

몸 없는 숨, 과잉된 몸: AI 보컬과 비트박스로 본 포스트휴먼 시대 목소리의 존재론적 의미

강 서 희
(한양대학교 강사)

I. 들어가며

“목소리는 어떻게 존재하는가?”라는 물음은 오랫동안 성악 담론의 자명한 전제 속에 가려져 있었다. 목소리는 ‘인간 신체’에서 발생하며, 언어, 감정, 선율을 전달하는 예술적 매개이자 주체의 정체성을 드러내는 직접적인 표현 방식으로 이해되었다. 롤랑 바르트(Roland Barthes, 1915-1980)는 목소리를 단순한 기호나 메시지의 전달 수단을 넘어 신체의 흔적과 감각적 질감이 실리는 ‘결’(grain)로 개념화하여 목소리의 촉각적·물질적 층위를 강조하였고,¹⁾ 월터 옹(Walter J. Ong, 1912-2003)은 구술문화의 맥락에서 목소리가 언제나 발화 주체의 현전(presence)을 전제한다는 점에 주목하며 목소리를 존재의 현존성과 긴밀히 결합된 청각적 체현으로 설명했다.²⁾ 2014년 본 연구자는

1) 바르트의 글은 『뒤희장죄』(*Musique en jeu*) 9호에 발표되었고, 1977년 『이미지-음악-텍스트』(*Image-Music-Text*)에 영문으로 번역되어 수록되었다. 같은 해 『뒤희장죄』에 “음악, 목소리, 언어”(La musique, la voix, la langue)라는 제목으로 다시 실렸다. Roland Barthes, “Le grain de la voix,” *Musique en jeu* 9 (1972), 57-63, 영문 번역은 *Image-Music-Text*, tans. Stephen Heath (London: Fontana Press, 1977), 179-189.

2) 옹의 저서는 1982년에 ‘Methuen & Co. Ltd’에서 초판되었고, 이후 2002년에 ‘Routledge’에서 재판되었다. 저서에서 ‘소리는 존재하는 동시에 사라져야만 존

“오페라에서 가사가 사라진 목소리의 의미와 역할”이라는 글을 통해 목소리가 언어의 의미를 전달하는 도구를 넘어 악기 자체로서의 기능을 수행하는 이중적 매체임을 보여주었다.³⁾ 그러나 이 선행 연구들은 여전히 ‘인간의 신체’를 중심으로 목소리의 의미를 탐색한다는 점에서 휴머니즘적 사고를 벗어나지 못했다.

문제는 오늘날 우리가 접하는 목소리는 더 이상 인간의 발화 기관에만 한정되지 않으며, 심지어 인간의 음성조차도 기존의 한계를 벗어나 기계음과의 경계를 넘나들고 있다는 것이다. 특히 음악 영역에서 일어나는 이러한 현상은 기존의 ‘성악’ 개념과 ‘언어적, 음악적 의미 전달자’로서의 목소리 개념을 와해한다는 점에서 주목할 만하다. 예를 들어 수노(Suno) AI 같은 합성 보컬은 실제 성대와 호흡이 없음에도 음성을 만들어내며, 반대로 비트박스는 한 사람의 신체만으로 드럼, 관악기, 전자 음향을 동시에 구현한다. AI와 비트박서 목소리의 서로 다른 양상은 목소리에 대한 기존의 정의를 재검토하게 만든다. 이들의 목소리는 디지털 전환 이후에 등장한 새로운 목소리 타입이며, ‘포스트휴먼’이라는 단어로 함축되는 현재의 상황을 반영한다. 여기서 ‘포스트휴먼’은 인간 중심의 고정된 위계를 가정하지 않으며, 신체와 기술, 유기체와 비유기체가 얹히는 자리에서 목소리를 하나의 사건으로 이해하려는 시각을 의미한다. 다시 말해, 포스트휴먼은 인간과 비인간, 유기체와 비유기체, 신체와 기술 사이의 선형적 위계나 고유성을 전제하지 않고, 목소리를 다중적 재료와 맥락을 관통하며 발생하는 사건(event)으로 사유하는 관점이다. AI 보컬은 물리적 신체가 사라진 자리에서 발화를 재현하고, 비트박스는 하나의 몸이 과잉된 방식으로 여러 가상 밴드 사운드를 구현한다. 두 경우는 서로 다른 조건에서 출발하지만, 결과적으로 목소리가 성립하는 경로가 다양하다는 점을 보여준다. 전자는 물리적 신체 흔적의 부재 속에서, 후자는 과

