(XRR, X-ray Reflectivity), 엘립소메트리(Ellipsometer, 물질의 표면에 광이 반사할 때 편광 상태의 변화(입사와 반사)를 관측하고, 그로부터 물질에 관한 정보를 구하는 방법), 엑스선 광전자 분광법(XPS, X-ray Photoelectron Spectroscopy) 장비를 활용하여 분석하였다. UV-Vis와 Ellipsometer를 통해 분석된 투과도(transmittance)와 광학적 밴드갭(optical band gap)을 통해 해당 SnO 박막은 2.8 eV의 optical band gap을 갖는 것을 알 수 있었다. 이는 n형 IGZO(밴드갭 3.3 eV)와 집적되어 투명 전자 CMOS 소자를 이룰 수 있다는 특성을 나타낸다. XRR분석 결과, 밀도는 약 5.6 g/cm^3 (이론 밀도 6.4 g/cm^3)으로 피팅되어 기존 보고된 SnO 박막의 밀도 값과 유사함을 확인하였다. 추가로 XPS를 통해 Sn 3d core electron의 binding energy를 분석한 결과 Sn $3d_{5/2}$ peak가 486 eV (C 1s C-C bond 284.8 eV calibrated) 에서 검출되는 것을 확인하였다. 이는 Sn 원자가 분석된 박막 내에서 +2의 산화 상태(oxidation state)를 갖는 것을 의미한다. +4가의 산화 상태(oxidation state)를 의미하는 486.7 eV는 관찰되지 않았으므로 phase-pure한(상순도가 높은) SnO가 적절히 합성되었다고 볼 수 있다.

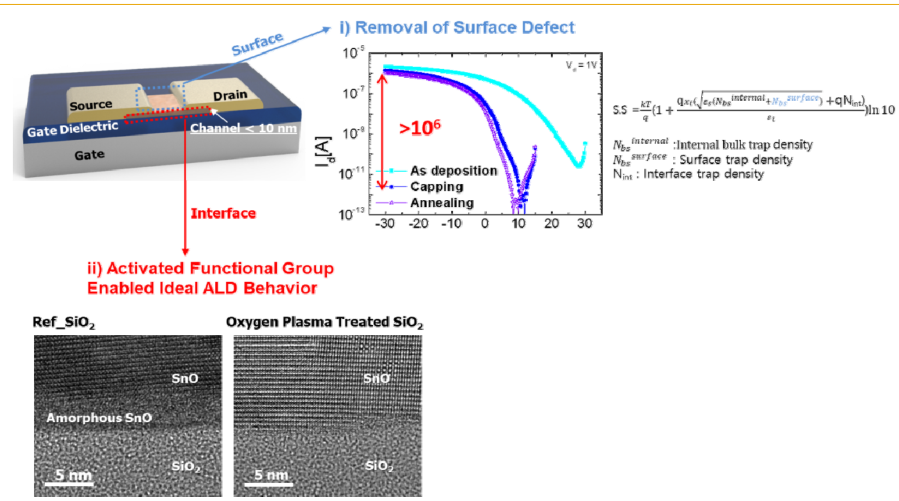
그림 8. 원자층 증착법으로 합성된 SnO 박막의 기초 물성(UV-Vis, XRR, Ellipsometry, XPS)



* 출처: 저자 작성

SnO 박막 트랜지스터(thin film transistor)에 사용되는 SnO 박막의 두께는 10 nm를 넘어가면 정공의 농도가 지나치게 높아져 turn-off가 제어되지 않는다는 문제점이 있었다. 또한 앞서 언급했듯 스케일링을 돌파하기 위한 우회기술인 Monolithic 3D integration을 위해서도 두께가 제한적이어야 한다는 점을 들어 SnO 두께는 10 nm 이내로 제어되는 것이 유리하다. 하지만 이러한 극박막 형태의 박막 트랜지스터를 제조할 경우, 표면과 계면이 차지하는 박막에서 차지하는 비율이 높아지며, 표면과 계면 특성이 전체 소자의 전기적인 성능에 영향을 끼치기 쉬워지는 것을 예상할 수 있다. 실제 선행 연구 결과를 통해 해당 사실을 증명할 수 있었다. 〈그림 9〉에는 SnO 박막의 표면과 계면을 각각 제어한 결과에 대해 나타나 있다. 우선 SnO 박막의 표면에는 표면 결함(surface defect)이 높은 농도로 존재한다. 알루미늄 옥사이드 박막을 통한 패시베이션(passivation)과 추가적인 열처리 공정을 통해 표면 결함(surface defect) 농도를 크게 낮출 수 있었으며 이는 개선된 문턱전압 이하 스윙(subthreshold swing) 값(문턱 전압 밑에서 게이트 전압에 따라 얼마나 빠르게 전류가 증가하는지를 나타낸 값)을 통해 유추할 수 있었다. 박막의 두께가 얇든지 두껍든지 간에 상관없이 표면 제어를 통해 문턱전압 이하 스윙(subthreshold swing) 값이 매우 작아진다는 사실은 벌크 부 대비 표면에 고농도로 존재하는 결함(defect)이 확실히 제어되었음을 의미한다. 두 번째로 SnO 박막과 유전체 사이 계면의 제어를 통해서도 향상된 전기적 성능을 얻을 수 있었다(Baek et al, 2021). SnO 박막과 유전체의 계면을 제어한다는 것은 유전체의 표면을 제어하는 것과 같다. 해당 선행 연구에서는 유전체 표면에 산소 플라즈마를 처리한 뒤 SnO 원자층 증착법을 진행하는 방식으로 계면 특성을 향상시키는 것을 꾀하였다. 〈그림 9〉의 투과전자현미경(TEM, Transmission Electron Microscope) 이미지에서 산소 플라즈마를 통해 크게 개선된 계면을 확인할 수 있다. 기존 산소 플라즈마 처리를 하지 않은 기판에서는 비교적 두꺼운 비정질 SnO 계면성분 및 배향이 틀어진 SnO 박막이 얻어진 반면, 산소 플라즈마 처리를 진행한 기판에서는 비정질 SnO 계면성분 없이 깔끔한 결정 구조가 얻어진 것을 볼 수 있다. 전계효과 이동도 또한 30% 증가한 결과를 얻을 수 있었다.

그림 9. p-SnO 박막 트랜지스터의 표면 및 계면 제어 결과

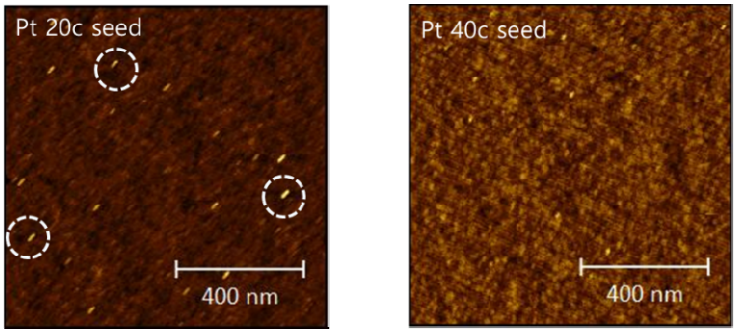


* 출처: 저자 작성

2. 금속닷을 사용한 기판 표면(유전체-반도체 계면)처리가 계면 특성에 미치는 영향 연구

금속 닷을 사용하여 유전체 표면을 우선 데코레이션하기 위해 아일랜드 성장(island growth)이 실제로 일어났는지 확인해볼 필요가 있었다. 〈그림 11〉은 유전체 대표로 SiO_2기판을 사용하여 백금(Pt)을 증착하고 원자간힘현미경(AFM, Atomic Force Microsope)로 분석한 결과를 보여준다. 20 cycle이 증착된 경우 기판 중간 중간에 섬(island) 또는 씨앗(seed) 형태로 백금(Pt) 닷(dot)이 위치하는 것을 볼 수 있다. 반면 40 cycle이 진행된 경우 비교적 넓은 범위에 모두백금 백금(Pt) 닷(dot)이 위치하고 있는 것을 확인할 수 있었다. 즉 40cycle은 아일랜드 성장(island growth)이 거의 끝나고 continuous한 금속 박막이 얻어지기 시작하는 시점이라고 여겨진다.

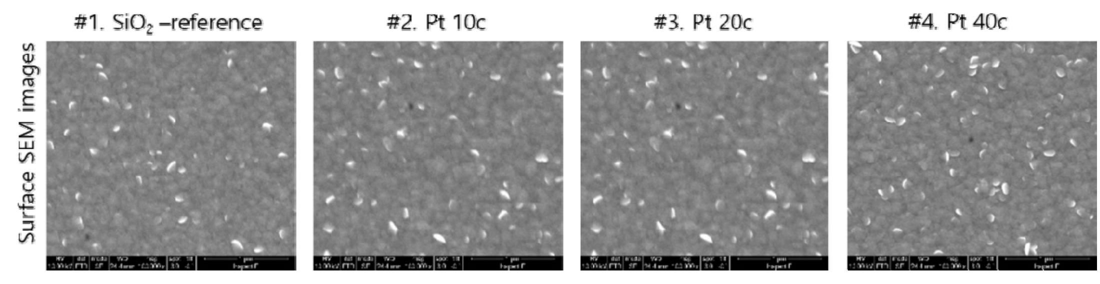
그림 10. SiO2 기판에 각각 20 cycles, 40 cycles 증착된 Pt metal dot의 AFM 이미지



* 출처: 저자 작성

〈그림 11〉은 SiO_2 기판에 백금 씨앗(Pt seed (island))를 증착한 직후 SnO 박막을 형성 한 샘플의 상부를 주사전자현미경(SEM, Scanning Electron Microscope)으로 분석한 것이다. SnO 상부에 백금 씨앗(Pt seed)를 증착해서 결정립계에 선택적 증착을 관찰하는 것이 초기 가설 수립 시의 실험 계획이었지만, 더 빠른 기간 안에 결과를 얻을 수 있는 백금 씨앗(Pt seed) 상 SnO를 증착하는 실험을 우선 수행하였다. 백금 씨앗(Pt seed)을 형성하지 않은 참고(reference) 박막 대비 씨앗(seed)을 형성했을 때 눈에 띄는 변화는 관찰되지 않았다. 이는 〈그림 5〉의 선행 연구 결과와 상반되는데, 씨앗(seed)의 양 자체가 다르기 때문이다. 〈그림 5〉에는 스퍼터링(sputtering, 건식방식을 사용해 얇은 박막을 코팅하는 공정)으로 백금(Pt)을 형성하여 거의 연속적인 백금(Pt) 박막 상에 SnO가 증착된 것이나 마찬가지이므로 배향에 매우 큰 차이가 보였음을 유추할 수 있다. 반면 〈그림 11〉은 아일랜드(island) 형태의 백금(Pt) 닷(dot)이 형성된 상부에 SnO가 증착된 것이기 때문에 참고(reference) 시편 대비 급격한 배향 차이가 관찰되지는 않은 것으로 여겨진다.

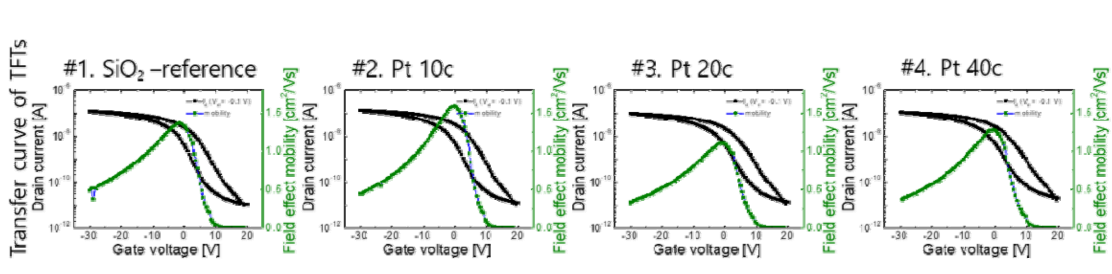
그림 11. SiO2 기판, Pt 10 cycles, 20 cycles, 40 cycles 가 증착된 SiO2 기판에 증착된 SnO morphology 의 SEM images



* 출처: 저자 작성

실제 SnO 결정의 경우 van der Waals면의 수직 방향으로 정공의 유효질량이 가장 작기 때문에 정공의 이동도가 가장 빠르다고 알려져 있다(Hu et al, 2019). 그러나 〈그림 11〉에서 관찰했듯이 백금 씨앗(Pt seed)의 양에 따른 박막 배향 차이가 거의 없었으므로 배향의 차이로부터 오는 SnO 트랜지스터(transistor)의 전기적 성능 변화는 무시할만한 수준이라고 가정한다. 따라서 〈그림 12〉에서 관찰된 백금 씨앗(Pt seed)의 양에 따른 SnO 트랜지스터(transistor)의 이동도 변화는 백금 씨앗(Pt seed)으로부터 야기된 것이라고 볼 수 있다. 특히 10 cycle을 통해 백금 씨앗(Pt seed)을 형성하고 SnO를 성장시킨 트랜지스터(transistor)의 경우 전계 이동도(field effect mobility)가 1.36 cm^2/Vs 에서 1.60 cm^2/Vs으로 약 18 % 증가한 결과를 보였다. 이는 앞서 제시된 성능 향상 가설과 부합하는 연구 결과라고 볼 수 있다. 하지만 백금 씨앗(Pt seed)이 이보다 더 많아지는 경우, 즉 20 cycle이상 수행한 뒤 SnO를 증착하여 트랜지스터(transistor)를 제작했을 때는 전계 이동도가 오히려 감소하는 결과가 관찰되었다. 10 cycle의 백금(Pt)은 실제 결정립계에서 정공의 산란을 억제했지만 백금(Pt) 양이 더욱 증가할 때는 상대적으로 SnO의 계면 접촉 면적이 줄어드는 것으로 보인다. 이에 따라 계면에 형성되는 정공 축적층(hole accumulation layer)의 유효 폭(width)을 감소시켜 전계 이동도가 감소한 것으로 예상된다.

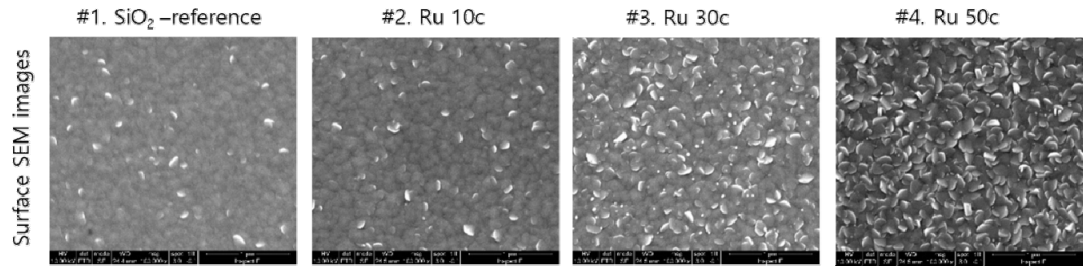
그림 12. Pt seed상 증착된 SnO 박막을 통해 제작된 트랜지스터의 Transfer curve



* 출처: 저자 작성

〈그림 13〉에는 〈그림 11〉에서 확인했던 것과 같은 방법으로 SiO_2 기판에 루테늄 씨앗(Ru seed(island))를 증착한 직후 SnO 박막을 형성 한 샘플의 상부 SEM 이미지이다. 참고(reference) 시편 대비 루테늄(Ru) 양이 증가할수록 SnO의 배향이 횡 방향(lateral direction)에서 종 방향(vertical direction)으로 변화하며 점차 하얗게 관찰되는 에지(edge)의 비율이 늘어나는 것을 알 수 있다. 이는 〈그림 5〉에서 관찰한 바 있듯, 루테늄(Ru)의 밀도(density)가 너무 높아서 아일랜드(island)가 기판 상 차지하는 비율이 매우 크거나 루테늄 씨앗(Ru seed) 자체가 백금 씨앗(Pt seed) 대비 SnO의 성장 방향(growth direction)을 더욱 원활하게 제어할 수 있다는 것을 의미한다.

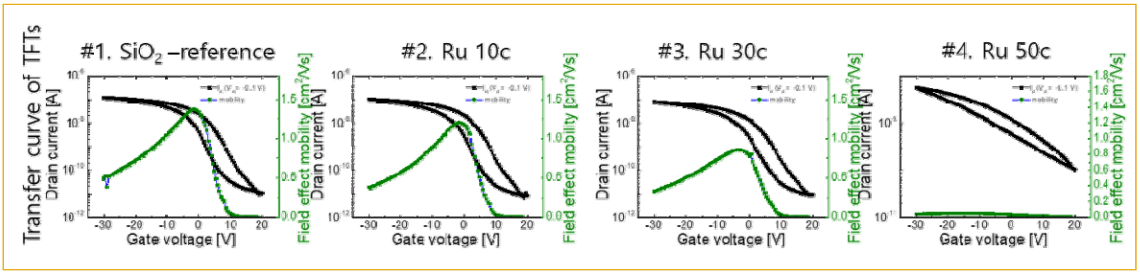
그림 13. SiO2 기판, Ru 10 cycles, 30 cycles, 50 cycles 가 증착된 SiO2 기판에 증착된 SnO morphology 의 SEM images



* 출처: 저자 작성

〈그림 14〉의 전달 곡선(transfer curve)을 참고해보면 루테늄(Ru) 씨앗(seed)이 10 cycle만 되어도 너무 과하게 증착되었기 때문에 전계효과 이동도의 감소가 관찰되는 것을 알 수 있다. 50 cycle에서는 SnO의 누설 전류(off-current)가 높아지며 스위칭(switching) 특성이 매우 열화되는 것이 관찰되는데 이는 〈그림 13〉의 SEM 이미지에서 SnO의 배향이 틀어지는 것에서 확인했듯 기판상 루테늄(Ru) 밀도(density)가 너무 높아서 정공 축적층의 유효 면적이 감소했기 때문으로 볼 수 있다. 두 번째 원인으로는 배향이 틀어지고 표면 거칠기(roughness)가 증가함에서 오는 전기적 특성의 열화를 제시할 수 있다. 루테늄(Ru) 씨앗(seed)을 통해서는 해당 연구의 가설인 결정립계에서의 정공 산란 감소를 관찰하지 못했다고 결론지을 수 있다.

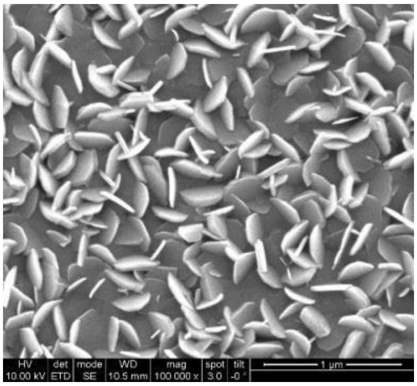
그림 14. Ru seed상 증착된 SnO 박막을 통해 제작된 트랜지스터의 Transfer curve



* 출처: 저자 작성

증착 온도를 기존 210℃에서 더 높여 270℃로 설정 후 루테늄(Ru) 씨앗(seed) 위에 증착하여 모폴로지(morphology, 박막 형태 및 특성)를 관찰해보았다. 이 때도 물론 초기 가설에서 의도한 대로 루테늄(Ru) 씨앗(seed)을 통해 정공 이동도가 증가하지는 않았지만 증착온도 210℃인 〈그림 13〉에서 관찰되었던 것보다 SnO 결정 배향이 매우 크게 변화하는 것이 관찰되었다(〈그림 15〉 참고). 이는 증착 온도가 높아짐에 따라 SnO crystal의 성장 중 흡착된 분자(adsorbed molecules)의 확산(diffusion)과 이동(migration)이 더욱 활발해졌기 때문에 얻어진 결과로 볼 수 있다. 수직으로 배향된 SnO는 직관적으로 알 수 있듯 계면 부에 접한 결정립계의 비율을 높이게 되어 정공 산란을 더 크게 유도할 것이다. 즉 트랜지스터 응용을 위해서는 수직 배향된 박막은 적절하지 않다. 하지만 해당 박막의 응용 타겟을 가스 센서로 변경한다면 단점을 오히려 장점으로 승화할 수 있게 된다. 가스 센서의 경우 상대적으로 불안정한 에지(edge)가 많이 드러날수록, 비표면적이 클수록 감도가 높아진다. 따라서 같은 양을 증착했을 때 수직으로 성된 박막이 가스 센서에는 오히려 유리할 수 있는 것이다. 이는 금속 닷을 통한 계면 제어 전략을 해당 연구의 범위(scope)에서 더 확장 시켜 또 다른 응용 분야인 고감도 센서 개발로까지 이어 나갈 수 있는 특성으로 여겨진다.

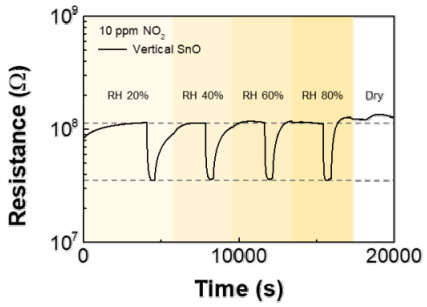
그림 15. Ru seed 50 cycles 상 270℃의 상대적 고온에서 증착된 SnO 박막의 SEM image



* 출처: 저자 작성

SnO의 NO_2 가스 센싱 특성을 관찰함으로써 유해가스 센싱용 채널 물질로의 적용 가능성을 평가하였다(〈그림16〉 참고). 상온에서 측정되었음에도 불구, 10 ppm의 미량 NO_2 가스에도 half-order 이상의 저항 변화를 보였다. 이것을 통해 200℃ 이상의 고온에서만 동작하는 n-type 산화물 기반 센싱 물질을 대체할 수 있는 저전력 p-type SnO 채널의 활용 가능성을 살펴볼 수 있었다. 더욱 특이한 점은 기존 보고되었던 센싱 물질과는 다르게 상대 습도(relative humidity)의 변화에도 불구, 항상 일정한 감도를 보인다는 것이다. 이는 계절 및 시간 등 실제 사용 환경에 제한 없이 고정밀도를 갖는 가스 센서를 제작할 수 있는 특성이라고 볼 수 있다. 따라서 p형 SnO 채널의 가스 센서 응용 또한 매우 가치 있는 연구 및 개발이 될 것으로 예상한다.

그림 16. 수직 성장한 SnO 박막을 통해 제작한 가스 센서의 상대 습도별 NO_2 가스 노출에 따른 저항 변화 특성

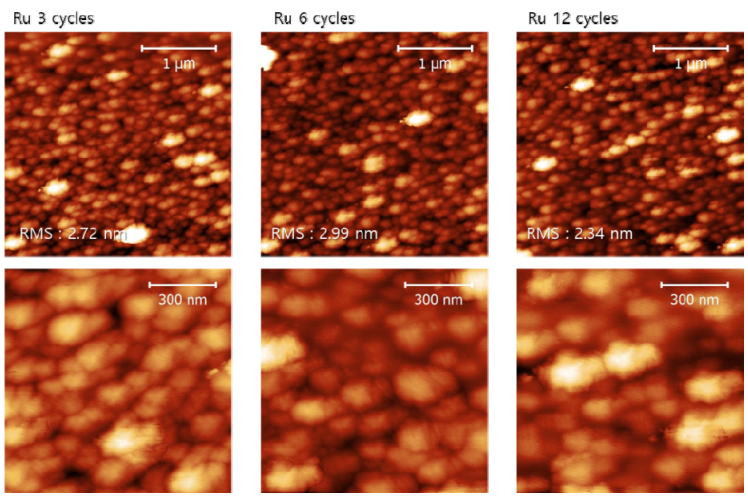


* 출처: 저자 작성

3. 금속 닷을 사용한 p형 SnO 박막 포면 특성 연구

이전 장에서는 금속 닷이 유전체와 SnO 반도체 박막 사이에 위치할 때, 즉 계면에서의 효과를 관찰하였다. 본 장에서는 금속 닷이 SnO 반도체 박막 상부(표면)에 위치할 때의 SnO 성능 변화를 관찰한 결과에 대해 다룬다. 〈그림 17〉은 SnO 박막 상부에 루테늄(Ru) 금속을 165℃의 합성 온도에서 원자층 증착을 진행한 시편의 표면 원자간힘현미경(AFM, Atomic Force Microscopy) 이미지를 보여준다. 루테늄(Ru) 금속 합성을 위해 RuO_4 전구체와 H_2 상대 반응물(counter reactant)이 사용되었다. 각각 3 cycle, 6 cycle, 12 cycle이 진행된 결과인데 눈에 띄는 루테늄(Ru) 씨앗(seed)은 전혀 관찰되지 않았다. 또한 cycle 수에 따른 모폴로지(morphology)의 변화도 관찰되지 않았다. 실제 루테늄(Ru)이 증착된 것인지 물리적으로 확인할 수 없었으므로 화학적 분석의 일종인 엑스선 광전자 분광법(XPS) 분석을 추가로 진행하였다.