재한다’는 표현은 소리가 단순한 대상이 아닌 사건으로 발생하며 발화 순간이 주체의 현존을 불가피하게 동반함을 보여준다. 또한, 웅은 소리가 다른 감각과 달리 내면성(interiority)과 결부되어 있으며 이는 인간 의식과 의사소통의 내면성과 관련된다고 말한다. 이 역시 목소리가 발화 주체의 내적 현존성과 불가분적으로 연결됨을 강조한다. Walter J. Ong, *Orality and Literacy: The Technologizing of the Word* (London: Routledge, 2002), 42-43, 80-81.

- 3) 강서희, “오페라 속 가사가 사라진 목소리의 의미와 역할,” 『서양음악학』 17/3 (2014), 32-33.

잉된(확장된) 신체성 속에서 목소리를 드러내며, 그 결과 성악/기악, 어쿠스틱/디지털, 실연/가상이라는 기존의 이분법을 교란하고, 성악의 개념뿐만 아니라 목소리의 존재 조건 자체를 다시 생각하게 한다. 이러한 문제의식은 목소리를 경계 횡단의 사건으로 읽는 새로운 프레임을 제시한다.

AI 보컬 생성 기술은 실재하는 성대나 호흡, 미세한 불완전성 없이도 인간의 노래를 알고리즘적으로 모사하며, 생성된 목소리는 디션, 프레이징, 감정 표현에 이르기까지 인간 가수의 발성을 높은 정밀도로 재현한다. AI의 무신체적 발화와 인간 비트박서의 자기-보철화된 신체가 생성하는 목소리는, 기술 발전의 결과 그 이상으로, 목소리 현상이 근본적으로 전환되고 있음을 분명히 보여준다. 특히 AI 보컬은 물리적 신체의 흔적 없이 존재 없는 발화를 시뮬레이션함으로써, 웅이 제시한 구술성의 현전성, 즉 지금-여기에서 발화 주체가 말하고 있다는 감각을 해체한다. 이는 전통적인 구술-텍스트 이론의 철학적 기반을 전환시키며, 목소리와 존재 사이의 관계를 새롭게 사유하도록 이끈다.

반대로, 비트박스는 단일한 인간 신체가 드럼, 트럼펫, 베이스, 효과음 등 다성적인 전자 사운드를 구현하는 방식으로, 기계와 신체의 얹힘을 보여주는 대표적인 사례다. 비트박스의 어원 자체가 이 점을 잘 드러낸다. 원래 ‘비트박스’(beat box)는 1960년대에 등장한 드럼 머신을 가리키는 은어였다. 1964년에 발매된 드럼 머신인 윌리처 사이드맨(Wurlitzer Sideman)은 큰 나무 상자 속에 들어 있는 기계였는데, 당시 사람들은 이를 흔히 ‘비트박스’라고 불렀다.⁴⁾ 초기 힙합 문화의 주체들은 값비싼 드럼 머신이나 신시사이저를 대신하여 자신의 입과 신체를 활용해 비트를 만들어냈다. 이처럼 비트박스는 애초부터 신체와 기계의 기능이 얹혀 탄생한 장르이며, 신체의 과잉된 발화를 통해 기계음을 모사하며, 인간의 발성 기관이 디지털 사운드를 재현하는 기술적·음향적 기계로 작동할 수 있음을 보여준다. 더 나아가 비트박서가 밴드 사운드를 구사하는 행위는 성악적 악기로서 단선율을 연주하는 전통적인 목소리 연주 개념을 넘어 다층적 사운드를 생성하는 복합적 장치로의 전환이자 급진적 실천이기도 하다.

4) 서동재, “휴먼 비트박스를 활용한 라이브 루핑 주법,” 『대중음악콘텐츠』 1/2 (2017), 38.

AI보컬과 비트박서는 각각 신체의 부재와 과잉이라는 양극단에 위치하지만, 목소리를 하나의 주체나 정체성에 귀속된 현상으로 정의하기 어렵게 만든다는 점에서 한 지점에서 만난다. 실재하는 신체가 없는 목소리, 그리고 인간적인 발화를 비워낸 신체는 더 이상 전통적 이분법으로 해석될 수 없다. 이에 본 연구는 디지털 전환 이후, 즉 포스트휴먼 시대에 등장한 신체 없는 목소리(AI 보컬)와 신체로 구현된 비인간적 음향(비트박스)을 하나의 프레임에 두고 분석한다. 이를 통해 포스트휴먼 시대의 목소리가 신체, 기술, 공간, 청취 규범의 교차 지점에서 발생하는 사건임을 밝히고자 한다. 특히 사례 분석은 앞서 제시한 이론적 틀과 긴밀히 호응하도록 구성되며, 개념들이 실제 청취 경험과 어떻게 만나는지를 드러내는 데 중점을 둔다.

본 연구의 핵심 질문은 다음과 같다. 1) 디지털 시대의 보컬 사운드는 전통적인 음악 분류 체계(성악 대 기악)를 어떻게 전복하는가? 2) 실재하는 신체 없이 생성된 AI 보컬은 음악적으로 어떤 존재론적 정체성을 가지는가? 3) 인간의 발성 기관으로 디지털 밴드 사운드를 구현하는 비트박스는 기존 음악 장르 체계에 어떤 균열을 일으키는가? 4) 이러한 사례들은 ‘목소리’라는 개념 자체를 어떻게 재정의하게 하는가?

본문에서는 목소리 및 신체성과 관련된 선행 연구와 개념들을 정리하고, 이를 AI 보컬이 부른 노래와 비트박스의 연주 사례에 비추어 분석한다. 참고한 이론은 바르트의 ‘결’과 신체적 흔적, 옹의 현전성, 장 보드리야르(Jean Baudrillard, 1929-2007)의 시뮬라크르(simulacre), 하이디 파스트(Heidi Fast)와 밀라 티아이넨(Milla Tiainen)의 신체, 기술, 문화, 환경의 얽힘, 새 무얼 윌슨(Samuel Wilson)의 보철 개념, 브라이언 마수미(Brian Massumi, 1956-)의 정동, 수전 큐직(Susan Cusick, 1954-)의 발화 규범론이다. 이러한 이론들은 목소리 연구의 연대기적 서술이라기보다는, AI 보컬과 비트박스라는 구체적 사례 분석을 위한 개념적 프레임으로 기능한다. 동시에 각각은 역사적·철학적 맥락 속에서 목소리를 사유하는 중요한 출발점을 제공한다. 본 연구는 이러한 맥락을 존중하면서, 각 개념이 AI 보컬과 비트박스 분석 과정에서 어떻게 구체적으로 작동하는지를 밝히고자 한다.

사례 분석에서는 본 연구를 위해 수노 AI로 제작한 노래 《몸은 없는데 숨을 쉬네》를 대상으로, ‘존재 없는 발화’, ‘결의 결핍’, ‘시뮬라크르적 현전’을,

비트박서 wing(Wing)의 《도파민》(*Dopamine*)을 대상으로 ‘신체, 기술, 문화, 환경의 얽힘’, ‘결의 과잉’, ‘분산된 행위자성’을 탐색한다. 나아가 두 사례를 ‘현전-결-정동-배치’의 스펙트럼 위에 위치시키고, 목소리 존재론의 확장된 지형을 제시한다.