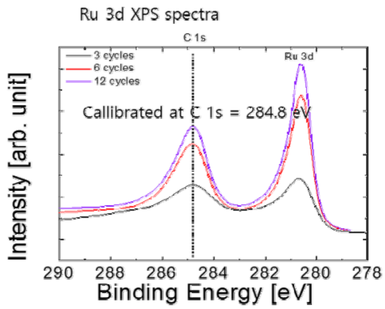
그림 17. SnO 박막 표면에 Ru의 원자층 증착 이후 AFM images (증착온도:165℃)



* 출처: 저자 작성

〈그림 18〉은 〈그림 17〉의 AFM에서 분석된 시편들의 XPS spectra를 보여준다. 특히 Ru 3d peak과 C 1s peak을 함께 플롯할 수 있도록 278 ~ 290 eV의 Binding energy 범위를 나타내었다. 모든 peak은 C 1s의 C-C bond위치인 284.8 eV로 캘리브레이션(calibration)되었다. Ru 3d peak의 화학적 이동(chemical shift)은 관찰되지 않아 산화가는 일정한 것을 알 수 있었고, 원자층 증착 사이클 수가 커짐에 따라 Ru 3d peak의 면적 또한 함께 커짐을 확인할 수 있었다. 이는 루테늄(Ru) 금속의 원자층 증착은 실제 잘 일어났음을 의미한다. 하지만 추가적으로 Sn 3d peak을 분석해본 결과 또 다른 문제가 발견되었다.

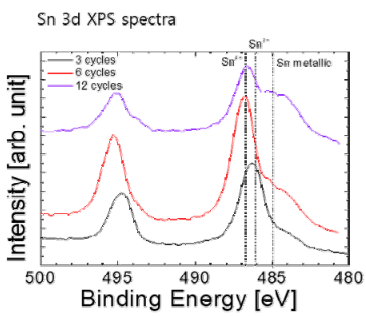
그림 18. SnO 박막 표면에 Ru 원자층 증착 이후 XPS 분석 결과(Ru 3d)



* 출처: 저자 작성

루테늄(Ru) cycle이 증가할수록 SnO 박막 내 Sn의 산화가가 +2에서 점차 +4와 0으로 변화하고 있는 것이 관찰된 것이다〈(그림 19) 참고〉. 이는 루테늄(Ru) 원자층 증착 사이클을 거듭할수록 SnO가 Sn과 SnO_2로 분리되는 불균등화 반응(disproportionation)이 일어났기 때문이다. 불균등화 반응은 화학적으로 reactive한 RuO_4가 SnO 박막의 표면에 흡착된 후 SnO를 산화시켰기 때문에 발생한 것으로 보인다. 수소의 SnO에 대한 영향은 배제할 수 있었는데 최근 SnO가 수소에 노출되어도 환원되지 않는다는 연구 결과를 확보했기 때문이다(data not shown). 따라서 RuO_4 전구체를 통한 Ru 금속 닷 증착 및 SnO 박막의 표면 개질은 본 연구에 부적절한 것으로 판명되었다.

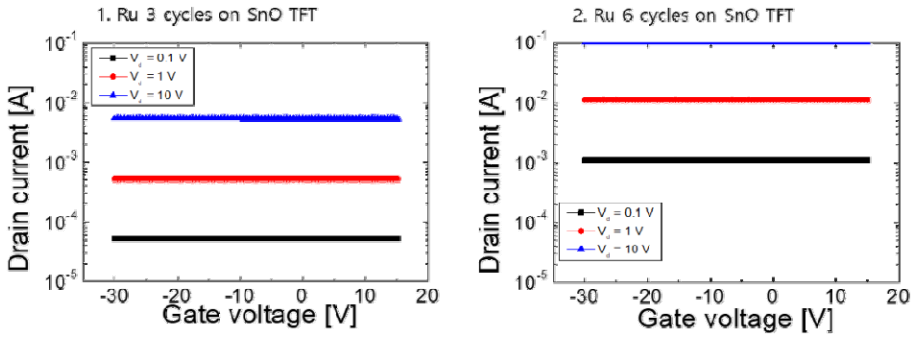
그림 19. SnO 박막 표면에 Ru 원자층 증착 이후 XPS 분석 결과(Sn 3d)



* 출처: 저자 작성

처음 의도와는 다르게 SnO 표면에 Ru 전구체로 인해 모두 산화 또는 환원되었다. 실제 소자를 제작하여 확인해본 결과 switching 특성이 아예 사라져버린 것을 확인할 수 있었으며(〈그림 20〉 참고), 이는 metallic한 Sn 및 전하 운송자 밀도가 높은 SnO2 상으로 인한 결과로 예상된다. 〈그림 18〉의 XPS를 통해서도 이미 확인했듯 루테늄(Ru) cycle 수가 증가할수록 불균등화 반응이 심화되는데 이에 따라 트랜지스터의 전류 크기(current level)도 함께 커지는 것을 확인할 수 있었다. 이는 XPS와 상응하는 결과로 사용된 Ru 전구체의 부적절성을 더욱 확실하게 보여준다. 추가 실험으로 Pt의 SnO 상부 증착을 통한 표면 개질을 진행해보았다.

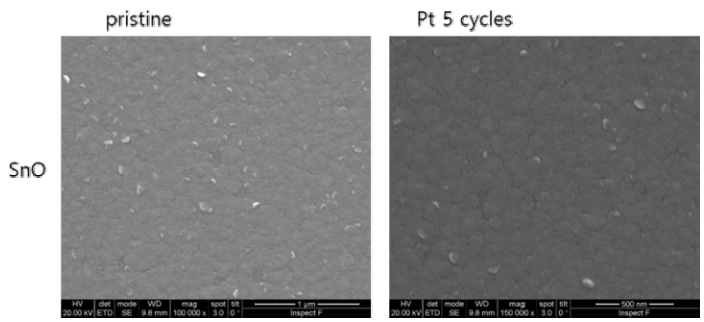
그림 20. 표면에 Ru가 증착된 SnO 트랜지스터의 transfer curve



* 출처: 저자 작성

백금(Pt) 원자층 증착법은 Trimethyl(methyl-cyclopentadienyl)-platinum(IV)과 산소를 사용하여 220℃의 증착온도에서 진행되었다. Ru에서도 금속 닷을 잘 관찰할 수 없었듯 Pt 또한 SnO위에 증착된 금속 닷을 관찰하는 데에는 어려움이 있었다. 〈그림 21〉을 참고해보면 pristine SnO 박막의 표면과 Pt 5 cycle이 증착된 SnO 박막의 표면이 크게 다르지 않음을 알 수 있다. 이에 루테늄(Ru)을 관찰했을 때와 같은 맥락으로 좀 더 미량의 원소를 볼 수 있도록 화학적인 분석을 추가로 진행하였다. 원자 증착양을 면밀도(layer density)로 분석해낼 수 있는 WDXRF(Wavelength Dispersive X-ray Fluorescence)를 활용하였다.

그림 21. SnO 및 Pt 5 cycle이 진행된 SnO 표면의 SEM images



* 출처: 저자 작성

<표 1>은 백금(Pt)과 루테늄(Ru) 금속 닷을 SnO 상부에 각각 90 cycle, 180 cycles 증착했을 때의 면밀도를 보여준다. 루테늄(Ru)을 90 cycles만 증착해도 1.97 $\mu g/cm^2$의 면밀도가 관찰되었음에 비해 백금(Pt)의 경우는 180 cycle을 진행해도 0.00 $\mu g/cm^2$의 면밀도가 관찰되었다. 즉 실제로 SnO 박막 위에 백금(Pt) 증착이 일어나지 않았음을 의미한다. 이는 표면 배향된 *(001)*면의 SnO의 표면 에너지(surface energy)가 매우 낮았기 때문으로 사료된다. 따라서 *(001)*면의 작용기(functional group, 분자들의 특징적인 화학 반응을 담당하는 분자 내의 특정 부분)을 치환하여 표면 에너지(surface energy)를 제어함으로써 백금(Pt) 증착을 가능하게 하기 위해 암모니아 열처리 공정을 진행하였다.

표 1. Metal dot의 XRF layer density

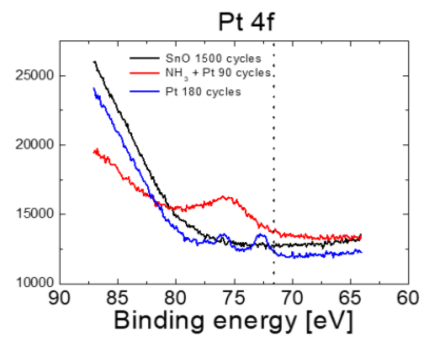
(unit: $\mu g/cm^2$)

SnO 위 증착	90 cycles	180 cycles
Pt	0	0
Ru	1.97	-

* 출처: 저자 작성

백금(Pt)을 바로 SnO 상에 증착했을 때와 백금(Pt)을 증착하기 직전 암모니아로 pre-treatment를 해준 경우의 시편을 각각 XPS로 분석한 Pt 4f spectra 결과를 <그림 22>에 나타내었다. 실제 의도한 대로 암모니아 열처리가 SnO 박막 표면의 작용기(functional group)를 치환하여 표면에너지를 높여주었기 때문에 바로 백금(Pt)을 증착했을 때와는 다르게 Pt가 미량 증착된 것을 관찰할 수 있었다.

그림 22. Pt 4f core electron XPS spectra



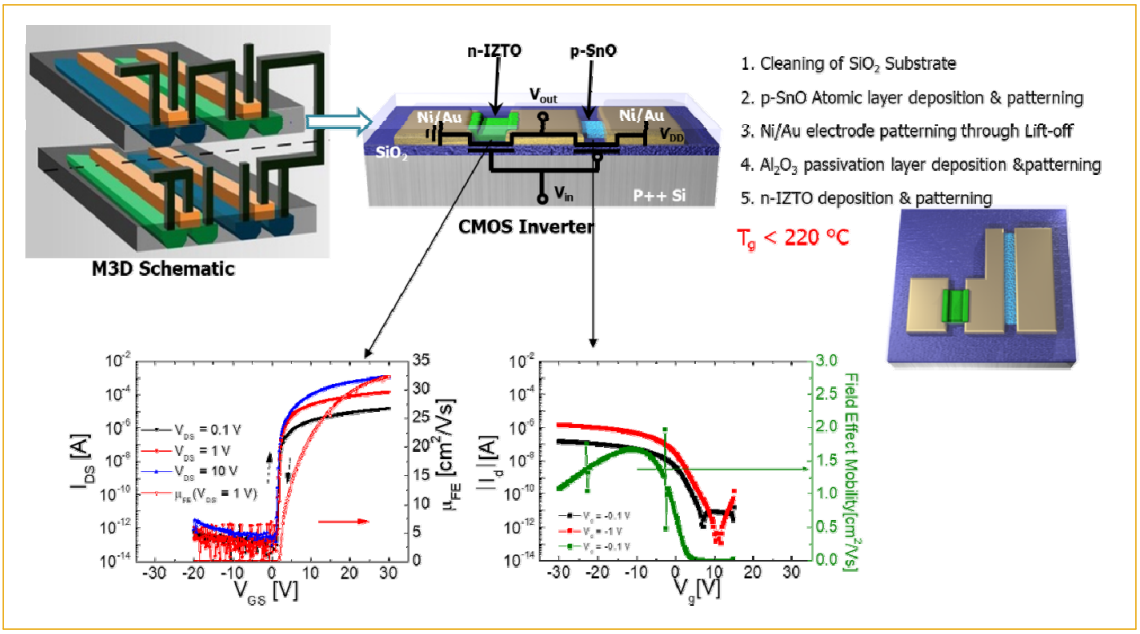
* 출처: 저자 작성

해당 샘플을 선별해 추가 실험으로써 미량의 백금(Pt) 금속 닷이 증착된 SnO 트랜지스터를 제작하여 전기적 성능을 관찰할 필요가 있다. 추가로 현재까지 최적화된 표면/계면처리 기술을 통해 제작한 p형 SnO 트랜지스터와 기개발된 n형 InZnSnO 트랜지스터를 조합하여 CMOS 박막 인버터를 형성한 결과에 대해 다룬다. CMOS 박막 인버터는 M3D integration 구현 가능성을 보여줄 수 있는 가장 기초단위 소자로써 본 연구를 통해 확보된 p-type SnO 트랜지스터의 활용성을 명확히 보여줄 수 있을 것이다.

4. 추가 연구 사항(CMOS 박막 인버터)

인하대학교 화학공학과 연구그룹에서 2019년 발표한 바 있는 n-type 반도체인 IZTO (InZnSnO) 기술(Baek et al, 2019)과 본 연구를 통해 개선된 특성을 갖는 p-type SnO 트랜지스터를 집적하여 하나의 CMOS 박막 인버터 소자를 구현하였다. 해당 CMOS 박막 인버터 소자는 전체 프로세스가 220℃ 이하인, 저온에서 제작되는 것을 목표로 하였다. 열처리가 전혀 들어가지 않은 저온 공정을 통해 M3D integration의 가능성을 제시하기 위함이다. 확보된 인버터 공정을 〈그림 23〉 오른쪽에 나타내었다. p-type SnO 박막을 먼저 증착 및 패터닝 한 뒤 전극을 형성하고 봉지막 공정을 진행하였다. 이후 IZTO를 증착하고 필요한 부분을 제외한 모든 부분을 식각한 뒤 전극을 패터닝하면 소자가 완성된다. 이 때 p-type SnO 박막의 봉지막은 n형 소자를 패터닝할 때 에치 스톱 층(etch-stop layer) 역할을 동시에 하는 것을 특징으로 한다.

그림 23. M3D integration의 기초 단위인 CMOS 박막 인버터 모식도와 각 n형, p형 소자에서의 transfer curve

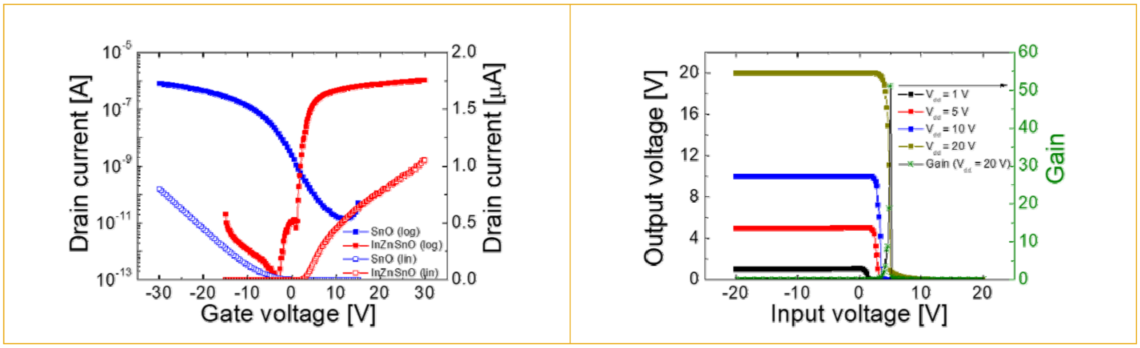


* 출처: 저자 작성

〈그림 24〉의 좌측 그림은 제작된 CMOS 박막 인버터에서 각각 측정된 n-type IZTO 박막 트랜지스터의 전달 곡선(transfer curve)과 p-type SnO 트랜지스터의 전달 곡선(transfer curve)을 보여준다. 〈그림 23 〉 내의 전달 곡선(transfer curve)과 비교했을 때 공정의 복잡도로 인해 문턱전압 이하 스윙(subthreshold swing) 값이 약간 증가되는 열화를 보였으나 switching 특성에는 큰 변화가 없음을 알 수 있다. 여러 단계의 공정 후에도 각각의 n, p 소자의 전달 곡선(transfer curve)은 정상적인 특성을 보이기에 인버터 또한 잘 작동할 것임을 알 수 있었다. 실제 〈그림 24〉의 우측 그림은 CMOS 박막 인버터의 전압 전달 특성(voltage

transfer characteristic)을 보여주는데, V_{dd}가 20 V인 경우 gain 값이 약 50에 달하는 매우 좋은 특성이 관찰되었다. 즉 표면과 계면이 개선된 p형 소자와 기존 개발된 n형 소자를 조합하여 우수한 성능의 CMOS 박막 인버터를 제작함으로써, 본 연구그룹에서 개발된 재료를 통해 M3D 저온 integration의 가능성을 제시하였다고 볼 수 있다.

그림 25. (좌)CMOS 박막 인버터 상 n-IZTO 및 p-SnO 트랜지스터의 transfer curve. (우)CMOS 박막 인버터의 voltage transfer characteristic



* 출처: 저자 작성

IV. 결과

인하대학교의 화학공학과의 원자층 증착 공정 및 반도체 소자 제조 기술과 한국화학연구원의 전구체 합성 기술, 한국과학기술연구원의 금속 합성 및 분석기술을 융합하여 p-type SnO 박막의 계면과 표면을 제어하는 연구를 수행하였다. 원자층 증착법으로 합성된 금속 닷 데코레이션을 통해 p-type SnO의 성능을 개선함으로써 새로운 패러다임의 소재를 합성하였다. SnO 박막의 표면과 계면에 필연적으로 위치하는 결정립계는 정공의 산란을 일으켜 이동도의 저하를 야기한다. 이러한 고질적인 문제를 해결하기 위해 일함수가 높은 백금(Pt) 또는 루테늄(Ru) 금속 닷을 결정립계에 데코레이션 함으로써 금속-반도체(Metal-Semiconductor) 오믹 접합을 형성할 수 있었고, 결정립계 운송자 산란이 억제되는 결과를 얻었다. 추가적으로 금속 닷의 양과 종류에 따라 SnO 박막의 결정 성장 배향이 달라지는 현상을 면밀히 관찰하였다. 금속 닷이 많아질수록 박막 트랜지스터에는 불리한 수직형 결정 배향이 관찰되었는데 이는 역으로 가스 센서에 응용될 경우 매우 좋은 특성으로 나타날 수 있다. 실제 가스 센싱 특성을 관찰해본 결과 상온에서도 고 민감도를 보임으로써 차세대 저전력 센서로써의 가능성을 증명하였다. 따라서 추가적인 융합연구로 고감도 가스 센서 개발까지 진행해볼 수 있을 것으로 예상한다. 본 연구를 통해 표면과 계면이 개선된 p형 SnO 박막의 활용성을 더욱 명확히 제시하기 위해 기개발된 n형 IZTO 박막과 집적하여 CMOS 박막 인버터 공정을 수행하였다. 집적을 위해 4개 이상의 반도체 공정이 추가됨에도 불구하고 각 각의 n-type과 p-type

박막들은 switching 특성이 크게 열화되지 않았으며 결과적으로 성공적인 인버터의 전압 전달 특성(voltage transfer characteristic)을 확보할 수 있었다. 모든 공정이 220℃이하의 온도에서 진행되었으므로 차세대 저온 M3D 공정 및 재료로 응용될 수 있는 충분한 가능성을 시사한다.

저자소개 **백인환**(In-Hwan Baek)

• 학력

서울대학교 재료공학 박사
연세대학교 신소재공학 학사

• 경력

現) 인하대학교 화학공학과 조교수
前) 삼성전자 반도체연구소 책임연구원

참고문헌

〈국내문헌〉

1) 권혁인. (2017). 고성능 고신뢰성 p-type 산화물 박막트랜지스터 제작 및 산화물 박막트랜지스터 기반 complementary 로직 시스템 구현에 관한 연구. 과학기술정보통신부. https://doi.org/10.23000/TRKO201800004237

2) 최민기, 전다희, 황인홍, 백인환. (2023). 트랜지스터용 p-형 산화물 반도체 (니켈 산화물, 주석 산화물, 구리 산화물) 최신 동향 분석. 세라미스트, 26(1), 75-89. https://doi.org/10.31613/ceramist.2023.26.1.06

〈국외문헌〉

3) Baek, I. H., et al. (2019). High-Performance Thin-Film Transistors of Quaternary Indium-Zinc-Tin oxide Films Grown by Atomic Layer Deposition. ACS Applied Materials & Interfaces, 11(16), 11874-11881.

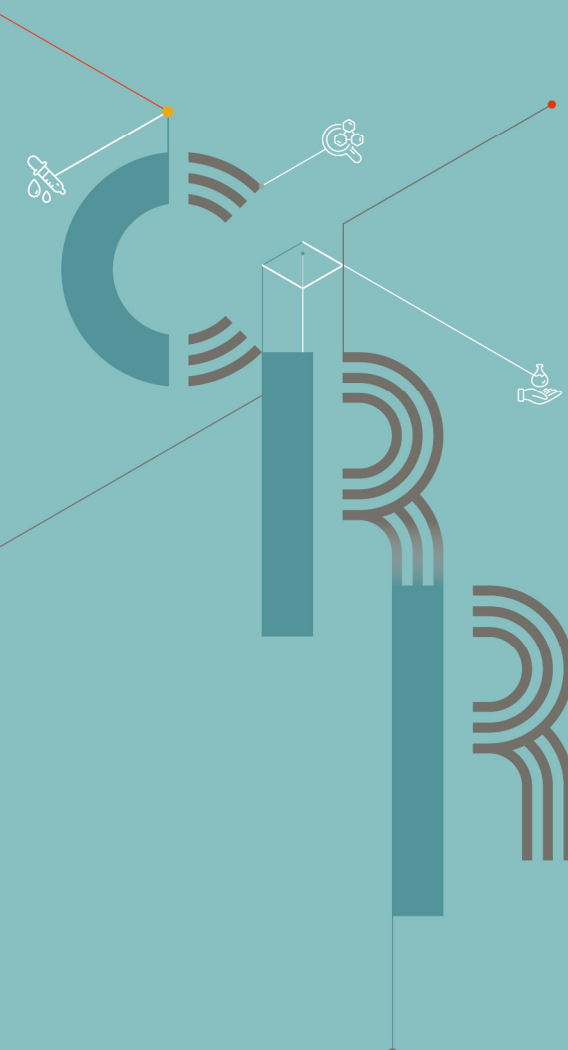
4) Baek, I. H., et al. (2021). Enhancement of Electrical Performance of Atomic Layer Deposited SnO Films via Substrate Surface Engineering. Journal of Materials Chemistry C, 9, issue 36, 12314-12321.

5) Hagen, D. J., et al. (2017). Island Coalescence during Film Growth: An Underestimated Limitation of Cu ALD. Advanced Materials Interfaces, 4(18), 1700274.

6) Kim, H. I, et al. (2021). Highly Dense and Stable p-Type Thin-Film Transistor based on Atomic Layer Deposition SnO Fabricated by Two-Step Crystallization. ACS Applied Materials & Interfaces, 13(26), 30818-30825.

7) Pyeon, J. J., et al. (2020). Highly Sensitive Flexible NO_2 Sensor Composed of Vertically Aligned 2D SnS_2 Operating at Room Temperature. Journal of Materials Chemistry C, 8, issue 34, 11874-11881.