본 연구는 음악학 중 목소리 연구, 사운드·퍼포먼스 연구, 음악 존재론의 교차점에서 논의된다. 목소리를 언어·선율의 매개가 아닌, 신체적·음향적 현상으로 접근하는 최근의 연구 흐름과 발맞추면서, 마이크, PA,⁵⁾ 플랫폼 알고리즘과 같은 기술 장치가 목소리의 성격을 구성하는 과정을 다룬 선행 연구, 그리고 목소리를 단순한 기호나 표현을 넘어, 실재, 현전, 정동의 사건으로 탐구하는 철학적 논의와도 호응한다. 다시 말해, AI 보컬과 비트박스라는 서로 다른 연구 유형을 통해 “음악학에서 목소리를 어떻게 이해할 것인가?”라는 근본 질문을 재구성함으로써, 성악 연구를 확장하고 사운드 연구 및 철학적 존재론과의 접속을 시도한다.

목소리는 더 이상 인간 신체의 고유한 발화가 아니라, 기술, 신체, 공간, 규범이 교차하는 배치 속에서 발생하는 사건이다. 이러한 시각은 음악 존재론, 성악 연구, 사운드·퍼포먼스 연구의 교차 지점에서 새로운 지형을 열며, 디지털 포스트휴먼 시대의 목소리를 이해하는 하나의 통합적 틀을 제시한다.

II. 포스트휴먼 목소리의 이론적 틀: 결, 정동, 시뮬라크르, 가상블라주

1. 목소리의 결과 발화의 규범성: 톨랑 바르트와 수전 쿠직

전통적으로 목소리는 언제나 신체와 긴밀히 연결되어 있으며, 음악적 기호로 기능하거나, 가사와 선율을 전달하는 매개로 이해되어왔다. 그러나 목소리는 단순한 소리가 아니라 복합적인 문화적, 물질적 조건 속에서 수행되므로, 단순히 음높이, 음색, 말소리의 수준으로 환원하는 이해는 그것이 지닌 감각적 현

5) PA(Public Address system)는 공연장, 강당, 야외 행사 등에서 활용되는 확성·음향 시스템 전체를 의미한다. 믹서(mixer), 앰프(amplifier), 스피커(speaker), 모니터 등이 포함된다.

전성과 사회적 조건성을 충분히 포착하지 못한다. 이에 본 장에서는 롤랑 바르트의 ‘목소리의 결’ 개념과 수전 쿠직이 제기한 발화의 사회 문화적 조건을 중심으로, 목소리가 신체적 흔적이자 규범적 수행의 현장으로 확장된 과정을 살펴본다.

1.1. 바르트의 ‘결’ 개념과 감각의 층위

1972년 바르트는 “목소리의 결”(Le grain de la voix)이라는 제목의 글에서 ‘결’을 ‘노래하는 목소리가 있는 몸, 글을 쓰는 손, 연주하는 사지’라고 정의한다.⁶⁾ 이 개념은 목소리를 의미의 전달이나 음악적 음색이라는 틀 안에 가두는 기존의 접근을 넘어, 신체의 감각적 흔적으로서의 목소리를 포착하려는 시도이다. 바르트에게 목소리의 결이란 곧 발화 기관의 육체성이 청자의 몸에 도달하는 감각의 마찰이다. 다시 말해, 목소리는 말하거나 노래하는 자의 성대, 호흡, 입술, 혀, 치아 등의 물질적 요소가 만들어내는 촉각적 흔적으로서 존재하며, 청취자는 그 흔적을 듣는 동시에 몸으로 감지한다.⁷⁾

바르트의 논의는 목소리를 언어나 음악이라는 틀에 고정하지 않고, 그 사이에서 발생하는 감각적 흔적과 마찰에 주목하도록 이끈다. 이때의 목소리는 더 이상 언어의 하위 개념이나 선율의 전달자가 아니라, 의미화되기 이전의 차원에서 청자와 직접적으로 접촉하는 신체의 표면이 된다. 바르트는 프랑스 가수 샤를 팡제라(Charles Panzéra, 1896-1976)의 목소리를 분석 대상으로, 단순한 음정이나 해석의 문제를 넘어 가창의 과정에서 불현듯 솟구치는 신체적 감각의 압력에 주목하였다. 바르트는 이를 설명하기 위해 ‘페노송’(phenosong)과 ‘제노송’(genosong)이라는 구분을 제시한다. 페노송은 곡의 해석, 발음의 명료성, 스타일적 적합성 등 사회적으로 규범화된 음악적 표층을 가리키는 반면, 제노송은 기호화 이전에 드러나는 목소리의 신체적 흔적과 감각적 마찰을 뜻한다. 바르트가 관심을 두었던 것은 바로 이 제노송, 즉 언어와 의미를 넘어서는 ‘몸의 결’이었다.

그는 이 지점을 보다 분명히 하기 위해, 독일 가곡의 대가 디트리히 피셔디

6) Roland Barthes, *Image-Music-Text*, trans. Stephen Heath (New York: Hill and Wang, 1977), 188.

7) Barthes, *Image-Music-Text*, 185.

스카우(Dietrich Fischer-Dieskau, 1925-2012)와 팡제라를 비교한다. 피셔 디스카우의 노래는 완벽한 디션, 세밀한 뉘앙스 조절, 텍스트에 대한 치밀한 해석으로 대표된다. 바르트의 용어로 이는 페노송의 극치, 곧 의미 전달과 음악적 규범의 정점이다. 그러나 바르트에게 이러한 노래는 “너무 완벽하게 기호화된 목소리”로서, 청자에게 신체적 감각을 직접적으로 전하지는 못한다. 반면 팡제라의 목소리에서는 때로는 거칠고 불완전하게 들릴 수 있는 발성의 혼적 속에서 ‘결’, 즉 신체의 압력과 질감이 돌출하며 청자의 몸에 닿는다. 이 비교는 바르트의 목소리 개념이 단순히 ‘잘 부른 목소리’와 ‘못 부른 목소리’의 평가 구도를 넘어, 목소리를 기호적 층위와 신체적 층위로 분리하여 사유하려는 시도임을 보여준다. 목소리를 규범화된 의미 작용(페노송)과 물질적 감각과 신체성(제노송)으로 이분화하고, 후자에 주목함으로써 바르트는 목소리를 텍스트와 음악을 넘어서는 존재론적 사건으로 규정한다.

바르트는 ‘결’이 기표 이전의 감응성이라고 강조한다. 결은 목소리가 언어화되기 이전의 상태, 즉 주이상스(jouissance)에 가까운 상태에서 경험되는 신체의 잔향으로,⁸⁾ 단순한 쾌감(plaisir)과는 다른 차원의 경험이며, 해석되거나 통제되기 이전의 감각적 충돌이자 내면화되지 않은 소리의 파편이다. 이처럼 바르트의 ‘결’은 목소리를 지시적 기능으로부터 해방시키고, 청취자의 신체와 직접 마찰하는 감각의 접점으로 재구성한다.

이러한 관점은 AI 보컬이나 비트박스와 같은 새로운 목소리 유형에서도 중요한 통찰을 제공한다. 예컨대 AI가 생성한 목소리는 인간의 신체 없이도 음색과 디션, 감정 표현을 정밀하게 시뮬레이션할 수 있지만, 스피커를 통해 전달되는 AI의 목소리에는 성대의 미세한 떨림, 호흡의 비대칭성, 긴장과 이완의 신체 리듬과 같은 ‘결’의 차원이 부재한다. 이는 바르트가 말한 “존재의 혼적으로서의 목소리”가 탈신체화될 수 있는가라고 묻게 만든다. 반면, 비트박서의 목소리는 극단적으로 신체화된 음향으로, 발화의 의도와 음색의 실현 사이의 마찰이 그대로 드러나는 사례다. 이 경우 ‘결’은 과잉으로 표출되며, 기계음을 재현하려는 시도 속에서도 인간 신체의 특이성이 배제되지 않고 드러난다.