8) Hu, Y., et al. (2019). First Principles Calculations of Intrinsic Mobilities in Tin-based Oxide Semiconductors SnO, SnO_2, and Ta_2SnO_6. Journal of Applied Physics, 126(18), 185701.



융합연구리뷰

Convergence Research Review

02

들리는 얼굴:
이미지를 활용한 가상 인물의 목소리 생성 및 변환 인공지능

이정식(영남대학교 전자공학과 석사과정생)

02

들리는 얼굴: 이미지를 활용한 가상 인물의 목소리 생성 및 변환 인공지능

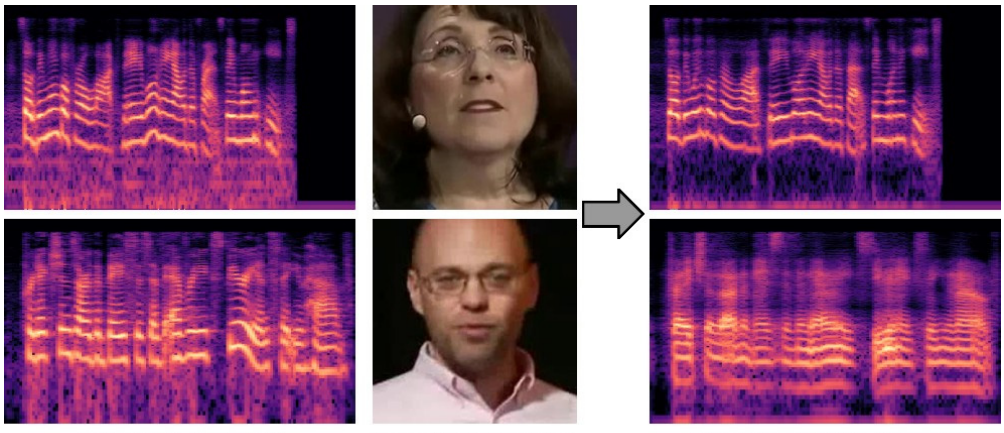
이정식(영남대학교)

I. 서론

음성 변환(VC, Voice Conversion) 은 주어진 음성 파형(waveform)을 목표하는 화자의 음성 스타일로 변형하는 음성 처리의 핵심 태스크 중 하나이다. 수많은 생성 모델들(generative models)이 목표 화자의 음성 파형(레퍼런스 목소리)으로부터 고품질의 음성 변환 결과를 얻기 위해서 연구되어 왔다(Lee et al, 2021; Qian et al, 2019; Qian et al, 2020). 그러나 이러한 방법론들은 레퍼런스 목소리가 존재하지 않는 새로운 목소리 생성에 제한이 있어, 콘텐츠 사업에서 활용이 어렵다. 예를 들어, 게임과 애니메이션 같은 콘텐츠 산업의 가상 인물 디자인 과정에는 외형 디자인과 더불어 외형과 어울리는 목소리를 디자인하는 과정이 필연적이다. 그러나 기존 방법론들은 이미 존재하는 음성 신호에서 음성 스타일의 정보를 추출하기 때문에 새로운 음성이 필요한 가상 인물 디자인에 활용이 제한된다.

이러한 기존 음성 변환의 한계를 극복하기 위해, 최근 음성으로부터 화자의 스타일을 추출하는 방법뿐만 아니라, 〈그림 1〉에서 보이듯이 목표 화자의 얼굴 이미지를 활용하는 교차-모달 음성 변환(cross-modal voice conversion) 방법론들이 연구되고 있다(Lu et al, 2021). 얼굴 기반의 음성 생성 및 변환은 음성과 얼굴 외형 간의 강력한 상관관계가 존재한다는 연구에 기반을 두며(McGurk et al, 1976; Kamachi et al, 2003; Smith et al, 2016), 이러한 상관관계를 딥러닝 모델로 모델링하는 것이 매우 중요하다. 이를 위해 기존의 방법론들은 음성과 이미지 두 모달 간의 공동 표현 공간(joint representation space) 학습과 얼굴 이미지로부터 음성 스타일을 추출하기 위해 매우 복잡한 신경망의 구조와 학습 방법을 채택하고 있다. 또한 학습 데이터에서 사용한 얼굴에 대해서만 음성을 생성할 수 있는 한계가 있으며, 이는 새로운 음성 디자인 및 생성에 적합하지 않다.

그림 1. 얼굴 이미지를 활용한 음성 변환



* 주어진 음성 신호의 스타일을 주어진 얼굴 이미지에 대응되게 변형한다.

** 출처: 저자 작성

음성 변환 외 다른 분야에서 최근 확산 모델(diffusion model)이 교차-모달 생성(cross-modal generation)에 우수한 생성 결과를 보이고 있다(Ho & Salimans, 2020; Song et al, 2021). 특히, 잠재 확산 모델(LDM, Latent Diffusion Model)이라는 모델(Rombach et al, 2022)이 주어진 문장으로부터 이미지를 생성하는 Text-to-Image(T2I) 분야에서 주목할 만한 결과를 보이고 있다. 기존 확산 모델이 이미지 원본 공간에서 이미지를 생성한 것에 반해(Ho & Abbeel, 2020), LDM은 압축된 잠재 공간(latent space)에서 서서히 노이즈(noise)를 제거하면서 텍스트에 대응하는 이미지 잠재 벡터(image latent vector)를 생성한다. 이를 통해 LDM은 기존 확산 모델이 많은 컴퓨팅 리소스를 요구한다는 문제를 해결했으며, 본 연구에서 또한 얼굴-음성 모달 음성 변환을 위해 LDM을 활용한다.

그러나 Audio-X modal generative 분야에서 몇 가지 해결되지 않은 문제들이 존재한다. 첫 번째로 기존의 교차-모달 생성 모델(cross-modal generative models) (Rombach et al, 2022; Ramesh et al, 2021)은 고성능의 사전 학습된 교차-모달 모델(pre-trained cross-modal model)의 잠재 공간(latent space) (Radford et al, 2021; Wu et al, 2023)을 활용하지만 얼굴-음성 모달에 대해서는 이러한 고성능의 사전 학습된 모델이 없다. 이로 인해 얼굴 이미지에 대응하는 음성 표현(speech representation)을 생성하는 LDM 학습이 어렵다. 두 번째로 기존 교차-모달 학습에 사용된 데이터들(Schuhmann et al, 2021; Wu et al, 2023)에 비해 face-speech data(Afouras et al, 2018; Chung et al, 2018)의 스케일이 매우 작기 때문에 고품질의 교차 표현 공간(joint representation space) 학습이 쉽지 않다. 마지막으로 음성 변환은 스타일 변환(style transfer) (Gatys, 2016; Huang et al, 2017)과 같이 음성의 콘텐츠와 스타일을 분리하고 다른 음성의 스타일과 분리된 콘텐츠를 합쳐 음성 변환을 수행하지만, 확산 모델에서는 콘텐츠와 스타일 분리에 대한 연구가 아직 미흡하다.

앞서 언급한 문제들을 해결하기 위해, 새로운 face-conditioned voice conversion model인 Grad-FVC를 제안한다. 기존 CLIP(Radford et al, 2021) 방식의 교차-모달 학습은 연속된 공간(continuous space)에서 아무런 제약(regularization)이 없기 때문에 작은 스케일의 데이터 셋으로 제약이 없는 공간에서의 교차 표현(joint representation) 학습이 어렵다. 이에 따라, 연속된 공간에서 학습하는 것 대신 벡터 양자화(VQ, Vector Quantization) (Van, 2017)을 도입하여 불연속적 공간(discrete space)에서 교차 표현을 학습한다. 또한 두 모달의 표현 공간을 동시에 학습하면서 교차 표현 공간을 구성하는 것은 shortcut problem을 야기할 수 있기 때문에 이미 잘 학습된 하나의 모달의 표현 공간에 다른 모달의 표현 공간을 정렬(align)하는 feature-based knowledge distillation(Gou et al, 2021)을 수행한다. 우선, 음성신호에 대해 유의미한 코드를 가지고 있는 오디오 코드북(codebook) 을 학습시킨다. 학습 완료된 오디오 코드북은 얼굴 표현 공간 학습에서 얼굴 이미지의 잠재 벡터를 양자화 하는데 사용되며, 이를 통해 얼굴 잠재 벡터가 오디오 잠재 공간에 투영되도록 한다. 이런 식으로 얼굴-음성 모달들에 대해서 하나의 공유하는 코드북(shared codebook)을 구성하고 각 모달의 잠재 벡터가 하나의 코드북으로 강제하여 교차 표현 공간을 구성할 수 있게 된다.

그러나 단순히 공유하는 코드북을 도입하는 것으로는 교차 표현 공간의 구성을 보장할 수 없다. 예를 들어, 얼굴과 음성 모달 각각의 표현 학습이 코드북의 일부분에 편향되어 학습되는 코드북 편향화(codebook collapse) 문제가 생길 수 있기 때문이다. 이러한 문제를 해결하기 위해, 교차-모달 학습에 효과적이라고 알려진 대조 학습(contrastive learning) (Radford et al, 2021; Wu et al, 2023)을 채택하여 얼굴-음성 모달의 특징을 정렬(align) 한다. 공유하는 코드북, 대조 학습 그리고 일반적인 VQGAN의 손실함수(loss)로 학습된 얼굴-음성 모달의 인코더, 디코더, 그리고 코드북으로 구성된 cross-modal VQGAN (CrossVQGAN)을 제안한다. CrossVQGAN은 VQGAN의 손실함수로 모달 간의 상관관계가 존재하는 모달 특성화 특징 맵을 추출할 수 있으며, 단순히 대조 학습만 수행한 방법론들 보다 뛰어난 특징 맵 추출 능력을 가진다. 이렇게 학습된 오디오 인코더, 디코더, 코드북 그리고 이미지 인코더를 활용하여 LDM을 학습한다. 학습된 LDM은 오디오 인코더를 통해 구성된 얼굴-음성 교차 공간에서 이미지 인코더를 통해 추출된 얼굴 잠재 벡터와 어울리는 음성 잠재 벡터를 생성하게 된다. 생성된 음성 잠재 벡터는 오디오 디코더를 통해 다시 음성 신호로 변환된다.

이렇게 학습된 LDM은 주어진 얼굴 이미지와 어울리는 음성을 생성하는 것이 가능하지만, 랜덤 노이즈로부터 음성을 생성하는 확산 모델의 한계로 생성된 음성의 발화내용을 조절하는 것이 불가능하다. 이미지 분야에서 확산 모델을 사용하여 이미지의 속성을 변경하는 선행 연구(Meng et al, 2022)에 영감을 받아, 음성에서 발화 내용만 분리하여 노이즈로 변환하는 Content-Aware Encoding을 제안하고, 제안한 방법으로 추출된 노이즈와 얼굴 이미지를 사용하여 원하는 발화 내용을 가지면서 얼굴 이미지와 대응되는 음성을 생성한다.

본 연구의 기여한 바는 다음과 같다.

• 노이즈가 존재하는 잠재 벡터를 노이즈리스 잠재 벡터로 변형시키는 LDM(Latent Diffusion

Model)을 사용하면서 얼굴 이미지로부터 음성 스타일을 추출하여 음성을 생성하는 Grad-FVC를 제안한다.

- 확산 모델의 확률론적 인코딩(stochastic encoding) 대신 음성의 스타일 정보는 제거하고 발화 정보만을 가지고 있는 노이즈를 생성하는 Content-Aware Encoding을 제안한다.
- Content-Aware Encoding과 학습한 확산 모델을 사용하여, 원하는 발화 내용을 가진 높은 품질의 음성을 생성할 수 있다.

II. 기존 연구 동향

1. 확산 모델

확산 모델(Ho & Abbeel, 2020; Song et al, 2021; Dhaiwal et al, 2021)은 이미지 생성 분야에서 강력한 성능을 보이는 생성 모델이다. 이는 복잡한 분포를 모델링할 수 있기 때문이며, 이러한 장점으로 생성된 이미지의 품질(fidelity)과 다양성(diversity) 측면에서 이전 GAN(Generative Adversarial Network) (Brock et al, 2019; Karras et al, 2019; Karras et al, 2020) 보다 우수하다. 하지만 확산 모델은 해상도가 작은 잠재 공간(latent space)에서 점진적으로 업-샘플링 하여 이미지를 생성하는 GAN과 다르게 이미지 공간(또는 픽셀 공간)에서 이미지를 생성하기 때문에, 확산 모델을 학습하는 데는 수백 개의 그래픽 처리 장치(GPU, Graphics Processing Unit)를 요구하고, 단계별로 노이즈를 제거하기 때문에 추론 비용 또한 매우 많이 든다. 이러한 문제점을 해결하기 위해 Rombach 외 저자들은 픽셀 공간보다 매우 작은 해상도를 가진 잠재 공간에서 동작하는 LDM(Latent Diffusion Model)을 제안하였다(Rombach et al, 2022). 이를 통해 최근 확산 모델들이 텍스트-이미지(Rombach et al, 2022), 텍스트-비디오(Blattmann et al, 2023) 등 다양한 입출력 관계를 모델링하고 뛰어난 생성 결과를 보이고 있다. 또한 최근에는 오디오(Liu, Chen et al, 2023; Liu, Tian et al, 2023; Luo et al, 2023)와 음성(Popov et al, 2022; Lee, 2023) 생성 분야에서 확산 모델이 다양하게 사용되고 있다.

그러나 확산 모델은 랜덤한 노이즈와 텍스트와 같은 유저 가이드를 통해 이미지를 생성하기에 원하는 이미지에서 의미론적 정보(semantic information)만 남기고 새로운 이미지로 변환하는 이미지 수정 태스크에 취약하다. 이를 해결하기 위해, 최근 이미지로부터 의미론적 정보만을 가지고 있는 노이즈를 생성하는 diffusion inversion 연구가 수행되고 있다. Diffusion inversion은 확산 모델을 사용하여 노이즈를 제거하는 것이 아니라 추정된 노이즈를 다시 더해주는 것을 반복하여, 해당 이미지의 의미론적 정보를 가지는 노이즈를 생성한다. 그 중 Meng 외 저자들은 inversion 과정 중에 적당한 양의 노이즈만을 이미지에 주입해 생성한 노이즈가 이미지의 사물의 외각선과 콘텐츠와 같은 의미론적 정보만을 가지고

있음을 보였고, 이를 활용하여 확산 모델로 이미지를 변경하는 SDEdit을 제안했다(Meng et al, 2022). SDEdit은 의미론적 의미만 가진 노이즈와 텍스트를 확산 모델의 입력으로 하여 원본 이미지의 콘텐츠만 유지하고 텍스트에 대응되는 스타일로 이미지를 변형 가능함을 보였다.

2. 음성 변환

음성 변환은 타겟 화자의 음성을 사용하여 입력 음성의 스타일을 원하는 화자의 음성 스타일로 변경하는 작업이다. 음성 변환 작업 중 zero-shot voice conversion으로 알려진 non-parallel many-to-many voice conversion은 원하는 화자의 음성과 입력 음성간의 일 대 일 매칭이 존재하지 않는 경우를 의미하며, 정답 샘플 쌍이 존재하지 않기에 매우 어려운 문제이다. 이러한 zero-shot voice conversion 문제를 해결하기 위해, 몇몇 연구들이 존재한다(Qian et al, 2019; Chou et al, 2019; Wu et al, 2020).

AutoVC(Qian et al, 2019)는 오토인코더(autoencoder) 구조를 활용하여 음성으로부터 화자의 스타일 정보와 내용을 분리하는 방법을 제안하였다. 분리된 스타일 정보와 내용을 다시 합쳐 원본 음성을 재구성하는 재구성 손실함수(reconstruction loss)를 활용하여 모델의 학습 복잡성을 낮추었으며, 학습 데이터에 없는 화자에 대한 고품질의 음성 변환에 성공하였다. AutoVC에 이어 AdaIN-VC(Chou et al, 2019)는 스타일 정보와 내용을 추출하고 변형하기 위해, IN(Instance Normalization) (Ulyanov et al, 2016)과 AdaIN(Adaptive Instance Normalization) (Huang & Belongie, 2017)을 활용한다. AdaIN-VC는 추출된 특징 맵을 정규화 하는 IN을 통해 정규화 된 특징 맵을 내용으로 가정하고, 정규화 되지 않은 특징 맵의 평균과 표준편차로 정규화 된 특징 맵을 변조하는 것으로 음성 변환을 수행한다. VQVC(Wu et al, 2020)는 음성으로부터 내용 정보를 추출하기 위해 벡터 양자화(VQ, Vector Quantization) 방법을 활용하였고, VQ를 통해 얻은 벡터와 원본 벡터와의 차이를 스타일 정보로 사용하였다.

이러한 오토인코더 기반의 방식에 이어, 최근 확산 모델을 사용하여 좀 더 좋은 품질의 음성 변환을 시도하려는 연구들이 존재한다(Popov et al, 2022). Popov 외 저자들은 음성 신호에서 발화 내용을 추출하기 위해 평균 음성을 예측하는 인코더와 예측된 평균 음성으로부터 변환된 음성을 생성하는 확산 모델을 디코더로 채택하였다. 또한 확산 모델의 느린 추론 속도를 해결하기 위해, 새로운 확산 모델의 샘플링 방식을 제안하였다.

또한 최근 타겟 화자의 음성 대신 타겟 화자의 얼굴 이미지를 활용하여 음성을 변환(Lu et al, 2021)하고 생성(Goto et al, 2020; Lee et al, 2023)하는 연구들이 나타나고 있다. FaceVC(Lu et al, 2021)는 얼굴 이미지로부터 음성 스타일을 추출하기 위해, 3단계로 문제를 나누어 학습을 시도한다. 그러나 이러한 노력에도 불구하고 얼굴-음성 간의 상관관계 학습이 부족하여 기존 오디오 기반의 음성 변환보다 좋지 못한 성능을 보인다. Face2Speech(Goto et al, 2020)는 얼굴 이미지로부터 추출된 특징 맵과 얼굴 이미지와 대응되는 음성 스타일 특징 맵이 유사한 값을 가지도록 얼굴 이미지 인코더를 학습하고, 이를 TTS(Text-To-Speech) 모델의 입력으로 활용한다. 간단한 방법으로 얼굴 기반의 음성 생성이 가능했지만,

여전히 좋지 못한 음성 생성 결과를 보인다. Face-TTS(Lee et al, 2023)은 음성 생성 모델을 확산 모델로 변경하여 기존 얼굴 기반 음성 변환 및 생성 모델들 보다 뛰어난 성능을 보인다.

3. 교차-모달 학습

교차-모달 학습은 둘 이상의 모달(modal)에서 얻은 정보 간의 유의미한 상관관계성을 파악하는 것을 기본으로 한다. 이렇게 교차 모달 학습을 통해 파악된 상관관계성은 크게 두 가지 방면으로 활용된다. 첫 번째로 하나의 모달의 정보를 사용하여 다른 모달의 정보를 강화(Liu et al, 2019; Nawaz, 2021; Hong et al, 2022)하거나, 두 번째로 하나의 모달 정보를 활용하여 다른 모달의 정보를 변경하거나 새로운 정보를 생성한다(Rombach et al, 2022; Blattmann et al, 2023; Liu et al, 2023; Liu et al, 2023; Luo et al, 2023; Popov et al, 2022; Lee et al, 2023.

이러한 교차-모달 학습을 위해, 자가 지도 표현 학습 (self-supervised representation learning) 중 하나인 대조 학습(contrastive learning)이 널리 사용되고 있다(Oord et al, 2018]; Wu et al, 2018; Chen et al, 2020; He et al, 2020). 이미지-모달에 한정하여 대조 학습은 하나의 이미지로부터 서로 다른 증강법(augmentation)을 통해 얻은 두 이미지를 뷰(view)라고 정의하고, 동일한 이미지로부터 생성된 두 뷰의 특징 맵은 유사하게, 서로 다른 이미지로부터 생성된 뷰 간의 특징 맵은 다르게 학습한다. 이러한 대조 학습은 여러 교차-모달 학습에서도 활용되었다(Radford et al, 2021; Wu et al, 2023). Radford 외 저자들은 이미지와 이에 상응되는 텍스트를 각각의 뷰로 정의하여 대조 학습을 수행하여, 이미지와 텍스트 간의 교차-모달 학습을 수행하는 CLIP 모델을 제안하였다(Radford et al, 2021). Wu 외 저자들은 CLIP 모델과 유사하게 오디오와 이에 상응되는 텍스트를 각각의 뷰로 정의하여 교차-모달 학습을 수행하는 CLAP 모델을 제안하였다(Wu et al, 2023). 이러한 교차-모달 학습은 많은 데이터양과 컴퓨팅 리소스가 요구된다.

그러나 오디오-얼굴 데이터 셋의 스케일(Afouras et al, 2018; Chung et al, 2018)은 텍스트-이미지(Schuhmann et al, 2021) 또는 텍스트-오디오(Wu et al, 2023)의 스케일에 반해 매우 적기 때문에 유의미한 상관관계를 학습하기에 어려움이 존재한다. 또한 이러한 모델들은 학습을 위해 수많은 GPU를 요구하며 일반적으로 대량의 컴퓨팅 리소스를 사용이 불가능하다. 이러한 컴퓨팅 리소스 문제를 해결하기 위해 최근 MAE(Masked Autoencoder) (He et al, 2022)를 활용하여 오디오-이미지 간의 교차-모달 학습이 이루어지고 있다(Gong et al, 2022). 그러나 여전히 적은 수의 데이터 셋에 대한 교차-모달 학습의 연구는 이루저지 않고 있다.

III. 방법론

A_i 는 음성 오디오, V_i 는 얼굴 이미지라 하고 얼굴-음성 데이터 쌍 ($D = \{A_i, V_i\}_{i=1}^{N}$) 이 있을 때, 본 연구의 목표는 오디오 인코더(audio encoder)와 이미지 인코더(vision encoder)가 교차 표현 공간(joint-representation space)에서 유의미한 정렬된 오디오-이미지 특징(audio-vision aligned feature)을 추출하는 것이다. 학습된 표현 공간을 활용하여 음성 생성 모델을 학습하여 얼굴 이미지와 연관된 음성을 생성하고, 더 나아가 음성에서 발화 내용과 발화자의 스타일을 분리하여 , 최종적으로 분리된 발화 내용 벡터(content feature)와 임의의 얼굴 이미지를 사용하여 임의의 얼굴 이미지와 연관된 음성을 생성하는 것이 목표이다.