8) Barthes, *Image-Music-Text*, 182-183.

1.2. 발화의 규범성과 사회적 배치: 쿠직의 비판

바르트가 목소리를 신체적 감각의 층위에서 사유했다면, 쿠직은 그 목소리가 누구에 의해, 언제, 어떻게 발화될 수 있는가라는 조건을 사회 문화적 맥락 속에서 비판적으로 고찰하였다.⁹⁾ 쿠직은 “젠더와 성의 음악적 수행에 관하여”(On Musical Performances of Gender and Sex)라는 제목의 글에서 기존의 청취 중심의 음악 비평을 넘어 퍼포먼스 중심의, 신체화된 음악 비평을 제안한다. 그녀에게 음악 수행, 곧 목소리는 단지 소리의 문제나 음악적 형식의 문제가 아니라, 사회적 공간과 권력 관계에 의해 규정되는 수행적 행위이다.¹⁰⁾ 다시 말해, 목소리는 자연적으로 주어진 보편적 권리가 아니라 특정한 사회적·문화적 맥락에서 허용되거나 배제되는 발화라는 것이다.

쿠직은 이 지점을 주디스 버틀러(Judith Butler, 1956-)의 수행성(performativity) 개념과 연결하여 성·젠더가 본질적 속성이 아니라 반복된 수행을 통해 구성되듯, 음악 수행 역시 사회적 규율과 담론에 의한 특정한 방식으로만 가능하다고 설명한다.¹¹⁾ 따라서 어떤 목소리가 아름답다고 판단되거나 적절하다고 여겨지는 기준은 특정한 역사적, 문화적 맥락 속에서 구성된 규범의 산물이라는 것이다.

이 통찰은 우리가 AI 보컬이나 비트박스를 듣는 경험에도 유효하다. AI가 생성한 목소리가 ‘자연스럽다’거나 ‘사람 같다’고 평가될 때, 그 기준은 이미 사회적으로 구성된 청각적 규범을 반영한다. 반대로, 비트박서의 발성이 ‘음악적이지 않다’라거나 ‘기괴하다’라는 평가 역시 전통적 음악 규범과 발화 규칙을 기준으로 한 판단이다. 중요한 점은 이러한 사회적 규범이 단순히 목소리를 허용하거나 배제하는 기준에 그치지 않고, 포스트휴먼 시대의 새로운 목소리 실천을 이해하는 토대가 된다는 점이다.

왜냐하면, AI 보컬은 발화 주체의 부재 속에서 ‘자연스러움’이라는 규범을 재현하며, 비트박스는 과잉된 신체를 통해 기존 규범을 교란한다. 이처럼 쿠직

9) Suzanne G. Cusick, “On Musical Performances of Gender and Sex,” in *Audible Traces: Gender, Identity, and Music*, eds. Elaine Barkin and Lydia Hamessley (Zürich and Los Angeles: Carciofoli Verlagshaus, 1999), 25-48. 이 글은 1992년에 학술대회 발표문으로 처음 발표되었다.

10) Cusick, “On Musical Performances of Gender and Sex,” 25.

11) Cusick, “On Musical Performances of Gender and Sex,” 27-28.

의 논의는 본 연구가 제시하는 포스트휴먼 시대의 목소리, 즉 몸을 넘나드는 목소리, 경계 너머의 음악을 이해하는 데 적용될 수 있으며, 목소리를 단순히 사회적 규범에 의해 제한되는 수행이 아니라, 포스트휴먼적 조건에서 재배치되고 새롭게 구성되는 수행으로 사유할 수 있는 관점을 제공한다.