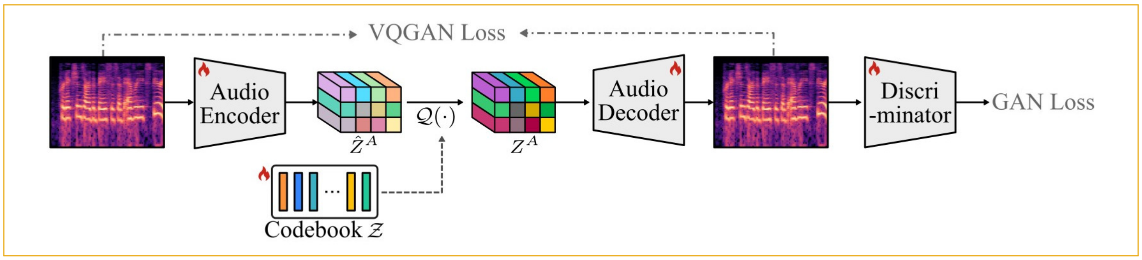
교차 표현 공간을 학습하기 위해, 많은 연구들이 각 모달(modal)의 잠재 벡터(latent vector)에 대조 학습(contrastive learning)을 수행하는 교차-모달 학습(cross-modal learning)을 사용한다. 이 때 많은 기존 방법들은 주로 연속된 공간에서 대조 학습을 수행한다(Radford et al, 2021; Wu et al, 2023). 그러나 이러한 제한되지 않은 공간에서 서로 다른 두 모달의 잠재 벡터는 학습과정 중에 지속적으로 업데이트 되기 때문에, 각 모달의 잠재 공간 학습과 동시에 교차 공간을 학습하는 것은 매우 어려운 일이다. 이에, SwAV(Caron et al, 2020)에 영감을 받아 벡터 양자화(VQ, Vector Quantization)을 활용하여 제한된 양의 코드들로 구성된 코드북(codebook)으로 불연속 공간에서 교차 표현 공간을 학습한다. 구체적으로 하나의 코드북을 두 모달에 공유하여 두 모달의 잠재 벡터들을 동일한 벡터 양자화 된 공간에 투영하여 교차 표현 공간을 구성한다. 그러나 두 모달 간의 유의미한 교차-모달 학습은 여전히 어려운 작업이면서 많은 컴퓨팅 자원을 요구하기 때무에, 문제를 분리하여 단계별로 해결한다. 우선적으로 음성 잠재 공간을 사전학습(pre-train) 시키고(〈그림 2〉 참고), 그 다음 사전 학습된 오디오 잠재 공간에 이미지 잠재 공간을 정렬(align)한다(〈그림 3) 참고〉. 이러한 방법으로 적은 컴퓨팅 자원으로 효과적으로 얼굴-음성 교차 공간을 구성할 수 있다. 그런 다음 구성된 교차 표현 공간에 음성 생성 모델을 학습하고(〈그림 5〉 참고), 교차 표현 공간에서 발화 내용만을 분리하여 음성 변환(〈그림 6〉 참고)을 수행한다. 앞에서 언급한 내용들을 남은 하위 장에서 자세히 설명한다.

1. 얼굴-음성 표현 공간 학습

1.1 음성 표현 공간 사전 학습

두 모달에 특화된 표현 학습(modal-specific representation learning)과 교차-모달 학습(cross-modal learning)을 동시에 수행하는 대신, 잘 학습된 오디오 표현 공간(well pre-trained audio representation space)을 구성하고 해당 공간에 얼굴 표현 공간(visual representation space)을 정렬한다. 이를 위해, 〈그림 2〉에서 보이는 바와 같이, VQGAN(Esser et al, 2021)을 채택하여 오디오 특화 표현 공간 학습(audio-specific representation learning)을 수행한다.

그림 2. 음성 표현 공간 학습 도식도



* VQGAN(Esser, 2021)의 VQGAN 손실함수를 사용하여 음성 표현 공간을 학습한다.

** 출처: 저자 작성

멜-스펙트로그램 (mel-spectrogram) $\mathbf{A} \in \mathbb{R}^{T \times F}$ 를 오디오 인코더 $E^{\mathbf{A}}(\cdot)$를 통해 C와 f가 각각 채널과 다운 샘플링 인자로 정의되는 오디오 특징 맵 $\hat{\mathbf{Z}}^{\mathbf{A}} \in \mathbb{R}^{C \times H/f \times W/f}$ 를 추출한다. 추출된 오디오 특징 맵의 공간 코드 (spatial code) $\hat{\mathbf{Z}}^{\mathbf{A}}_{ij}$ 를 각각 코드북의 코드들 중 가장 가까운 코드 $\mathbf{z}_k$ 로 치환하는 벡터 양자화 (vector quantization) 을 수행하여 양자화 된 특징 맵 (quantized feature map) $\mathbf{Z}^{\mathbf{A}}$를 얻는다 (수식 1). 이 때 i와 j는 각각 공간 위치 인덱스를 나타낸다.

$$\mathbf{Z} = \mathcal{Q}(\hat{\mathbf{Z}}) := \left(\underset{z_k \in \mathcal{Z}}{\operatorname{argmin}} \left\| \hat{\mathbf{Z}}_{ij} - \mathbf{z}_k \right\| \right), \tag{1}$$

이 때, $\mathcal{Q}(\cdot)$ 는 벡터 양자화 함수를 뜻한다. 벡터 양자화 후, 우리는 오디오 디코더(audio decoder) $G^{\mathbf{A}}(\cdot)$ 를 통하여 입력을 복원(reconstruction)한다.

$$\hat{\mathbf{A}} = G^{\mathbf{A}}(\mathbf{Z}^{\mathbf{A}}). \tag{2}$$

벡터 양자화 함수는 미분이 불가능(non-differentiable)하기 때문에, 디코더의 gradient를 인코더로 복사하여 인코더의 gradient를 근사화 하는 straight-through gradient estimator(Bengio et al, 2013)를 사용하여 수식 (3)의 손실함수를 최소화하여 인코더, 디코더, 그리고 코드북을 학습한다.

$$\mathcal{L}_{VQ}(E, G, \mathcal{Z}) = \left\| \hat{\mathbf{X}} - \mathbf{X} \right\|^2 + \| \mathrm{sg}\,[E(\mathbf{X})] - Z \|_2^2 + \| \mathrm{sg}\,[Z] - E(\mathbf{X}) \|_2^2, \tag{3}$$

수식 (3)의 $E, G, \mathcal{Z}$는 각각 인코더, 디코더, 그리고 코드북을, $sg[\cdot]$은 stop-gradient 연산을, $\mathbf{X}$와 $\hat{\mathbf{X}}$은 각각 인코더의 입력과 디코더의 출력을 나타낸다. 추가로 코드북 학습을 용이하기 위해 VQGAN에서 제안한 GAN 손실함수 (수식 4)를 적용한다.

$$\mathcal{L}_{GAN} = \log D(\mathbf{X}) + \log(1 - D(\hat{\mathbf{X}})), \tag{4}$$

수식 4의 $D(\cdot)$은 패치 기반의 판별기(patch-based discriminator) (Isola et al, 2017)를 의미한다.

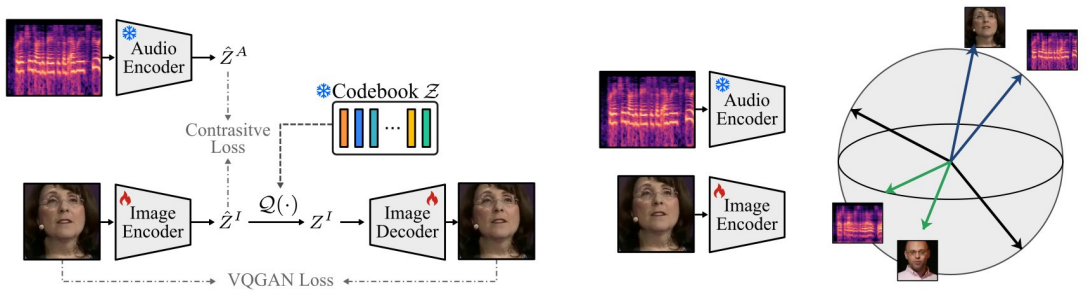
1.2 음성과 연관된 비전 표현 공간 학습

음성 표현 공간 학습과 동일하게 VQGAN을 사용하여 얼굴 표현 공간을 학습한다. 이 때의 목표는 단순히 얼굴 표현 공간을 학습하는 것이 아니라 음성 표현 공간과 얼굴 표현 공간이 동일한 표현을 가지는 얼굴-음성 교차 표현 공간을 학습하는 것이다. 이를 위해 VQGAN 손실함수를 통한 이미지 재구성 학습 과정에 얼굴 표현 공간을 위한 새로운 코드북을 학습하는 것이 아니라, 미리 학습된 오디오 코드북을 활용한다.

〈그림 3〉의 (a)는 제안한 방법의 학습 과정을 나타낸다. 우선 이미지 인코더를 통하여 얼굴 이미지 얼굴 특징 맵 $\hat{\mathbf{Z}}^{\mathbf{V}} = \mathbf{E}^{\mathbf{V}}(\mathbf{V})$ 를 얻는다. 그 후, 학습된 오디오 코드북을 사용하여 양자화 된 얼굴 특징 맵 $\mathbf{Z}^{\mathbf{V}} = \mathcal{Q}(\hat{\mathbf{Z}}^{\mathbf{V}})$ 를 얻는다. 이렇게 오디오 코드북을 사용하여 얼굴 특징 맵을 오디오의 불연속 공간(discrete space)으로 투영하여 얼굴-음성 교차 공간을 구성한다. 양자화 된 얼굴 특징 맵 $\mathbf{Z}^{\mathbf{V}}$ 은 이미지 디코더 를 통하여 이미지 스페이스로 복원되고, 이미지 디코더의 출력은 입력 얼굴 이미지를 복원하게 학습이 진행된다. 이미지 인코더와 디코더 학습을 위해 수식 (3)에서 두 번째 텀(codebook loss term)을 제외하고 이미지 인코더와 디코더 학습을 위한 손실함수로 사용한다 (수식 5).

$$\mathcal{L}_{recon}(E, G) = \left\|\hat{\mathbf{X}} - \mathbf{X}\right\|^2 + \|\mathrm{sg}\,[Z] - E(\mathbf{X})\|_2^2. \tag{5}$$

그림 3. 얼굴-음성 교차 공간 학습 도식도



(a) Audio-aligned Vision Learning (b) Vision-Audio Contrastive Learning

* (a) VQGAN 손실함수와 대조 학습을 통한 얼굴-음성 교차 공간 학습. (b) 얼굴-음성 대조 학습. 대응되는 얼굴 특징 맵과 음성 특징 맵은 유닛 구에서 서로 가깝게, 대응되지 않은 특징 맵들은 서로 멀게 학습한다.

** 출처: 저자 작성

이렇게 하나의 코드북을 사용하여 두 모달의 표현 공간을 제한(regularization)하는 것을 통해, 얼굴-음성 표현 공간을 구성할 수 있다. 그러나 학습 초기에 서로 다른 두 모달 간의 특징 맵이 큰 차이를 보이기 때문에 각 모달이 코드북의 전체 코드를 활용하는 것이 아니라, 일부 코드만 사용할 가능성이 있다. 이러한 현상은 코드북을 하나가 아닌 두 개를 나누어 학습한 것과 유사하며, 결과적으로 교차 표현 공간 학습에 실패한다. 이를 코드북 붕괴(codebook collapse)라 칭하며, 이러한 코드북 붕괴 문제를 해결하기

위해 벡터 양자화되기 전 두 모달의 표현 공간을 강제로 정렬(align) 한다. 서로 다른 모달 간의 공간을 정렬하기 위해 대조 학습(contrastive learning)이 널리 사용(Oord et al, 2018; Wu et al, 2018; Chen et al, 2020; He et al, 2020)되며 , 두 모달을 정렬하기 위해 InfoNCE(Oord et al, 2018) 손실함수를 사용한다. InfoNCE는 많은 오디오-비전 표현 학습에 사용(Afouras et al, 2020; Chen et al, 2021; Chen et al, 2021)되고 있으며 , 아래 수식 6과 같이 정의된다.

$$\text{InfoNCE}(\mathbf{x}_i, \{\mathbf{y}\}_{j=1}^{N}) = -\log \frac{\exp(-f(\mathbf{x}_i, \mathbf{y}_i)/\tau)}{\sum_{j=1}^{N} \exp(-f(\mathbf{x}_i, \mathbf{y}_j)/\tau)}, \tag{6}$$

$\mathbf{x}, \mathbf{y}$ 는 동일한 차원을 가진 각각의 뷰(view)로부터 얻은 벡터 ($\mathbf{x}, \mathbf{y} \in \mathbb{R}^C$), $f(\cdot, \cdot)$는 코사인 유사도(cosine similarity), 그리고 τ 는 temperature 인자를 나타낸다. InfoNCE는 두 뷰의 양의 쌍(positive pair) 벡터들을 C-차원의 단위 구(unit sphere)에서 거리가 가깝게, 음의 쌍(negative pair) 벡터들은 거리가 멀게 학습된다. 오디오-비전 표현 학습에서 대조 학습은 오디오와 비전을 각각의 뷰로 정의하고, 이미지와 이미지에 대응되는 오디오 쌍을 양의 쌍(positive pair), 그 외 대응되지 않은 쌍들을 음의 쌍(negative pair)으로 정의한다. 화자의 얼굴 이미지와 화자의 음성을 각각의 뷰로 정의하고, 서로 대응되는 화자 이미지와 음성을 양의 쌍으로 그 외의 랜덤한 조합들을 음의 쌍으로 정의하여, 오디오-중심 정렬 손실함수 (audio-centric align loss)를 수식 7과 같이 정의한다.

$$\mathcal{L}_i^{\mathbf{A}} = \text{InfoNCE}(\mathbf{z}_i^{\mathbf{A}}, \{\mathbf{z}^{\mathbf{V}}\}_{j=1}^{N}), \tag{7}$$

이 때, $\mathbf{z}^{\mathbf{A}}$와 $\mathbf{z}^{\mathbf{V}}$ 는 각각 $\mathbf{Z}^{\mathbf{A}}$와 $\mathbf{Z}^{\mathbf{V}}$의 평균 벡터를 나타낸다. 추가로 손실함수의 대칭을 위해, 얼굴-중심 손실함수 (face-centric align loss) $\mathcal{L}_i^{\mathbf{V}} = \text{InfoNCE}(\mathbf{z}_i^{V}, \{\mathbf{z}^{\mathbf{V}}\}_{j=1}^{N})$ 를 정의한다. 마지막으로 정의한 두 손실 함수의 합이 미니-배치(mini-batch B)에 존재하는 모든 쌍에 대해 최소화하는 것으로 얼굴-음성 정렬 손실함수(Face-Audio align loss)로 정의한다.

$$\mathcal{L}_{align} = \mathbb{E}_B \left[\mathcal{L}_i^{\mathbf{A}} + \mathcal{L}_i^{\mathbf{V}} \right], \tag{8}$$

이때 $\mathbb{E}$은 평균 연산자 (expectation operator)를 뜻한다.

최종적인 이미지 인코더와 디코더의 손실함수는 λ를 가중치로 하는 수식 5와 수식 8의 가중치 합으로 정의된다.

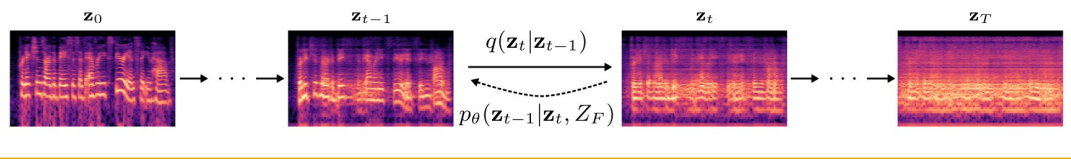
$$\mathcal{L}_{vision}(E^{\mathbf{V}}, G^{\mathbf{V}}) = \mathcal{L}_{recon} + \lambda \cdot \mathcal{L}_{align}\,. \tag{9}$$

이렇게 학습된 이미지 인코더와 디코더는 얼굴-음성 정렬 손실함수(수식 8)와 오디오 코드북을 통한 정규화 두 단계로 얼굴 특징 맵이 오디오 표현 공간으로 투영되기에 오디오와 잘 정렬된 얼굴 특징 맵을 추출할 수 있다.

2. Face-conditional Latent Diffusion Models

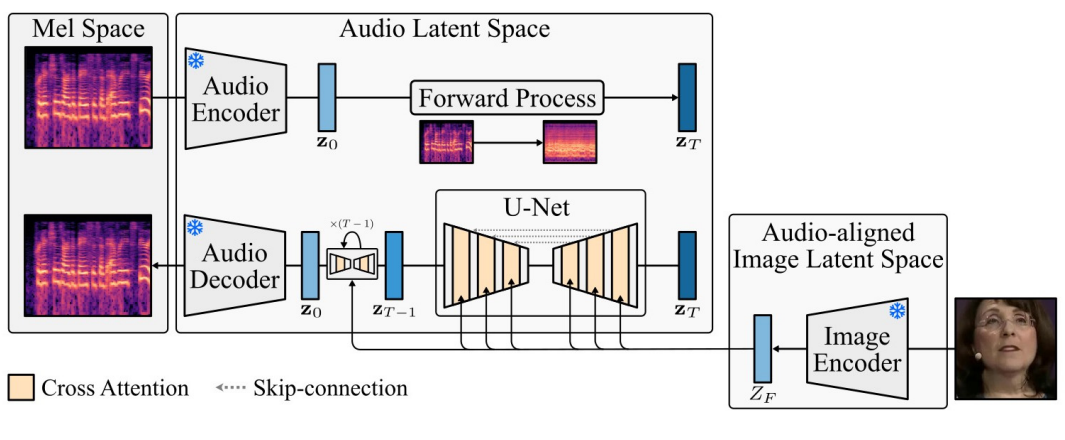
확산 모델 중 하나인 LDM(Rombach et al, 2022)을 사용하여 얼굴 조건부 음성 데이터 분포(face conditional speech data distribution, $p_{\text{data}}(\boldsymbol{z}_0|\mathbf{Z}_F)$)를 추정한다. 이때 $\boldsymbol{z}_0 \in \mathbb{E}^{n_z \times T/f \times F/f}$는 잠재 공간에서의 음성 샘플의 prior를 의미하며, T, F, f는 각각 멜-스펙트로그램의 길이, 주파수, 그리고 압축률을 의미한다. 확산 모델(Ho et al, 2020; Song et al, 2021)은 〈그림 4〉와 같이 원본 데이터 분포에 서서히 노이즈를 주입하여 가우시안 분포로 변형시키는 forward process와 노이즈를 추정하여 해당 노이즈를 서서히 제거하면서 데이터 샘플을 생성하는 reverse process로 구성된다.

〈그림 4〉 확산 모델의 Forward 및 Reverse 프로세스 가시화



* 출처: 저자 작성

〈그림 5〉 확산 모델의 학습 도식도



* 출처: 저자 작성

Forward process는 주어진 noise schedule ($\beta \in (0,1)$) 의 각 time step $t \in [1, \cdots, T]$ 마다 다음과 같이 정의된다.

$$q(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}) = \mathcal{N}(\boldsymbol{z}_t; \sqrt{\alpha_t}\boldsymbol{z}_{t-1}, (1-\alpha_t)\mathbf{I}), \quad \text{where } \bar{\alpha}_t := \prod_{u=1}^{t} \alpha_u \text{ and } \alpha_t := 1-\beta_t. \tag{10}$$

$$q(\boldsymbol{z}_t|\boldsymbol{z}_0) = \mathcal{N}(\boldsymbol{z}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{z}_0, (1-\bar{\alpha}_t)\boldsymbol{\epsilon}), \quad \text{where } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{11}$$

앞선 forward process를 통해 주입된 노이즈를 추정하는 것으로 조건부 확산 모델 $\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, Z_F)$ (conditional diffusion model)을 학습하며, 다음과 같이 손실함수를 정의한다.

$$\mathcal{L}^t(\theta) = w_t \cdot \mathbb{E}_{\boldsymbol{z}_0, \boldsymbol{\epsilon}} \|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, Z_F)\|_2^2 \tag{12}$$

Z_F는 학습 완료된 CrossVQGAN의 이미지 인코더를 사용하여 추출한 얼굴 이미지의 특징 맵을 나타내며, w_t는 각 노이즈 단계에 따른 손실함수의 가중치이다. 모든 노이즈 단계에 대해 1을 사용하는 simple objective(Ho et al, 2020] 대신 Min-SNR 가중치 방법 ($w_t = \min\left\{\frac{\gamma}{SNR(t)}, 1\right\}$) (Hang et al, 2023)를 사용한다. Min-SNR 가중치 방법은 각 노이즈 단계별로 신호 대비 잡음비 (SNR)을 활용하여 노이즈가 강할수록 큰 가중치를 부여하기에, 기존 simple objective에 비해 대략 3.4배 정도 빨리 확산 모델의 학습을 완료할 수 있다. 또한 조건부 확산 모델을 위해 CFG(Classifier-Free Guidance) (Ho et al, 2022)를 사용한다. CFG는 확산 모델에 컨디션을 주기위한 방법 중 하나로, 하나의 확산 모델에 대해 조건부 및 일반 확산 모델을 동시에 학습한다. 이를 통해 학습된 확산 모델은 특별한 컨디션 네트워크 없이 강력한 조건부 생성 능력을 가지며 수식 13과 같이 표현된다.

$$\tilde{\epsilon}_\theta(z_t, t, Z_F, \varnothing) = w_c \cdot \epsilon_\theta(z_t, t, Z_F) + (1-w_c) \cdot \epsilon_\theta(z_t, t, \varnothing), \quad \text{where } \varnothing \text{ is null condition.} \tag{13}$$

LDM 의 학습 후, 아래와 같은 reverse process를 통하여 음성 잠재 벡터를 생성할 수 있다.

$$p_\theta(\boldsymbol{z}_{t-1}|\boldsymbol{z}_t, Z_F) = \mathcal{N}(z_{t-1}; \mu_\theta(\boldsymbol{z}_t, t, Z_F), \sigma_t^2\mathbf{I}), \tag{14}$$

$$\mu_\theta(\boldsymbol{z}_t, t, Z_F) = \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{z}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, Z_F)), \tag{15}$$

$$\sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t. \tag{16}$$

최종적으로 LDM 을 통해 얻은 음성 잠재 벡터를 오디오 디코더 G_S를 통해 멜-스펙트로그램으로, 보코더(vocoder)를 사용하여 멜-스펙토그램을 waveform으로 변형시킨다.

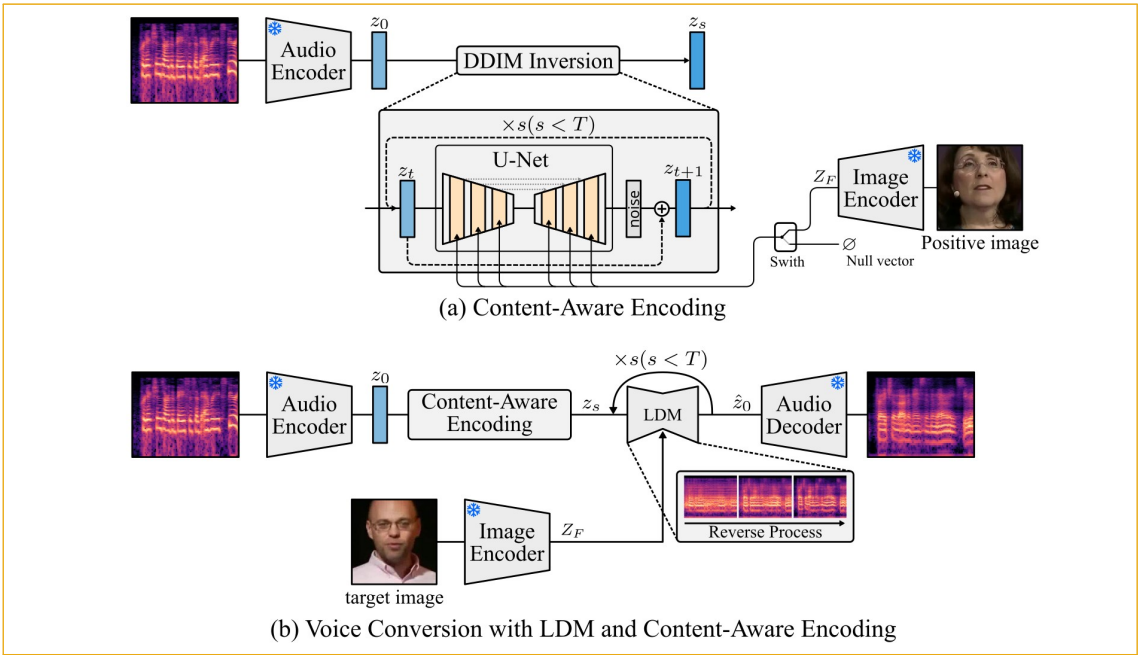
음성 생성 속도를 가속화하기 위해, Markov process인 reverse process를 non-Markov process로 변경하는 DDIM(Denosing Diffusion Implicit Model) (Song et al, 2021)을 사용한다. 추론 과정에서 DDIM(수식 17)을 통해 기존 1000 스텝의 디노이징과정으로 학습된 LDM 모델을 200 스텝으로 감소시켜 추론시간을 가속화한다.

$$z_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}} z_t + \left(\sqrt{\frac{1}{\alpha_{t-1} - 1}} - \sqrt{\frac{1}{\alpha_t} - 1} \right) \cdot \epsilon_\theta(z_t, t, Z_F) + \sigma_t \epsilon \qquad (17)$$

3. Content-Aware Encoding

학습된 LDM은 랜덤한 노이즈와 주어진 얼굴 잠재 벡터 Z_F로부터 음성 샘플 생성에 강한 성능을 보여준다. 그러나 랜덤한 노이즈로부터 음성을 생성하는 것은 생성된 음성의 발화 내용을 조절하는 것이 불가능하다. 발화 내용을 조절하기 위해, 이미지의 스타일을 변경하는 확산 모델(Meng et al, 2022)에 영감을 받아, 발화 내용만을 담고 있는 음성 노이즈를 추출하는 방법을 제안한다(〈그림 6〉 참고).

그림 6. (a) Content-Aware Encoding 도식도 (b) 제안 방법으로 얻은 발화 노이즈 특징 맵과 얼굴 특징 맵을 LDM의 입력으로 하여 음성을 생성하는 과정 및 변화된 멜-스펙트로그램 가시화



* 출처: 저자 작성

몇몇 이미지 생성 확산 모델 연구들은 reverse 프로세스 과정에서 노이즈 제거 중간 단계에서 의미론적 정보(semantic)가 결정되고, 그 이후 노이즈 제거 단계에서는 스타일과 디테일한 정보가 더해지는 것을 설명하였다(Meng et al, 2022; Choi et al, 2022). 제안한 오디오 생성 LDM 또한 이러한 현상이 있을 것으로 가정하고, DDIM inversion (수식 18) (Mokady et al, 2023)을 사용하여 일정 노이즈를 오디오 특징 맵에 가해 화자의 스타일은 제거되고 음성의 의미론적 정보인 발화 내용만 남은 노이즈 오디오 특징 맵을 생성한다(〈그림 6a〉 참고).

$$z_{t+1} = \sqrt{\frac{\alpha_{t+1}}{\alpha_t}} z_t + \left(\sqrt{\frac{1}{\alpha_{t+1} - 1}} - \sqrt{\frac{1}{\alpha_t} - 1} \right) \cdot \epsilon_\theta(z_t, t, Z_F) + \sigma_t \epsilon \tag{18}$$

이 때, 입력 음성에 대해 결정적 노이즈 (deterministic noise)를 얻기 위해, 수식 17의 스토케스틱 인자 (ϵ)을 0으로 설정한 deterministic DDIM inversion(수식 19)을 사용한다.

$$z_{t+1} = \sqrt{\frac{\alpha_{t+1}}{\alpha_t}} z_t + \left(\sqrt{\frac{1}{\alpha_{t+1} - 1}} - \sqrt{\frac{1}{\alpha_t} - 1} \right) \cdot \epsilon_\theta(z_t, t, Z_F) \tag{19}$$

원하는 발화를 가지고 있는 음성 z_0에 deterministic DDIM inversion을 통해 발화 정보만을 가지고 있는 노이즈 벡터 z_s를 얻을 수 있다. 이 때, 발화 노이즈를 얻기 위해 얼굴 정보에 대한 컨디션으로 발화 음성과 대응되는 얼굴 이미지(positive image) 또는 null vector를 사용할 수 있다. 또는 둘 다 사용하여 CFG를 통하여 Content-Aware Encoding이 가능하다. Content-Aware Encoding으로 발화 노이즈를 얻은 다음, 발화 노이즈와 원하는 얼굴 이미지(target image)의 특징 맵을 LDM의 입력으로 하여 입력 얼굴과 어울리면서 발화 내용을 조절하여 음성을 생성할 수 있다(〈그림 6b〉 참고).

이 때 인코딩할 스텝(s)에 따라 발화 내용만을 담고 있는 노이즈를 추출하거나 음성의 스타일과 발화 내용이 모두 있는 존재하는 노이즈를 추출할 수 있으며, 그리드 서치(grid search)를 통해 전체 디노이징 스텝의 70%를 인코딩할 스텝의 값으로 정하였다.

IV. 실험

1. 실험 설정

1.1 데이터 세트

CrossVQGAN과 Vocoder 학습을 위한 데이터로 VoxCeleb2(Chung et al, 2018)의 train/val을 사용했다. VoxCeleb2는 5,994명의 화자의 오디오와 비디오 쌍으로 구성되어 있는 데이터로 음성과 얼굴 이미지 간의 상관관계성을 학습하기에 적절하다. 이 두 네트워크의 평가를 위해 VoxCeleb2의 실험 세트(test set)를 사용하였다. LDM 학습을 위해, TED 비디오로부터 수집된 오디오-비디오 쌍 데이터 세트인 LRS3(Afouras et al, 2018)을 사용한다. LRS3는 VoxCeleb2와 다르게 발화내용에 대한 대본 또한 존재하기 때문에 변환된 목소리가 발화 내용을 정확하게 반영하고 있는지 평가하기에 용이하다. 또한 LRS3 데이터 세트에서 1.3초보다 짧은 데이터 쌍과 총 발화 시간이 10초 이하인 데이터는 학습용 데이터에서 제거하였다. 이런 필터링을 통해 최종적으로 사용된 학습 데이터는 14,114개의 발화와 2,007명의 화자로 구성되었으며, 테스트 데이터는 학습 데이터와 중복이 없는 412명의 화자로 구성되어 있다.

1.2 오디오 및 이미지 전처리

16kHz 샘플링 레이트를 가진 오디오를 64개의 빈으로 구성된 멜-스펙트로그램(64-bin mel-spectrogram)으로 변경하여 네트워크의 입력으로 사용했다. 멜-스펙트로그램으로 변경하기 위해, 윈도우 사이즈 64ms, 시프트 16ms, nfft 1024, 최소 주파수 0, 최대 주파수 8000, 64개의 Mel bin, 그리고 Hanning window을 사용하여 waveform을 멜-스펙트로그램으로 변형한다. 변형된 멜-스펙트로그램의 시간축이 256보다 짧다면 멜-스팩트로그램 우측에 제로-패딩을 수행하여 $R^{64\times256}$차원을 가진 멜-스펙트로그램을 학습 데이터로 사용한다. 얼굴 이미지의 경우, 비디오에서 하나의 프레임을 랜덤하게 샘플링하고 224×224로 해상도를 리사이즈하여 사용했다.

1.3 평가 지표

평가 지표로 객관적 지표(objective evaluation)를 사용한다. 객관적 평가를 위해, FAD(Frechet Audio Distance) (Bińkowski et al, 2020)과 KAD(Kernel Aduio Distance) (Bińkowski et al, 2020)를 사용하여 음성 변환 결과를 평가한다. 음성 변환 평가를 위해 LRS3 테스트 데이터세트에서 오디오와 타겟 화자를 랜덤하게 조합하여 50,000개의 샘플을 생성하여 생성된 샘플에 대한 각 평가지표의 평균값을 표기한다.

FAD[Bińkowski, 2020]는 이미지 생성 분야에서 사용하는 FID(Frechet Inception Distance) (Huesel et al, 2017)를 오디오 도메인에 맞게 변형한 것으로, 생성된 오디오 샘플과 정답 오디오 샘플들의 분포간의 유사도를 나타낸다. KAD(Bińkowski et al, 2020) 또한 이미지 생성 분야에서 사용하는

KID(Kernel Inception Distance) (Bińkowski et al, 2018)를 오디오 도메인에 맞게 변형한 것으로, 기존 FID에서 가우시안 분포로 데이터의 분포를 가정대신 비모수 분포 방식으로 두 샘플간의 분포를 비교한다. 세 개의 지표를 계산하기 위해 AudioSet(Gemmeke et al, 2017)로 학습된 PANN(Kong et al, 2020)을 특징맵 추출기(feature extractor)로 사용하고 FAD와 KDA는 PANN의 분별기(classifier) 이전의 2,048 차원을 가진 벡터(vector)를 사용하여 각각의 지표를 계산한다. 세 개의 지표는 다음과 같은 방식으로 계산된다. 1) 타겟 화자의 음성과 타겟 화자의 얼굴 이미지로 변형된 음성의 특징 벡터를 각각 PANN으로부터 추출하고, 2) 구해진 특징 벡터 쌍에 대해 FAD와 KAD를 계산하고 전체 테스테 세트에서의 평균값을 구한다. 생성 퀄러티 관점에서 FAD를 주요 평가 지표로 사용한다.

또한 억양, 강세, 리듬과 같은 특징을 나타내는 음율(prosody)과 기본 주파수(F0, Fundamental Frequency) 간에 강한 연관성이 있다는 것에 기반을 두어(Pell et al, 1999; Qian et al, 2020), 변환된 음성과 타겟 화자의 음성간의 F0를 비교하여 운율 유사도(prosody similarity)를 평가한다. 다음과 같은 단계로 운율 유사도를 계산한다. 1) 타겟 화자의 음성과 변환된 음성에서 음소 수준의 F0를 추출하고 (Maunch et al, 2014), 2) 각 음성 시퀀스에서 F0의 평균, 표준 편차, 기울기, 첨도(mean, standard deviation, skew, kurt)를 계산하고, 3) 쌍을 이룬 각 타겟 화자의 음성과 변환된 음성 간의 평균, 표준 편차, 기울기, 첨도의 차이를 계산하고 전체 테스트 세트 간의 차이를 평균값을 구한다.

음성 변화 결과가 원본 음성의 발화 내용을 잘 유지하는지 확인하기 위해, WER(Word Error Rate)을 채택하고 다음과 같은 방식으로 계산된다. 1) 변환된 음성의 발화 내용을 Whisper(Radford et al, 2023)를 통해 얻고, 2) 발화 스크립트와 추출된 발화 내용 간에 WER을 계산하고, 전체 테스트 세트에서의 평균값을 구한다. 발화 내용 유지 관점에서 WER를 주요 평가 지표로 사용한다.

1.4 비교 방법군

얼굴 이미지 기반의 음성 생성 및 변환의 선행 연구들의 공식 코드가 존재하지 않기 때문에(Lu et al, 2021), 제안 방법의 요소들에 대해 절제 평가(ablation study)를 수행한다. 우선 제안 모델의 상한 성능 (upper bound performance)을 계산하기 위해, 테스트 데이터 세트의 원본 오디오를 멜-스펙트로그램으로 변형시키고, 보코더를 통해 다시 waveform으로 변형시킨 결과물(GT Mel + Vocoder)에 대한 성능을 상한선으로 설정한다. 절제 평가 방법론으로는 1) 제안한 방법으로 학습한 오디오 표현 공간(〈그림 2〉 참고)과 사전 학습된 CLIP 이미지 인코더(Radford et al, 2021)를 사용하여 학습한 모델(Grad-FVC + CLIP), 2) 제안한 방법으로 학습한 오디오-이미지 표현 공간(〈그림 3〉 참고)를 사용하여 학습한 모델 Grad-FVC(Full)을 사용하여 제안 방법에 대해 평가한다. 이 때 Grad-FVC + CLIP의 CLIP 이미지 인코더는 DataComp(Gadre et al, 2023)로 사전 학습된 이미지 인코더를 사용한다.

2. 실험 결과

2.1 기존 기법과의 비교

〈표 2〉에 제안 기법과 기존 기법들과의 정량적 평가 수치를 나타내었다. 얼굴-음성 교차 표현공간을 구축하지 않고 음성을 생성하고 변환하는 모델 Grad-FVC + CLIP은 10이 넘는 FAD 수치로 좋지 않은 음성 변환 결과를 보여주고 있다. 또한 WER 또한 상한 성능의 5.28%보다 7배 가까이 큰 36.73%를 보이고 있다. 반대로 제안한 방법으로 얼굴-음성 교차 표현 공간을 구축한 모델 Grad-FVC(Full)은 기존 모델보다 모든 수치에서 좋은 성능을 보이고 있다. 특히 FAD와 WER 이 각각 5.91과 24.99% 감소하였으며, 이를 통해 교차 표현 공간을 구축하는 것이 얼굴 기반의 음성 생성 및 변환에 중요한 역할을 하는 것을 알 수 있다.

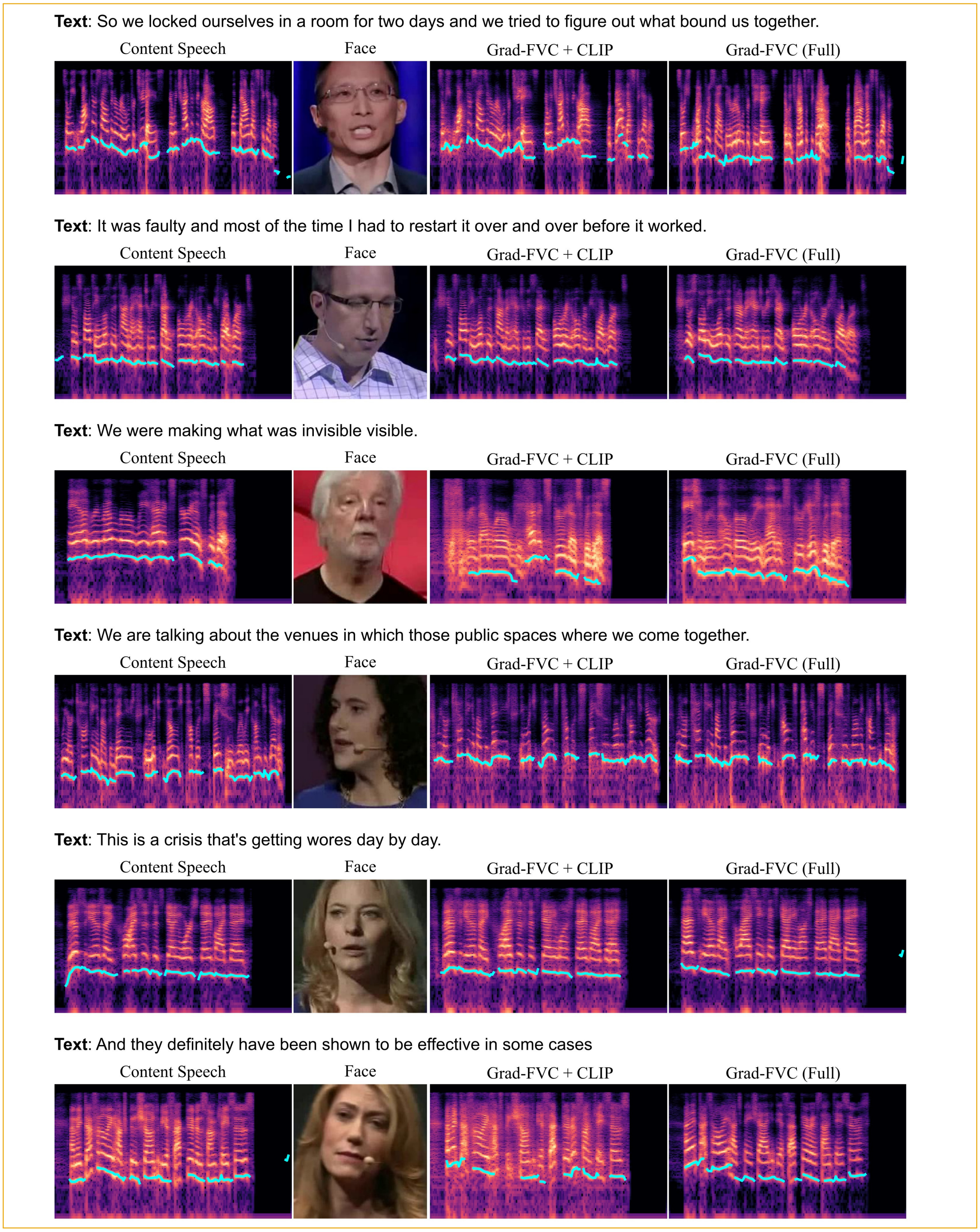
표 2. 제안 기법(Grad-FVC)와 기존 기법들과의 비교

Model	생성 퀄러티 관점		발화 내용 유지 관점
	FAD ↓	KAD ↓	WER(%) ↓
GT Mel + Vocoder	-	-	5.28
Grad-FVC + CLIP	19.15	0.24	36.73
Grad-FVC (Full)	13.24	0.19	11.74

* 출처: 저자 작성

정량적 평가 외에 원본 음성, 타겟 얼굴 이미지 그리고 변환된 음성의 멜-스펙트로그램을 〈그림 7〉에 나타내었다. 〈그림 7〉의 1,2,3행은 여성의 목소리를 남성 얼굴 이미지로, 4,5,6행은 남성의 목소리를 여성 얼굴 이미지로 각각 변환한 결과를 나타낸다. 음성의 높낮이 변화를 확인하기 위해 F0 주파수 또한 멜-스펙트로그램에 표시해두었다. F0 기준으로 Grad-FVC(Full)은 Grad-FVC + CLIP보다 남성과 여성 사이의 음높이 변화를 더 잘 표현하는 것을 확인할 수 있다. 특히, Grad-FVC + CLIP은 음높이 변화를 제대로 변화하지 못하고 그대로 유지하는 경향성을 〈그림 7〉의 2행, 4행, 그리고 5행을 통해 확인할 수 있다. 〈그림 7〉의 3행에서 Grad-FVC + CLIP 또한 여성의 음성을 늙은 남성의 목소리로 변환하는데 성공하였지만, 발화 내용을 유지하지 못하는 문제가 나타났다.

그림 7. 음성 변화 결과 비교



* 좌측 열부터 각각 컨텐츠 음성, 타겟 얼굴 이미지, 그리고 Grad-FVC+CLIP, Grad-FVC (full) 모델을 사용하여 음성을 변환한 결과를 나타낸다. Text는 각 행의 발화 내용을 나타낸다. 멜-스펙트로그램의 하늘색 라인은 F0 주파수를 의미한다.

** 출처: 저자 작성

〈표 3〉에 제안 기법 Grad-FVC와 기존 기법들과의 운율(prosody)을 비교하였다. F0 주파수를 구하고, 음성에 대항되지 않는 F0 주파수들은 제외하고 F0 분포의 특징값(moment)를 계산하였다. μ, σ, γ, K는 각각 평균, 표준편차, 왜도(skew), 첨도 (kurtosis)의 차이의 절댓값(absolute error)을 나타낸다. 가장 좋은 수치는 볼드체, 두 번째로 좋은 수치에 대해서는 밑줄로 표시하였다.

표 3. 제안 기법(Grad-FVC)와 기존 기법들과의 운율(prosody) 비교

Model	Pitch			
	$\mu\downarrow$	$\sigma\downarrow$	$\gamma\downarrow$	$K\downarrow$
Grad-FVC + CLIP	39.33	19.69	1.03	3.06
Grad-FVC (Full)	23.61	12.32	0.83	2.18

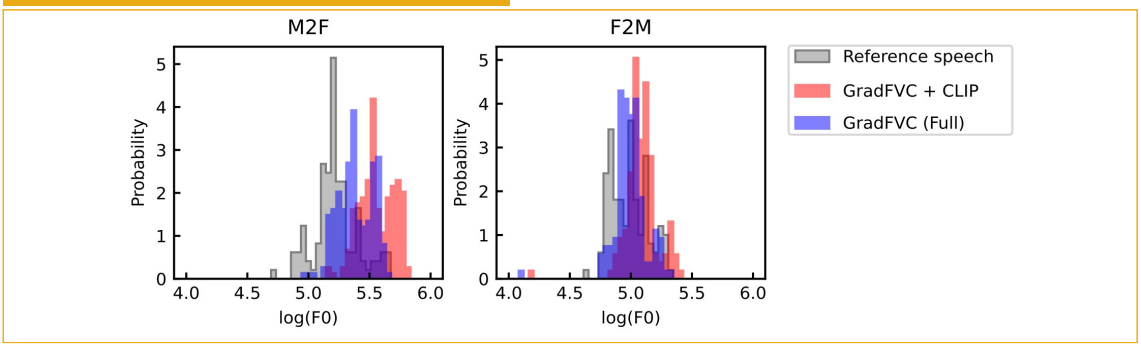
* 출처: 저자 작성

2.2 F0 분포 비교

F0가 음성의 스타일 중 하나인 운율과 연관성이 있다는 관점에서 변형된 음성과 타겟 화자의 음성 간의 F0 분포를 비교였고, 결과를 〈표 3〉에 나타냈다. 각 표의 값들은 생성된 음성과 레퍼런스 음성의 F0 분포들의 평균, 표준편차, 왜도(skew), 첨도(kurtosis) 차이 값을 나타낸다. 얼굴-음성 표현 공간을 구축하지 않은 Grad-FVC + CLIP은 F0 분포의 평균값 차이가 대략 40으로 이는 두 음성의 높낮이의 평균값이 40이 차이가 난다는 의미로 매우 많은 차이를 보이고 있다. 이에 반해 Grad-FVC(Full)은 비교적 낮은 값을 보이며, 얼굴 이미지로부터 성공적으로 음성의 스타일을 반영하는 특징 맵을 추출하고 있음을 확인할 수 있다.

추가로 남성에서 여성으로, 여성에서 남성음성으로 변화하는 결과에 대한 F0 분포를 M2F (남성→여성), F2M (여성→남성)으로 〈그림 8〉에 나타내었다. 〈그림 8〉을 통해, GradFVC + CLIP은 성별 간의 변화를 제대로 반영하고 있지 못함을 확인할 수 있다. 반면 Grad-FVC (Full)은 성별 간의 F0 분포를 어느 정도 잘 따라가고 있음을 확인할 수 있다. 이를 통해, 제안한 얼굴-음성 표현 공간이 얼굴에 대응되는 음성 스타일을 잘 추출하고 생성할 수 있음을 알 수 있다.

그림 8. 변환된 음성과 레퍼런스 음성간의 F0 분포 비교



* PYIN(Maunch et al, 2014)을 사용하여 F0 주파수를 구하고, 음성에 해당하지 않는 F0 주파수들은 제외하고 히스토그램으로 나타내었다.
** 출처: 저자 작성

2.3 Content-Aware Encoding 비교

제안한 Content-Aware Encoding이 발화 내용을 잘 유지하는지 평가하기 위해, 제안 방법으로 인코딩된 노이즈와 원본 음성에 소량의 랜덤 노이즈를 주입한 stochastic encoding 방법과 비교한다. 본 실험에서는 각 방법이 발화 내용을 얼마나 잘 유지하는 지를 평가하는 것이 목적이기 때문에 입력 얼굴 이미지를 원본 음성의 화자 얼굴을 입력으로 하여, 각 방법으로 인코딩된 노이즈를 원본 음성으로 복원하여 실험을 수행했다. 복원된 음성 신호를 평가하기 위해, 음성의 복원 퀄리티 측면과 발화 내용 유지 측면에서 평가를 수행하였으며, 복원 퀄리티를 측정하기 위해 PSNR과 SSIM을 평가 지표로 사용하고 발화 내용 유지의 정도를 측정하기 위해 WER(Word Error Rate) 외에 추가로 CER(Character Error Rate)를 평가 지표로 사용하였다.

〈표 4〉는 앞서 언급한 평가 지표들로 제안 방법과 기존 방법의 수치적 비교를 나타낸다. 발화 내용 유지 측면에서 stochastic encoding은 WER과 CER 둘 다 100% 가 넘는 오류율을 보여주고 있으며, 이는 발화 내용을 전혀 유지하지 못한다는 것을 의미한다. 이에 비해 제안한 Content-Aware enoding은 비교적 낮은 27.74 WER과 17.30 CER을 보여주고 있으며, 이는 제안한 방법이 효과적으로 발화 내용을 유지하면서 노이즈 벡터를 만들고 있다는 것을 보여준다. 복원 퀄리티 측면에서도 stochastic encoding이 발화 내용을 유지하지 못했기 때문에 낮은 PSNR과 SSIM 수치를 보여주고 있다.

표 4. 제안 방법(Content-Aware Encoding)과 기존 방법의 비교

Method	복원 퀄리티		발화 내용 유지	
	PSNR ↑	SSIM ↑	WER(%) ↓	CER(%) ↓
Stochastic Encoding	26.18	0.8897	148.09	106.96
Content-Aware Encoding	30.66	0.9386	7.74	17.30

* 출처: 저자 작성

V. 결론

본 연구는 얼굴 외형과 음성 간의 상관관계가 존재한다는 연구에 기반하여 얼굴 이미지로부터 음성의 스타일을 추출하고 음성을 생성하는 얼굴 기반의 음성 생성 및 변환 인공지능 모델을 제안한다. 얼굴 이미지로부터 음성의 스타일을 추출하기 위해, 하나의 코드북을 사용하는 shared codebook과 얼굴-음성 간의 대조 학습을 통하여 고품질의 음성 스타일 정보를 얼굴 이미지로부터 추출할 수 있었다. 또한 제안한 방법을 통해 적은 컴퓨팅 리소스만으로 얼굴-음성 교차 모달 학습이 가능했다. 추출된 얼굴-음성 정보로부터 음성을 생성하기 위해 확산 모델을 사용하였고 이를 통해 학습 데이터세트에 존재하지 않은 음성 샘플에 대해서 높은 품질의 음성을 생성 성능을 보였다. 발화 내용을 컨트롤하기 위해 음성 정보로부터 발화

내용만을 추출하는 Content-Aware Encoding을 제안하였으며, 이를 통해 발화 내용은 유지하면서 음성의 스타일을 변경할 수 있는 얼굴 이미지 기반의 음성 변환할 수 있었다. 다양한 실험과 평가를 통해 제안한 방법이 얼굴 이미지로부터 음성 스타일을 추출할 수 있음을 보였으며, 제안한 모델이 모달 간의 다양성이나 전반적인 음성 생성 품질 측면에서 우수하다는 것을 보였다. 본 연구를 통해 얼굴-음성 교차 모달 연구에 대한 더 많은 가능성을 조명한다고 확신한다.

저자소개 **이정식**(Jeong-Sik Lee)

• **학력**

영남대학교 전자공학 학사

• **경력**

現) 영남대학교 전자공학과 석사과정 재학 중

참고문헌

〈국외문헌〉

1) Afouras, T., Chung, J. S., & Zisserman, A. (2018). LRS3-TED: A Large-Scale Dataset for Visual Speech Recognition. arXiv preprint arXiv:1809.00496.

2) Afouras, T., Owens, A., Chung, J. S., & Zisserman, A. (2020). Self-Supervised Learning of Audio-Visual Objects from Video. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16, 208-224. Springer International Publishing.

3) Bengio, Y., Léonard, N., & Courville, A. (2013). Estimating or Propagating Gradients through Stochastic Neurons for Conditional Computation. arXiv preprint arXiv:1308.3432.

4) Bińkowski, M., Donahue, J., Dieleman, S., Clark, A., Elsen, E., Casagrande, N., … Simonyan, K. (2020). High Fidelity Speech Synthesis with Adversarial Networks. International Conference on Learning Representations.

5) Bińkowski, M., Sutherland, D. J., Arbel, M., & Gretton, A. (2018). Demystifying MMD GANs. International Conference on Learning Representations.

6) Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S. W., Fidler, S., & Kreis, K. (2023). Align your Latents: High-Resolution Video Synthesis with Latent Diffusion Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 22563-22575.

7) Brock, A., Donahue, J., & Simonyan, K. (2019). Large Scale GAN Training for High Fidelity Natural Image Synthesis. International Conference on Learning Representations.

8) Chen, H., Xie, W., Afouras, T., Nagrani, A., Vedaldi, A., & Zisserman, A. (2021). Localizing Visual Sounds the Hard Way. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 16867-16876.

9) Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A Simple Framework for Contrastive Learning of Visual Representations. In International Conference on Machine Learning (pp. 1597-1607). PMLR.

10) Chen, Y., Xian, Y., Koepke, A., Shan, Y., & Akata, Z. (2021). Distilling Audio-Visual Knowledge by Compositional Contrastive Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7016-7025.

11) Choi, J., Lee, J., Shin, C., Kim, S., Kim, H., & Yoon, S. (2022). Perception Prioritized Training of Diffusion Models. 2022 IEEE. In CVF Conference on Computer Vision and Pattern Recognition(CVPR), 101462-11471.

12) Chou, J. C., Yeh, C. C., & Lee, H. Y. (2019). One-Shot Voice Conversion by Separating Speaker and Content Representations with Instance Normalization. arXiv preprint arXiv:1904.05742.

13) Chung, J. S., Nagrani, A., & Zisserman, A. (2018). Voxceleb2: Deep Speaker Recognition. arXiv preprint arXiv:1806.05622.

14) Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A Large-Scale Hierarchical Image Database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

15) Dhariwal, P., & Nichol, A. (2021). Diffusion Models Beat Gans on Image Synthesis. Advances in neural information processing systems, 34, 8780-8794.

16) Gadre, S. Y., Ilharco, G., Fang, A., Hayase, J., Smyrnis, G., Nguyen, T., ... & Schmidt, L. (2023). DataComp: In Search of the Next Generation of Multimodal Datasets. arXiv preprint arXiv:2304.14108.

17) Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image Style Transfer Using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2414-2423.

18) Gemmeke, J. F., Ellis, D. P., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., ... & Ritter, M. (2017, March). Audio Set: An Ontology and Human-Labeled Dataset for Audio Events. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 776-780. IEEE.

19) Gong, Y., Rouditchenko, A., Liu, A. H., Harwath, D., Karlinsky, L., Kuehne, H., & Glass, J. R. (2022). Contrastive Audio-Visual Masked Autoencoder. In The Eleventh International Conference on Learning Representations.

20) Goto, S., Onishi, K., Saito, Y., Tachibana, K., & Mori, K. (2020, October). Face2Speech: Towards Multi-Speaker Text-to-Speech Synthesis Using an Embedding Vector Predicted from a Face Image. In INTERSPEECH, 1321-1325.

21) Gou, J., Yu, B., Maybank, S. J., & Tao, D. (2021). Knowledge Distillation: A Survey. International Journal of Computer Vision, 129, 1789-1819.

22) Hang, T., Gu, S., Li, C., Bao, J., Chen, D., Hu, H., ··· Guo, B. (2023). Efficient Diffusion Training via Min-SNR Weighting Strategy. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 7441-7451.

23) He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked Autoencoders are Scalable Vision Learners. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 16000-16009.

24) He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum Contrast for Unsupervised Visual Representation Learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 9729-9738.

25) Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Advances in neural information processing systems, 30.

26) Ho, J., & Salimans, T. (2021). Classifier-Free Diffusion Guidance. NeurIPS 2021 Workshop on Deep

Generative Models and Downstream Applications.

27) Ho, J., & Salimans, T. (2022). Classifier-Free Diffusion Guidance. arXiv preprint arXiv:2207.12598.

28) Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. Advances in neural information processing systems, 33, 6840-6851.

29) Hong, J., Kim, M., Yoo, D., & Ro, Y. M. (2022). Visual Context-Driven Audio Feature Enhancement for Robust End-to-End Audio-Visual Speech Recognition. arXiv preprint arXiv:2207.06020.

30) Huang, X., & Belongie, S. (2017). Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. In Proceedings of the IEEE International Conference on Computer Vision, 1501-1510.

31) Iashin, V., & Rahtu, E. (2021). Taming Visually Guided Sound Generation. arXiv preprint arXiv:2110.08791.

32) Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1125-1134.

33) Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). Putting the Face to the Voice': Matching Identity across Modality. Current Biology, 13(19), 1709-1714.

34) Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4401-4410.

35) Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8110-8119).

36) Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., & Plumbley, M. D. (2020). Panns: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 2880-2894.

37) Lee, J., Chung, J. S., & Chung, S. W. (2023, June). Imaginary Voice: Face-Styled Diffusion Model for Text-to-Speech. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

38) Lee, S. H., Kim, J. H., Chung, H., & Lee, S. W. (2021). Voicemixer: Adversarial Voice Style Mixup. Advances in Neural Information Processing Systems, 34, 294-308.

39) Liu, H., Chen, Z., Yuan, Y., Mei, X., Liu, X., Mandic, D., ... & Plumbley, M. D. (2023). Audioldm: Text-to-Audio Generation with Latent Diffusion Models. arXiv preprint arXiv:2301.12503.

40) Liu, H., Tian, Q., Yuan, Y., Liu, X., Mei, X., Kong, Q., ... & Plumbley, M. D. (2023). AudioLDM 2: Learning Holistic Audio Generation with Self-Supervised Pretraining. arXiv preprint arXiv:2308.05734.

41) Liu, Y., Albanie, S., Nagrani, A., & Zisserman, A. (2019). Use What You Have: Video Retrieval using Representations from Collaborative Experts. arXiv preprint arXiv:1907.13487.

42) Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic Gradient Descent with Warm Restarts. arXiv pre-

print arXiv:1608.03983.

43) Loshchilov, I., & Hutter, F. (2017). Decoupled Weight Decay Regularization. arXiv preprint arXiv:1711.05101.

44) Lu, H. H., Weng, S. E., Yen, Y. F., Shuai, H. H., & Cheng, W. H. (2021). Face-Based Voice Conversion: Learning the Voice Behind a Face. In Proceedings of the 29th ACM International Conference on Multimedia, 496-505.

45) Lu, H. H., Weng, S. E., Yen, Y. F., Shuai, H. H., & Cheng, W. H. (2021). Face-Based Voice Conversion: Learning the Voice Behind a Face. In Proceedings of the 29th ACM International Conference on Multimedia, 496-505.

46) Luo, S., Yan, C., Hu, C., & Zhao, H. (2023). Diff-Foley: Synchronized Video-to-Audio Synthesis with Latent Diffusion Models. arXiv preprint arXiv:2306.17203.

47) Mauch, M., & Dixon, S. (2014, May). pYIN: A Fundamental Frequency Estimator using Probabilistic Threshold Distributions. In 2014 ieee International Conference on Acoustics, Speech and Signal Processing(icassp), 659-663. IEEE.

48) McGurk, H., & MacDonald, J. (1976). Hearing Lips and Seeing Voices. Nature, 264(5588), 746-748.

49) Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.-Y., & Ermon, S. (2022). SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations. International Conference on Learning Representations.

50) Mokady, R., Hertz, A., Aberman, K., Pritch, Y., & Cohen-Or, D. (2023). Null-Text Inversion for Editing Real Images using Guided Diffusion Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6038-6047.

51) Nawaz, S., Saeed, M. S., Morerio, P., Mahmood, A., Gallo, I., Yousaf, M. H., & Del Bue, A. (2021). Cross-Modal Speaker Verification and Recognition: A Multilingual Perspective. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1682-1691.

52) Oord, A. V. D., Li, Y., & Vinyals, O. (2018). Representation Learning with Contrastive Predictive coding. arXiv preprint arXiv:1807.03748.

53) Pell, M. D. (1999). Fundamental Frequency Encoding of Linguistic and Emotional Prosody by Right Hemisphere-Damaged Speakers. Brain and language, 69(2), 161-192.

54) Popov, V., Vovk, I., Gogoryan, V., Sadekova, T., Kudinov, M. S., & Wei, J. (2022). Diffusion-Based Voice Conversion with Fast Maximum Likelihood Sampling Scheme. International Conference on Learning Representations.

55) Qian, K., Jin, Z., Hasegawa-Johnson, M., & Mysore, G. J. (2020). F0-Consistent Many-to-Many Non-parallel Voice Conversion via Conditional Autoencoder. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), 6284-6288. IEEE.

56) Qian, K., Zhang, Y., Chang, S., Yang, X., & Hasegawa-Johnson, M. (2019). Autovc: Zero-Shot Voice Style Transfer with Only Autoencoder Loss. In International Conference on Machine Learning,

5210–5219. PMLR.

57) Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning Transferable Visual Models from Natural Language Supervision. In International conference on machine learning, 8748–8763. PMLR.

58) Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. In International Conference on Machine Learning, 8821–8831. PMLR.

59) Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10684–10695).

60) Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved Techniques for Training Gans. Advances in Neural Information Processing Systems, 29.

61) Schuhmann, C., Vencu, R., Beaumont, R., Kaczmarczyk, R., Mullis, C., Katta, A., ... & Komatsuzaki, A. (2021). Laion-400m: Open Dataset of Clip-Filtered 400 Million Image-Text Pairs. arXiv preprint arXiv:2111.02114.

62) Smith, H. M., Dunn, A. K., Baguley, T., & Stacey, P. C. (2016). Matching Novel Face and Voice Identity using Static and Dynamic Facial Images. Attention, Perception, & Psychophysics, 78, 868–879.

63) Song, J., Meng, C., & Ermon, S. (2021). Denoising Diffusion Implicit Models. International Conference on Learning Representations.

64) Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based Generative Modeling through Stochastic Differential Equations. arXiv preprint arXiv:2011.13456.

65) Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818–2826.

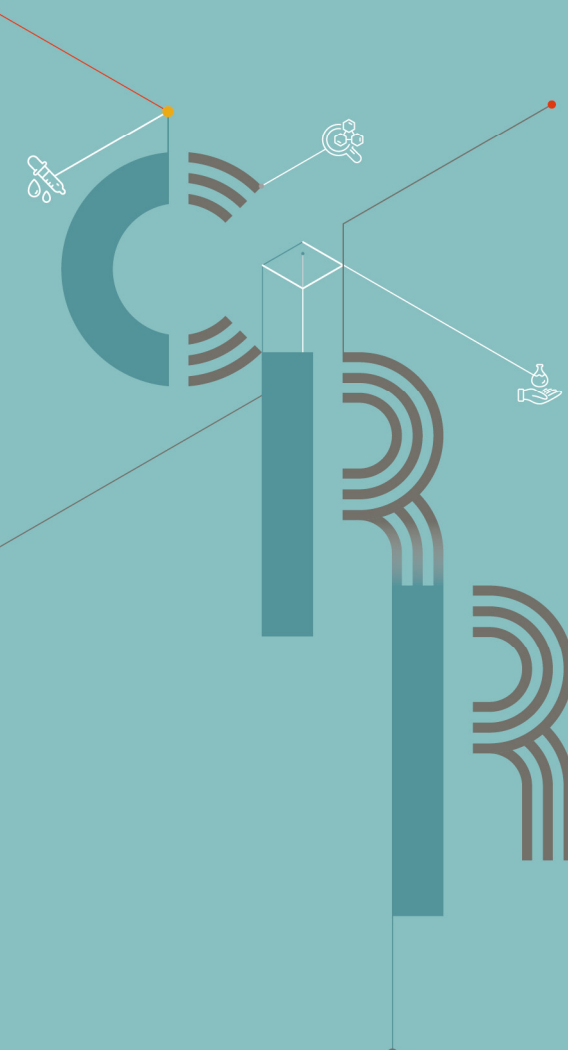
66) Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2016). Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022.

67) Wu, D.-Y., Chen, Y.-H., & Lee, H.-Y. (2020). VQVC+: One-Shot Voice Conversion by Vector Quantization and U-Net Architecture. Proc. Interspeech 2020, 4691–4695

68) Wu, Y., Chen, K., Zhang, T., Hui, Y., Berg-Kirkpatrick, T., & Dubnov, S. (2023, June). Large-Scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), 1-5. IEEE.

69) Wu, Z., Xiong, Y., Yu, S. X., & Lin, D. (2018). Unsupervised Feature Learning via Non-Parametric Instance Discrimination. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3733–3742.

70) Zhang, L., Rao, A., & Agrawala, M. (2023). Adding Conditional Control to Text-to-Image Diffusion Models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 3836–3847.



융합연구리뷰

Convergence Research Review

03

공통 데이터 모델(CDM)을 활용한 약물 치료 패턴 연구 동향

유보림(서울특별시보라매병원 데이터사이언스센터 부교수)
박은지(서울특별시보라매병원 데이터사이언스센터 박사 후 연구원)

03

공통 데이터 모델(CDM)을 활용한 약물 치료 패턴 연구 동향

유보림(서울특별시보라매병원)
박은지(서울특별시보라매병원)

I. 서론

데이터와 인공지능의 시대에서 특히 의료 빅데이터 분야는 지속적으로 발전하고 있으며, 이와 관련된 여러 새로운 기술과 응용 분야가 등장하고 있다. 의료 빅데이터는 환자가 병원에 방문하면서 생산되는 전자의무기록(EMR, Electronic Medical Records), 웨어러블 기기 등 장치로부터 수집되는 라이프로그(life-log), 건강보험 청구를 위한 보험 청구 자료(claims), 그리고 개인 유전체 데이터(genome)로 크게 나뉠 수 있다.

현재 대부분의 의료기관에서 전산화 된 통합 정보 시스템을 사용하여 환자 진료에 활용하고 있다. 병원의 정보 시스템의 대표적인 예로는 기존의 종이 차트에 기록되던 개인 환자의 진료정보를 전산화한 전자의무기록, 각종 검사 및 약물 처방 정보를 입력하여 전달하는 처방 정보 전달 시스템(OCS, Order Communication System), 의료 이미지와 영상을 위한 시스템(PACS, Picture Archiving and Communication System), 혈액 검사와 같은 각종 검사 정보를 처리하기 위한 검사정보시스템(LIMS, Laboratory Information Management System) 등이 있으며, 이외에도 환자의 예약을 관리하는 시스템이나 건강보험, 원무 정보를 처리하는 시스템 등 여러 종류의 전산 시스템이 하나의 병원정보시스템을 구성하는 구조로 운영된다.

이러한 병원정보시스템은 병원마다 다른 구조로 되어 있으며, 그로 인해 개별 의료기관마다 데이터 형식과 내용이 상이할 수 있다. 이러한 상황에서 여러 병원의 데이터로 임상 연구를 진행할 때에는 여러 제약이 생긴다. 이 문제점을 해결하기 위해 각 병원의 데이터를 하나의 공통적인 형식으로 변환하여 다양한 임상 다기관 연구를 진행하고자 하는 개념이 대두되었다. 이것을 공통 데이터 모델(CDM, Common Data Model)이라고 하며, 본 원고에서는 CDM 데이터에 대한 연구 사례, 특히 CDM 기반의 약물 치료 패턴 연구와 동향을 체계적인 문헌고찰 방법론을 활용하여 탐구해보기로 한다.

II. 공통 데이터 모델(CDM)

1. 공통 데이터 모델의 등장 배경

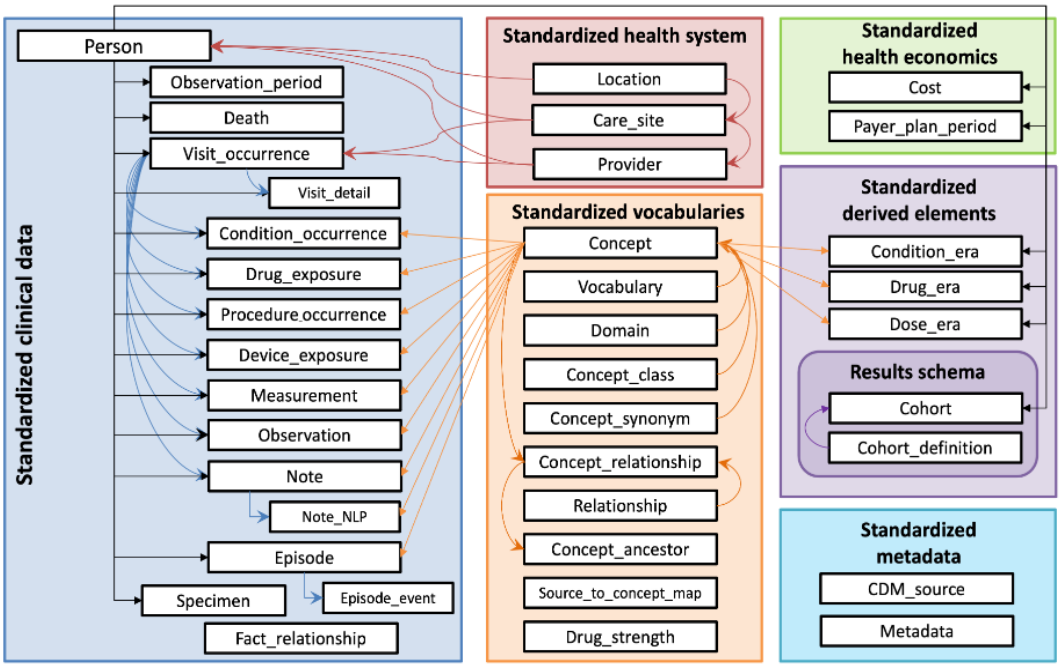
앞서 설명한 바와 같이, 병원의 임상 데이터를 활용한 다기관 연구 수행에 몇 가지 제약 사항이 있다. 의료 데이터는 데이터 구조, 형식의 이질성, 데이터의 질과 양 등 기술적인 어려움과 기관의 허락, 개인정보보호 문제 등 법적 문제 그리고 타인에게 제공하는 데이터가 자신에게 불리하게 사용될지 모른다는 두려움 등의 문제로 연구자 간 공유가 매우 어렵다. 현재까지 대부분의 기관 간 공동 연구는 극히 일부의 환자 데이터를 연구 주도 기관과 공유함으로써 진행하였는데, 한 번의 공동 연구를 위하여 막대한 노력과 시간, 자금이 들어가는 현실적인 문제와 개인 정보 공유를 제한하는 법적/제도적 문제들이 있다.

2. 공통 데이터 모델의 개념 및 동향

CDM은 공통 데이터 모델, 영어로 Common data model의 약자로, ICD-10, SNOMED CT, LOINC 등 보건의료 영역에서 사용하는 국제표준 용어를 기반으로 하는 관계형 데이터베이스 형태의 데이터 모델로, 각 의료기관에서 보유하는 데이터를 공통의 형식으로 변환한 데이터베이스이다. 다양한 종류의 CDM이 제안되어 왔으나, 그 중에서도 국제적으로 가장 많이 확산되어 활용되고 있는 모델은 2010년 다수의 이해관계자들이 모여 설립된 OMOP(Observational Medical Outcomes Partnership)에서 만든 CDM 모델이다. 현재는 OHDSI(Observational Health Data Sciences and Informatics)라는 비영리 단체에서 모델의 지속적 개발을 포함하여 CDM 데이터를 기반으로 누구나 사용할 수 있는 오픈소스(open-source) 도구를 개발하고 분산형 연구망 구축을 목표로 하는 국제 협력 연구 네트워크의 형태로 운영을 하고 있다. 특히, OHDSI의 여러 활동적인 커뮤니티들은 CDM 변환, 유지 보수 등을 위하여 서로 협력하고 있다. OHDSI는 다양한 데이터 세트를 CDM으로 변환할 수 있는 자원들을 제공할 뿐만 아니라 변환된 CDM을 활용할 수 있도록 100종 이상의 다양한 도구들을 제공한다. OHDSI의 가장 큰 특징 중 하나는 CDM 기반의 재현 가능한 연구(reproducible research)를 추구한다는 점이다. 연구의 투명성과 사전 정의를 위하여 모든 연구 프로토콜 및 분석 코드들을 깃 허브(Git Hub, 소프트웨어 개발 프로젝트를 위한 소스 코드 관리서비스)를 통해 공개한다. 공개된 자료에 대해 전 세계 연구자들에 의하여 신뢰성을 검토 받고 2차적 활용이 가능한 장점이 있다.

OHDSI의 CDM은 관계형 데이터베이스 구조를 가지는 여러 테이블들의 집합으로 구성된다. 개별 테이블은 환자가 병원을 방문함으로 인해 생성되는 임상 데이터를 중심으로 표준 진료 데이터(Standardized Clinical Data), 표준 용어(Standardized Vocabularies), 표준 보건 시스템(Standardized Health System), 표준 보건경제학(Standardized Health Economics), 표준 파생 요소(Standardized Derived Elements), 표준 메타 데이터(Standardized Meta-Data)의 영역으로 구분할 수 있다.

그림 1. CDM 데이터베이스 구조



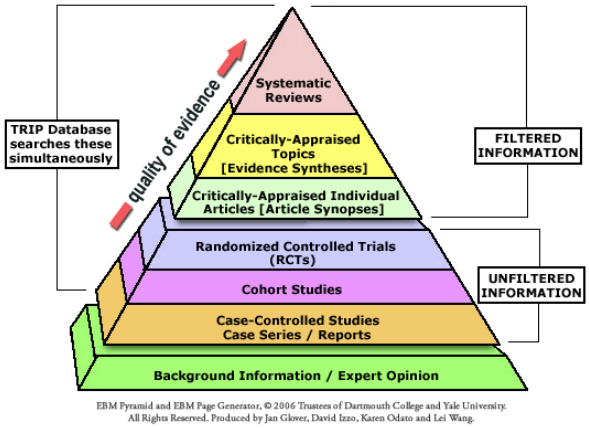
* 출처: OHDSI github

III. 근거기반 임상의사결정

1. 근거중심의학

근거중심의학이란 과학적인 방법을 통해 얻어진 표준화된 증거를 기반으로 의료 서비스 제공의 균일화를 꾀하는 현대의학의 과학적 방법론이다. 정리된 최신의 근거를 기반으로, 최선의 진료를, 모든 환자에게 제공하는 것이 근거중심의학의 목표이며 잘 수행된 임상 연구를 종합해서 통합된 결론을 도출하는 메타분석이나 체계적 문헌고찰을 최고 수준의 근거로 인정한다(김수영 외, 2011).

그림 2. 근거 수준 피라미드



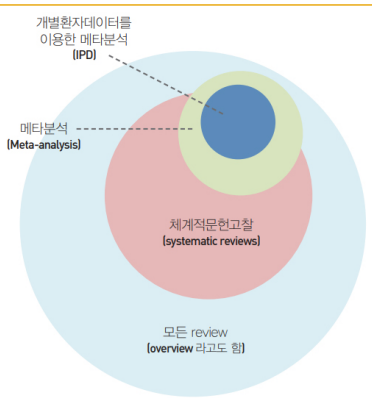
* 출처: Dartmouth College and Yale University

2. 체계적 문헌고찰 방법론

체계적 문헌고찰(systematic review)은 특정 연구 질문에 대해 최선의 가용 가능한 연구 결과를 모아 종합적인 결과를 도출하는 연구 방법이다. 체계적이고 포괄적인 문헌검색과 사전에 미리 정해진 포함 및 배제기준에 따른 문헌 선택 과정, 선정된 문헌에 대한 비뚤림 위험 평가(문헌의 질을 평가하는 것을 의미)등 객관적이며 과학적인 과정을 통해 수행된다(박병주 외, 2018).

일반적인 종설 또는 문헌고찰(review 또는 overview) 중에서도 체계적 문헌고찰은 비체계적 문헌고찰(narrative review)과 구분되는 개념이다. 또한, 체계적 문헌고찰과 메타분석(meta-analysis) 방법론이 혼용 되어 사용되기도 하지만, 정확하게는 〈그림 3〉에서와 같이 체계적 문헌고찰과 메타 분석이 같이 수행되는 연구도 있는 반면에 꼭 두 가지 연구방법이 동시에 사용되는 것은 아니다.

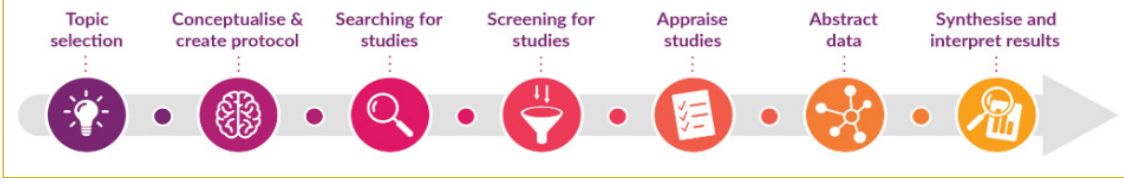
그림 3. 체계적 문헌고찰 개념도



* 출처: NECA 체계적 문헌고찰 매뉴얼

체계적 문헌고찰 연구의 수행 과정은 〈그림 4〉와 같이 표현될 수 있다. 가장 첫 번째 단계로, 연구를 수행하는 목적과 연구에 포함될 대상군을 확정하여 연구의 주제를 선정한다. 다음 단계에서는 연구 수행을 위한 프로토콜을 개발하고, 적절한 데이터베이스를 통하여 문헌을 검색한다. 검색된 문헌들을 대상으로 선정 및 제외 기준에 따라 문헌을 선정하고 분류하는 작업을 수행하며, 선정된 문헌의 질(quality)을 평가하는 비뚤림 위험 평가를 실시한다. 최종적으로 문헌이 선정되면 최종 선정된 문헌에 대한 자료를 추출하고, 마지막 단계에서는 메타분석 등 통계적 기법을 통하여 통합하여 추출된 자료들을 종합적인 결론을 도출하는 단계로 마무리한다.

그림 4. 체계적 문헌고찰 연구 수행 단계



* 출처: Karolinska Institutet University Library

IV. 약물 치료 패턴 연구 동향

1. 약물 치료·경로 패턴

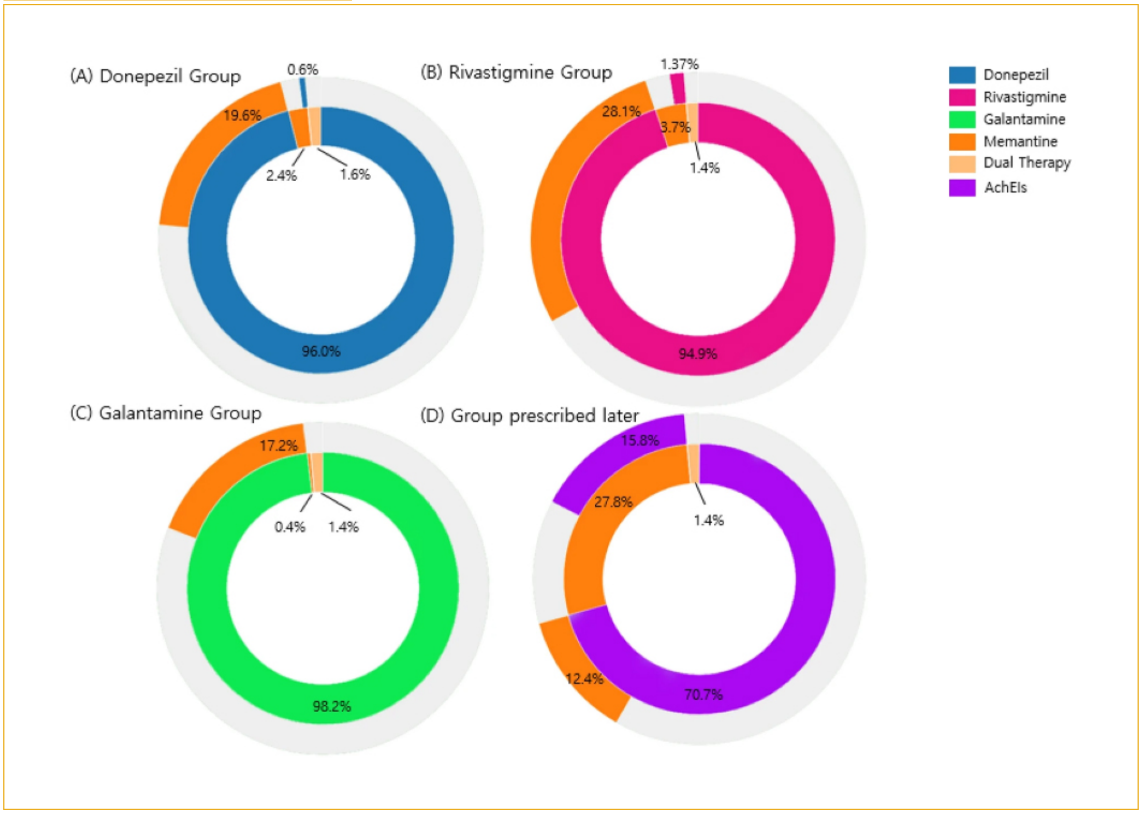
특정 환자군에서 1차 약제로 사용되는 약물은 효과가 좋다고 알려져 있고, 부작용이 적어 1차적인 치료제로써 주로 처방되는 약물이다. 1차 치료제를 투여 후 치료에 실패하거나 내성이 생겨 더 이상 사용이 어려운 경우에는 또 다른 치료제를 고려하게 되고 이때 사용되게 되는 치료제를 2차 약제라고 부른다. 환자의 임상적 특징이나 약제의 효능, 부작용, 비용 등을 전반적으로 고려하여 환자에게 적절한 치료 약제를 선택해야 한다.

그림 5. 1차 및 2차 약제

* 출처: 서북병원-서울시

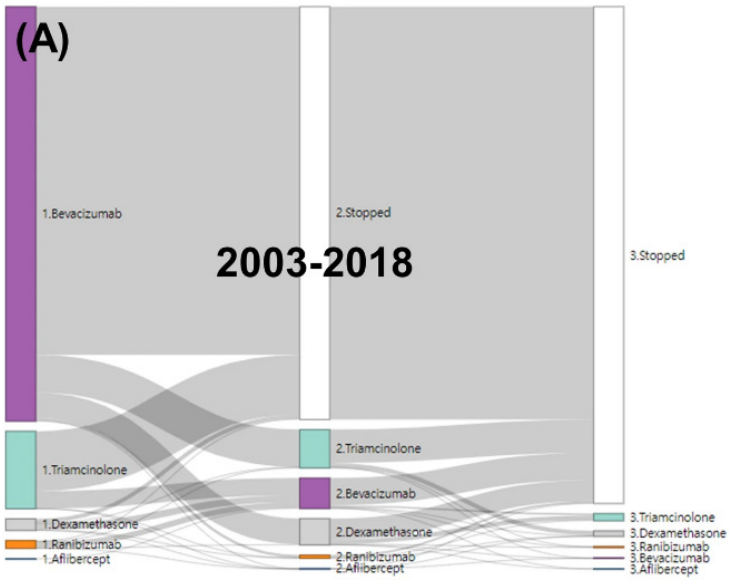
약물 치료·경로 패턴은 사전에 연구자가 정의한 관심 연구 대상 환자군에 대하여 약물 처방의 패턴을 조사함으로써 1차 치료 요법(first-line therapy), 2차 치료 요법(second-line therapy), 3차 치료 요법(third-line therapy) 등 특정 환자군의 치료 경로와 특성을 분석하기 위해 고안된 기술 통계적 방법이다. 선버스트 차트(Sunburst plot), 생키 다이어그램(Sankey diagram) 등 시각적인 그림들을 활용하여 치료의 경로와 패턴을 확인이 가능하며, 각 환자군을 그룹으로 나누거나 데이터베이스별로 소그룹(subgroup) 분석을 통해 그룹별 차이를 비교하는 것도 가능하다.

그림 6. Sunburst plot 그림 예시



* 출처: Byun et al.(2022)

그림 7. Sankey diagram 그림 예시



* 출처: Mun et al.(2022)

임상 현장에서 환자에게 투여되는 약물의 처방 및 치료 패턴을 분석하여 치료 행태를 파악하는 것은 의료진의 적절한 치료법 선택과 선호도에 대한 통찰력을 제공하며 현재 의료기관에서 사용되고 있는 치료 관행을 파악하는데 도움을 준다. 뿐만 아니라 관심 환자군에 대한 치료 이용 현황의 근거를 마련할 수 있고, 향후 질병 대상군에 대한 진료 지침을 제공하는데 활용될 수 있다.

그림 8. 당뇨병 진료지침 표지

* 출처: 대한당뇨병학회 2023

2. 체계적 문헌고찰 방법론을 통한 약물 치료 패턴 연구 동향 분석

2.1. 문헌 선정 기준

- 문헌 검색 데이터베이스 : PubMed, EMBASE 의학문헌 DB
- 문헌 검색 최종 수행일 : 2023-08-21
- 문헌 포함 기준(inclusion criteria)
  - OMOP, OHDSI, CDM(Common data model)을 이용하여 분석을 수행한 문헌
  - 치료 패턴(treatment patterns), 치료 경로(treatment pathways), 선버스트 차트(sunburst plot)를 결과로 제시한 문헌
  - Characterization을 연구 주제로 한 문헌
- 문헌 제외 기준 (exclusion criteria)
  - 2010년 이전 문헌

2.2. 문헌 검색어 및 결과

(1) PubMed

Search	Query	Results
#1	(OMOP[Title/Abstract] OR OMOP[Text Word]) OR (OHDSI[Title/Abstract] OR OHDSI[Text Word]) OR (CDM[Title/Abstract] OR CDM[Text Word]) OR(Common data model[Title/Abstract]) OR (Common data model[Text Word])	2,320
#2	"treatment outcome"[MeSHTerms] OR "drug prescriptions"[MeSHTerms]	1,283,383
#3	(treatment*[Title/Abstract]) OR (treatment*[Text Word]) OR (prescription*[Title/Abstract]) OR (prescription*[Text Word])	6,235,215
#4	(pathway*[Title/Abstract]) OR (pathway*[Text Word]) OR (pattern*[Title/Abstract]) OR (pattern*[Text Word]) OR (sequence*[Title/Abstract]) OR (sequence*[Text Word]) OR (sunburst[Title/Abstract]) OR (sunburst[Text Word])	4,568,255
#5	#3 and #4	714,180
#6	(characteri*[Title/Abstract]) OR (characteri*[Text Word])	3,810,505
#7	#1 AND (#2 OR #5 OR #6)	496
#8	#6 Filters: from 2010 – 2023	396

(2) EMBASE

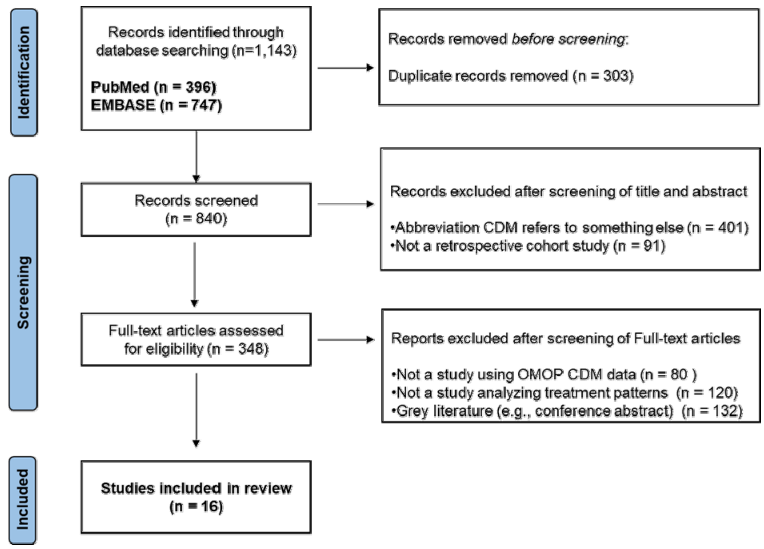
Search	Query	Results
#1	omop:ti,ab,kw OR ohdsi:ti,ab,kw OR cdm:ti,ab,kw OR 'common data model':ti,ab,kw	3,245
#2	'treatment'/exp OR 'prescription'/exp	260,505
#3	treatment$:ti,ab,kw OR prescription$:ti,ab,kw	7,881,948
#4	pathway$:ti,ab,kw OR pattern$:ti,ab,kw OR sequence$:ti,ab,kw OR sunburst:ti,ab,kw	4,626,663
#5	#3 and #4	894,363
#6	characteri*:ti,ab,kw	4,696,943
#7	#1 AND (#2 OR #5 OR #6)	839
#8	#7 AND [2010-2023]/py	747

2.3. 문헌 선정 과정

문헌 포함 및 제외 기준에 따른 문헌 선정 과정은 PRISMA(Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 가이드라인(Page et al., 2021)에 따라서 수행되었고, 문헌 선택 흐름도는 〈그림 9〉와 같이 나타낼 수 있다.

PubMed와 EMBASE 2가지 의학문헌 데이터베이스(DB)를 통하여 총 1,143건의 문헌이 검색되었고, 중복제거 후 840건에 대한 문헌 선별(screening) 과정을 진행하였다. 1단계 선별과정으로 문헌의 제목 및 목차를 살펴보았다. 후향적 코호트(retrospective cohort, 질병 발생의 위험요인을 확인하기 위하여 특정 요인에 대한 노출군과 비노출군을 설정한 후 추적 관찰을 통해 자료를 수집, 분석하는 관찰 연구)가 아니거나, CDM 약어가 Common Data Model이 아닌 다른 용어를 지칭하거나, 이미지 데이터를 포함하여 CDM 자료를 활용한 연구가 아닌 경우인 총 2,351건의 문헌이 1단계 선별과정에서 제외되었다. 남은 348건의 문헌에 대하여 2단계 선별과정으로 원문(full-text article) 검토 과정을 거쳤다. 이중 OMOP CDM 데이터를 활용한 연구가 아닌 문헌 80건, 치료 패턴 분석을 수행한 연구가 아닌 문헌 120건, 학회 초록 등 회색문헌 132건을 제외하였다. 최종적으로 본 연구에서 선정된 문헌은 16건이었다.

그림 9. 문헌 선택 흐름도(PRISMA flow chart)

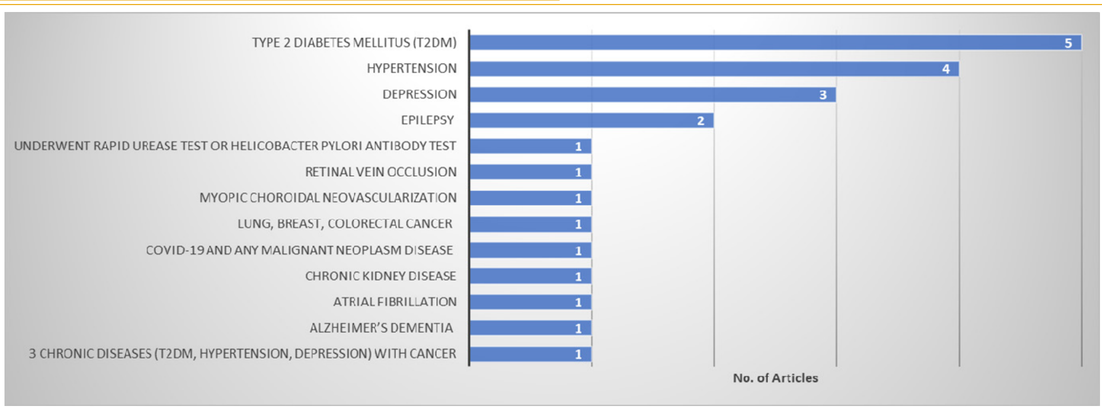


* 출처: 저자작성

2.4. 최종 선정 문헌

각 문헌에 대하여 논문 제목, 활용된 CDM 데이터베이스 명, CDM 데이터 기간, 코호트 기준 시점, 연구 대상 질환, 관심 대상 약물, 분석 방법, 분석 결과 그림, 결과 요약 내용을 추출하였다. 16건의 문헌에 포함된 연구 대상(target cohort) 질환군은 총 13개로, 그 중에 제 2형 당뇨병(type 2 diabetes mellitus)이 5건의 문헌에 포함되어 가장 많았으며, 고혈압(hypertension), 우울증(depression), 간질(epilepsy)이 각각 4건, 3건, 2건의 문헌에서 다루어져 연구 대상에 선정되었다.

그림 10. 최종 선정된 문헌에 포함된 연구 대상 질환 분포*

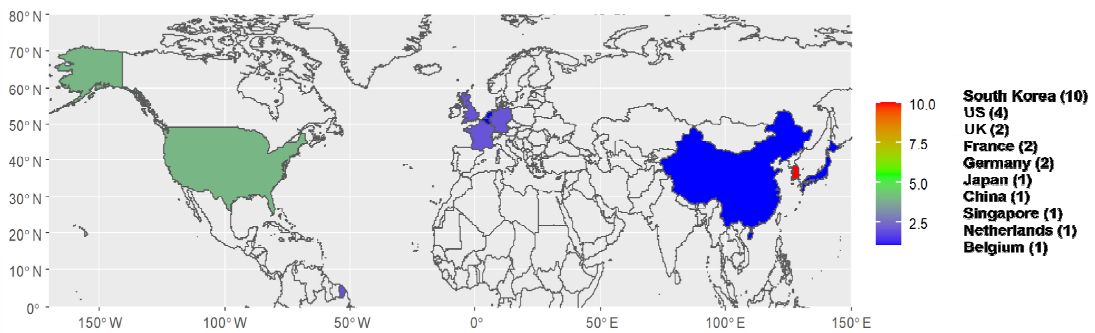


* 각 문헌은 중복해서 포함이 가능함

** 출처: 저자작성

최종 선정된 연구 문헌에 포함된 OMOP CDM 데이터베이스(database)를 국가별로 살펴보면 〈그림 11〉과 같다. 한국(South Korea)의 OMOP CDM 자료를 분석에 사용한 경우가 10건으로 가장 많았고, 다음으로는 미국(US)의 OMOP CDM 자료를 사용한 문헌이 4건, 영국(UK), 프랑스(France), 독일(Germany)의 OMOP CDM를 사용한 문헌이 각 2건으로 확인되었다.

그림 11. 최종 선정된 문헌에 포함된 국가별 OMOP CDM 데이터베이스

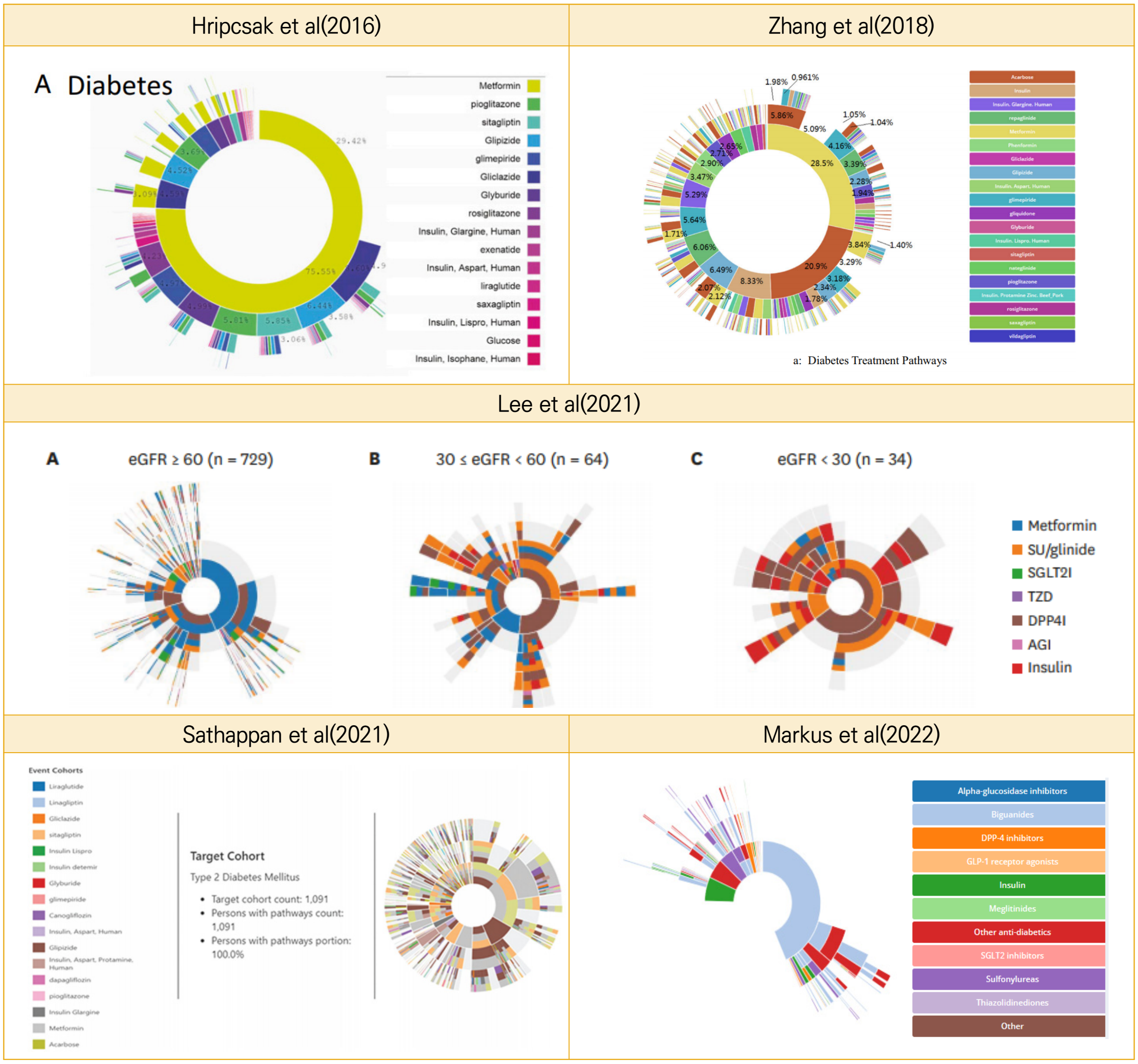


* 각 문헌은 중복해서 포함이 가능함

** 출처: 저자작성

16건의 문헌 중 가장 많은 연구 대상 환자군으로 포함되었던 제 2형 당뇨병 환자에 대한 약물 치료 패턴을 상세히 살펴보면 〈그림 12〉와 같다. 제 2형 당뇨병 환자에게 1차 치료제로 가장 많이 처방되는 약제는 메트포르민(metformin)으로 각 문헌별로 처방 비율에 차이가 있었다. 추정사구체여과율(eGFR, estimated Glomerular Filtration Rate, 신장에서 필터링 되는 혈액의 양을 추정하는 측정값) 수치를 기준으로 소그룹 분석을 실시한 Lee 외(2021) 문헌에서는 eGFR ≥ 60 인 그룹에서는 메트포르민(metformin)이 1차 치료제로 높은 비율로 사용되었으나 eGFR 수치가 낮은 타 그룹에서는 DPP4I (DiPeptidyl Peptidase-4 Inhibitors) 약제가 더 우세하였다. 메트포르민(metformin)은 비구아나이드(biguanides)계 약물(간에서의 포도당의 생성을 억제하고, 근육에서 포도당의 흡수 및 이용을 증가시키며, 소장에서 포도당의 흡수를 감소시켜 혈당을 조절)로써 Markus 외(2022) 문헌에서 1차 치료제로 가장 높은 비율로 사용된 비구아나이드(Biguanides) 약물의 대부분을 차지하는 것으로 추정된다. 2차 및 3차 치료 약제로 사용되는 약물의 사용 패턴은 기관별, 국가별로 상이함을 파악하였다.

그림 12. 제2형 당뇨병 환자에 대한 약물 치료 패턴 분석 문헌별 비교



* 출처: 저자작성

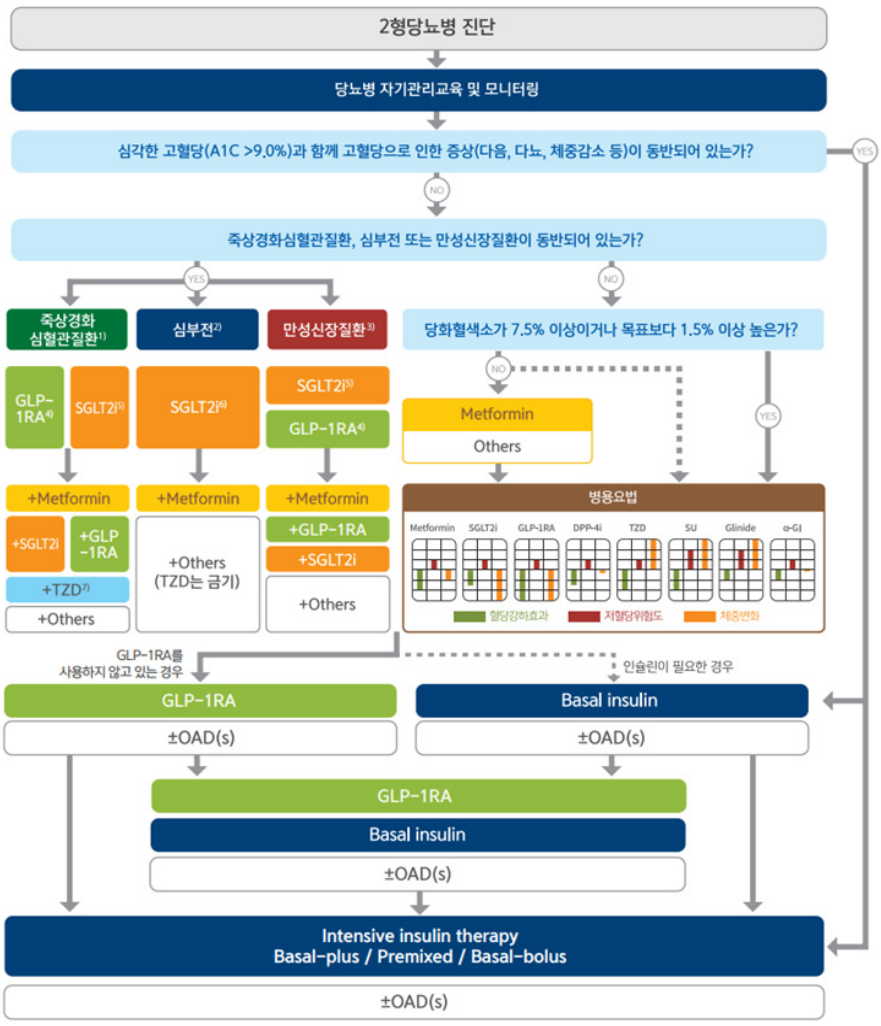
V. 약물 치료 패턴 연구 고찰

본 연구에서는 체계적 문헌고찰 연구 방법론을 사용하여 기존에 발표된 국내외 OMOP CDM 자료를 활용한 약물 등 치료 패턴 연구의 현황을 분석하였다. 2010년부터 2023년 8월 21일까지 발표된 국내외 OMOP CDM 자료를 활용한 약물 치료 패턴 연구는 총 16건의 문헌이었고, 이 중 한국의 OMOP CDM 자료를 활용한 문헌이 10건으로 가장 많았다. OMOP CDM 자료를 활용한 치료 패턴 연구 중 가장 먼저

시작된 연구(Hripcsak et al., 2016)에서 관심대상 질환은 제 2형 당뇨병, 고혈압, 우울증 3가지였으며, 이 연구에서 착안되어 치료 패턴 분석이 진행된 연구가 많았기 때문에 주로 제 2형 당뇨, 고혈압, 우울증 3가지 질환을 연구 대상으로 한 문헌이 많이 확인되었다.

제 2형 당뇨병 환자가 연구 대상으로 포함된 문헌이 5건으로 가장 많이 확인되었으며, 해당 연구들의 약물 치료 패턴 분석 결과에서 메트포르민(metformin)이 1차 치료 약제로 우세하게 사용되었음을 공통적으로 확인하였다. 이는 국내외 주요 당뇨병 진료 지침에서 당뇨병 환자의 초기 치료를 위한 경구혈당강하제로써 메트포르민(metformin)을 권고하고 있는 내용(Choi et al., 2023)과 실제 임상 데이터에서 사용 패턴이 유사하게 나타나는 것을 확인하였다 (〈그림 13〉 참고).

그림 13. 제 2형 당뇨병 약물치료 알고리즘



* 출처: 대한당뇨병학회 2023

2차 및 3차 치료 약제로 사용되는 약물의 사용 패턴은 기관별, 국가별로 이질적으로 나타났고, 서로 상이한 패턴을 보이는 것을 확인하였다. Sathappan 외(2021) 문헌에서는 여러 복합 약제를 치료 패턴 분석에 고려하였고, 싱가포르 내에 단일 상급종합병원 CDM 자료를 토대로 분석을 실시하여 다른 연구와의 이질성이 크게 나타났다.

약물 치료 패턴 연구에 사용된 약물들이 실제 임상 현장에서는 용량 및 용법에 따라 투약 방식이 상이하지만, 본 연구에 포함된 OMOP CDM 자료를 활용한 문헌들에서는 약물별 용량 및 용법의 차이를 반영하지 못한 한계점이 있다. 따라서, 추후에 약물의 용량 차이를 반영한 임상 자료가 확보된다면 더 나은 근거를 생성할 수 있을 것으로 기대한다. 또한, 본 연구에서 치료 패턴 분석에 포함된 OMOP CDM 자료는 대부분이 병원에 내원한 환자의 전자건강기록(EHR, Electronic Health Records)을 기반으로 하여 수집되고 가공하여 생성된 자료이고, 청구자료(claims database)를 기반으로 생성된 경우는 드문 편이었다. 한국의 OMOP CDM 자료 중에 청구자료를 기반으로 한 경우는 건강보험심사평가원(HIRA) COVID-19 환자 자료를 이용한 1건(Jeon et al., 2021)만 확인되었다. 앞으로 청구자료를 기반으로 한 OMOP CDM 변환 및 연구 데이터 활용이 증가한다면, 향후 청구자료와 EHR 자료를 기반으로 수집된 OMOP CDM 자료를 비교하여 추가 분석을 할 수 있을 것으로 기대된다.

Ⅵ. 결론 및 제언

지금까지 최근 활발하게 연구가 진행되고 있는 OMOP CDM 자료를 활용하여 약물 치료 패턴 분석 사례가 얼마나 발표되었는지에 대한 현황을 파악하였다. 임상 현장에서 환자에게 투여되는 약물의 처방 및 치료 패턴을 분석하여 관심 연구 질환에 대한 환자의 치료 행태를 파악하는 것은 현재 의료기관에서 사용되고 있는 치료 관행을 파악하는데 도움을 줄 뿐 아니라 의료진의 적절한 치료법 선택 등 임상 의사 결정에 대한 통찰력을 제공할 수 있다.

본 원고에서 다루었던 약물 치료 패턴 분석 연구에 포함된 CDM 데이터는 대부분이 병원에 내원한 환자의 EMR 데이터를 기반으로 하여 만들어진 자료인 반면, 건강보험 청구자료를 기반으로 만들어진 CDM을 활용한 연구 사례는 상대적으로 드물게 나타났다. 국내 CDM 자료 중에 건강보험 청구자료를 기반으로 했던 연구는 건강보험심사평가원의 COVID-19 특화 자료를 이용한 연구 1건(Jeon et al., 2021)이었다. 만약 건강보험 청구자료와 병원의 EMR 자료의 통합 분석이 가능하다면 향후 다양한 관점에서 약물 치료 패턴 분석이 가능할 것이다. 체계적 문헌고찰 검토를 통하여 도출한 지식을 토대로 임상 현장 및 제약 연구의 방향을 제시할 수 있고, 실제 임상 의사 결정에 활용할 수 있는 정책적 근거로 사용할 수 있을 것으로 기대한다.

저자소개 **유보림**(Borim Ryu)

• 학력

서울대학교 의과대학 의공학교실(의료정보) 박사
서울대학교 의과대학 의공학교실(의료정보) 석사
서울여자대학교 생명공학/컴퓨터공학 학사

• 경력

現) 서울특별시보라매병원 데이터사이언스센터 부교수
前) 서울특별시보라매병원 데이터사이언스센터 조교수
분당서울대학교병원 디지털헬스케어연구사업부 연구원, 선임연구원

저자소개 **박은지**(Eungee Park)

• 학력

서울대학교 의료정보학 박사
이화여자대학교 통계학 석사
이화여자대학교 통계학 학사

• 경력

現) 서울특별시보라매병원 데이터사이언스센터 박사 후 연구원

참고문헌

〈국내문헌〉

1) 김수영, 박지은, 서현주, 서혜선, 손희정, 신채민, 이윤재, 장보형, 허대석 (2011). NECA 체계적 문헌고찰 매뉴얼. NECA 연구방법 시리즈, 1-287
2) 대한당뇨병학회 (2023). 2023 당뇨병 진료지침 제8판. 서울메드쿠스
3) 박병주, 허대석, 이상무, 안형식, 지선하, 배은영 등 (2018). 근거기반 보건의료. 박영사

〈국외문헌〉

1) Ahmadi, N., Zoch, M., Kelbert, P., Noll, R., Schaaf, J., Wolfien, M., & Sedlmayr, M. (2023). Methods Used in the Development of Common Data Models for Health Data: Scoping Review. JMIR Medical Informatics, 11, e45116.
2) Bui, M. H., Lee, D. Y., Park, S. J., & Park, K. H. (2023). Real-World Treatment Intensity and Patterns in Patients With Myopic Choroidal Neovascularization: Common Data Model in Ophthalmology. Journal of Korean medical science, 38(23), e174.
3) Byun, J., Lee, D. Y., Jeong, C. W., Kim, Y., Rhee, H. Y., Moon, K. W., . . . Jang, J. W. (2022). Analysis of treatment pattern of anti-dementia medications in newly diagnosed Alzheimer's dementia using OMOP CDM. Scientific reports, 12(1), 4451.
4) Chen, R., Ryan, P., Natarajan, K., Falconer, T., Crew, K. D., Reich, C. G., . . . Hripcsak, G. (2020). Treatment Patterns for Chronic Comorbid Conditions in Patients With Cancer Using a Large-Scale Observational Data Network. JCO Clin Cancer Inform, 4, 171-183.
5) Choi, J. H., Lee, K. A., Moon, J. H., Chon, S., Kim, D. J., Kim, H. J., ... & Korean Diabetes Association. (2023). 2023 Clinical Practice Guidelines for Diabetes Mellitus of the Korean Diabetes Association. Diabetes & Metabolism Journal, 47(5), 575.
6) Chung, T. K., Jeon, Y., Hong, Y., Hong, S., Moon, J. S., & Lee, H. (2022). Factors affecting the changes in antihypertensive medications in patients with hypertension. Frontiers in Cardiovascular Medicine, 9.
7) Glover, J. (n.d.). Yale University Library Subject Guides. Evidence-Based Clinical Practice Resources. Pyramid.
8) Han, S., Son, M., Choi, B., Park, C., Shin, D. H., Jung, J. H., . . . Park, I. (2021). Characterization of Medication Trends for Chronic Kidney Disease: Mineral and Bone Disorder Treatment Using Electronic Health Record-Based Common Data Model. BioMed Research International, 2021.
9) Hripcsak, G., Ryan, P. B., Duke, J. D., Shah, N. H., Park, R. W., Huser, V., . . . Madigan, D. (2016). Characterizing treatment pathways at scale using the OHDSI network. Proceedings of the National Academy of Sciences of the United States of America, 113(27), 7329-7336.

10) Jeon, H., You, S. C., Kang, S. Y., Seo, S. I., Warner, J. L., Belenkaya, R., & Park, R. W. (2021). Characterizing the Anticancer Treatment Trajectory and Pattern in Patients Receiving Chemotherapy for Cancer Using Harmonized Observational Databases: Retrospective Study. JMIR Med Inform, 9(4), e25035.

11) Kim, H., Yoo, S., Jeon, Y., Yi, S., Kim, S., Choi, S. A., . . . Kim, K. J. (2020). Characterization of Anti-seizure Medication Treatment Pathways in Pediatric Epilepsy Using the Electronic Health Record-Based Common Data Model. Frontiers in Neurology, 11.

12) Lee, A., Yuan, Y., Eccles, L., Chitkara, A., Dalén, J., & Varol, N. (2022). Treatment patterns for advanced non-small cell lung cancer in the US: A systematic review of observational studies. Cancer Treatment and Research Communications, 100648.

13) Lee, K. A., Jin, H. Y., Kim, Y. J., Im, Y. J., Kim, E. Y., & Park, T. S. (2021). Treatment Patterns of Type 2 Diabetes Assessed Using a Common Data Model Based on Electronic Health Records of 2000-2019. Journal of Korean medical science, 36(36), e230.

14) Markus, A. F., Verhamme, K. M., Kors, J. A., & Rijnbeek, P. R. (2022). TreatmentPatterns: An R package to analyze treatment patterns of a study population of interest. medRxiv, 2022-01.

15) Martin, A., Bessonova, L., Hughes, R., Doane, M. J., O'Sullivan, A. K., Snook, K., ... & Harvey, P. D. (2022). Systematic Review of Real-World Treatment Patterns of Oral Antipsychotics and Associated Economic Burden in Patients with Schizophrenia in the United States. Advances in therapy, 39(9), 3933-3956.

16) Mun, Y., Park, C., Lee, D. Y., Kim, T. M., Jin, K. W., Kim, S., . . . Park, S. J. (2022). Real-world treatment intensities and pathways of macular edema following retinal vein occlusion in Korea from Common Data Model in ophthalmology. Scientific reports, 12(1), 10162.

17) Overhage, J. M., Ryan, P. B., Reich, C. G., Hartzema, A. G., & Stang, P. E. (2012). Validation of a common data model for active safety surveillance research. Journal of the American Medical Informatics Association, 19(1), 54-60.

18) Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. bmj, 372.

19) Sathappan, S. M. K., Jeon, Y. S., Dang, T. K., Lim, S. C., Shao, Y. M., Tai, E. S., & Feng, M. (2021). Transformation of Electronic Health Records and Questionnaire Data to OMOP CDM: A Feasibility Study Using SG_T2DM Dataset. Applied clinical informatics, 12(4), 757-767.

20) Seo, S. I., Kim, T. J., Choi, Y. J., Bang, C. S., Lee, Y. K., Lee, M. W., . . . Shin, W. G. (2022). Clinical Characteristics and Treatment Pathway of Patients Treated with Helicobacter pylori Infection-A Single Center Cohort Study Using Common Data Model. Korean Journal of Helicobacter and Upper Gastrointestinal Research, 22(3), 214-221.

21) Spotnitz, M., Ostropolets, A., Castano, V. G., Natarajan, K., Waldman, G. J., Argenziano, M., . . . Youngerman, B. E. (2022). Patient characteristics and antiseizure medication pathways in newly di-

agnosed epilepsy: Feasibility and pilot results using the common data model in a single-center electronic medical record database. Epilepsy and Behavior, 129.

22) Vora, P., Morgan Stewart, H., Russell, B., Asiimwe, A., & Brobert, G. (2022). Time Trends and Treatment Pathways in Prescribing Individual Oral Anticoagulants in Patients with Nonvalvular Atrial Fibrillation: An Observational Study of More than Three Million Patients from Europe and the United States. International journal of clinical practice, 2022, 6707985.

23) Zhang, X., Wang, L., Miao, S., Xu, H., Yin, Y., Zhu, Y., . . . Liu, Y. (2018). Analysis of treatment pathways for three chronic diseases using OMOP CDM. Journal of Medical Systems, 42(12).

〈기타자료〉

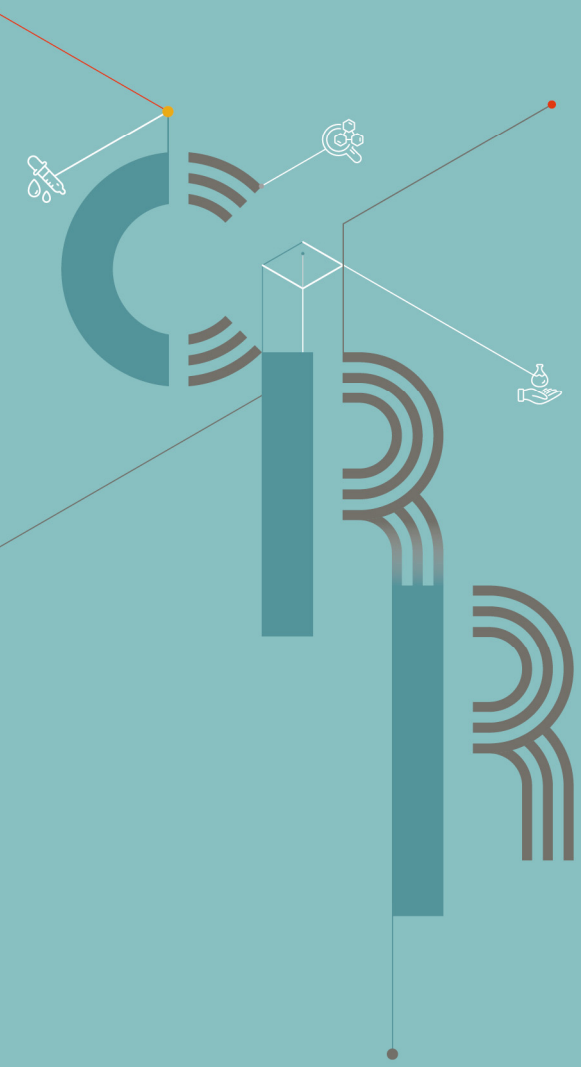
1) 서울특별시 서북병원 간호 1과. 다제내성결핵의 약물요법. https://sbhosp.seoul.go.kr/archives/14608

2) Karolinska Institutet University Library 사이트. Systematic reviews. https://kib.ki.se/en/search-evaluate/systematic-reviews

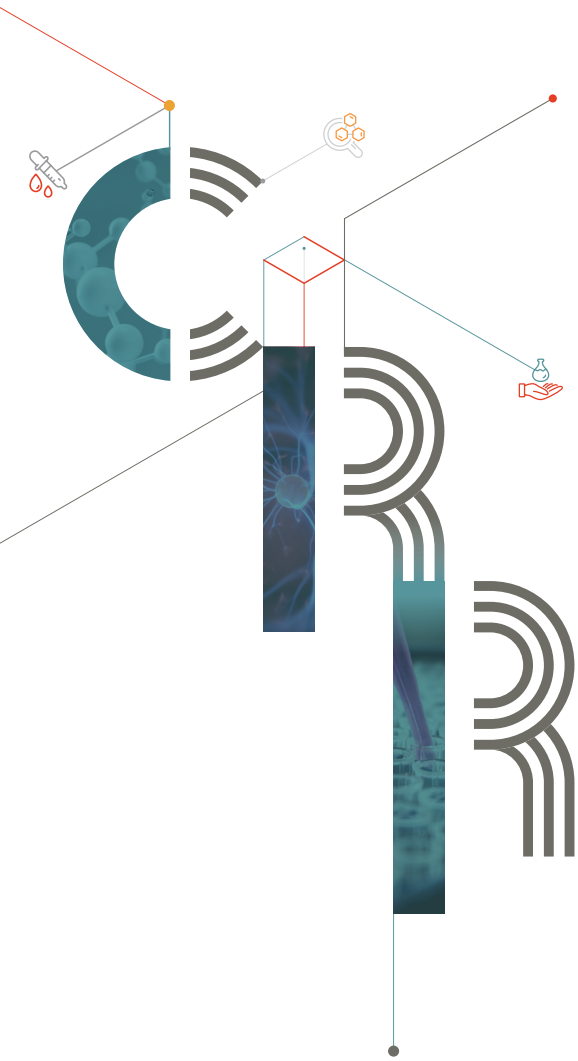
3) OHDSI(Observational Health Data Sciences and Informatics) 사이트. https://www.ohdsi.org/

4) OHDSI github 사이트. https://ohdsi.github.io/CommonDataModel/index.html

5) Treatment Patterns 분석 결과 shinyapp 예시 링크. https://mi-erasmusmc.shinyapps.io/TreatmentPatterns/



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



융합연구리뷰

Convergence Research Review

미래융합전략센터
Future Convergence Strategy Center

02792 서울특별시 성북구 화랑로 14길 5 **TEL.** 02.958.4973

ISSN. 2465-8456