1.3. 감각과 규범의 접점으로서의 목소리

바르트와 쿠직의 논의를 통합해보면, 목소리는 단순히 의미를 전달하는 매개가 아니라, 신체의 감각적 흔적이자, 사회적으로 허용된 수행의 현장임을 알 수 있다. 바르트가 주목한 것은 ‘결’이라는 비기호적 감각의 압력이며, 쿠직은 그 결이 어디에서, 누구에 의해 발화될 수 있는지를 규정하는 사회적 구조와 규율의 조건을 문제 삼는다. AI 보컬과 비트박스의 두 사례는 각각 다른 방식으로 이 긴장을 드러낸다. AI 보컬은 결이 제거된, 존재 없는 신체에서 만들어진 발화이며, 이는 감각의 충위를 상실한 채 규범화된 음색을 복제하는 방향으로 작동한다. 반대로 비트박스는 규범을 벗어난, 혹은 초과하는 신체의 소리이며, 기존의 음악 언어가 포착하지 못하는 감각의 과잉을 드러낸다.

바르트의 ‘결’ 개념은 AI 보컬에서 상실된 신체성의 흔적을 사유할 수 있게 하고, 쿠직의 분석은 비트박스 보이스가 제도적 음악 체계 밖에서 어떻게 실현되는지를 비판적으로 조망하게 한다. 이처럼 포스트휴먼 시대의 목소리는 신체와 기술, 감각과 규범, 발화와 청취가 얽힌 복합적인 실천이다.

2. 정동과 잠재적 실재성: 마수미의 ‘가상계’

브라이언 마수미는 저서 『가상계』(*Parables for the Virtual*, 2002)에서 정동(affect)을 단순히 개인의 심리 상태나 감정으로 환원하지 않고, 감정화되기 이전의 강도(intensity), 즉 신체적·비언어적 차원에서 발생하는 미세한 변이의 흐름으로 정의한다.¹²⁾ 정동은 주체의 내면에서 발생하는 정서가 아니라, 세계와 신체 사이에서 돌발적으로 발생하는 비인칭적·비의도적 사건으로, 감정(emotion)과 구분된다. 감정은 이미 언어화·심리화되어 사회적으로 공유

12) Brian Massumi, *Parables for the Virtual: Movement, Affect, Sensation* (Durham: Duke University Press, 2002), 27-28.

가능한 상태이지만, 정동은 그러한 명명 이전에 몸을 스치고 지나가는 순간적 강도의 변이다.¹³⁾ 음악을 듣는 순간 소름이 돋거나 숨이 멎는 듯한 긴장, 설명하기 힘든 경이감은 감정이라기보다는 정동의 작용에 가깝다. 이는 청취자가 ‘슬프다’, ‘감동적이다’라고 말하기 전에 이미 몸에서 발생하는 현상이다.

마수미는 정동을 설명하면서 가상계(virtuality)라는 개념을 제시한다. 가상계는 현실에 드러나지 않았지만 언제든지 현실화될 수 있는 잠재적 실재성(potential reality)의 영역이다.¹⁴⁾ 정동은 바로 이 가상계에서 발생하며, 특정한 사건으로 현실화되기 전의 미묘한 움직임을 포착한다. 예컨대 석악가가 숨을 들이마시는 순간, 청중이 아직 노래가 시작되지 않았음에도 다가올 소리를 예감하며 긴장한다. 이는 발화 이전의 잠재적 정동이 청중의 몸에 작용하는 대표적 사례이다.

그는 『가상계』의 마지막 장 “지나치게 파란”(Too-blue)에서 이러한 논의를 한층 더 밀고 나간다. 그는 프랭크 시나트라(Frank Sinatra, 1915-1998)의 눈이 지나치게 파랗게(too-blue) 기억되는 사례를 통해 물리적 색채의 사실 여부와 무관하게 공연을 경험한 이들의 몸에 남는 감각적 과잉의 흔적에 주목한다. 청중은 시나트라의 목소리와 시선, 몸짓 전체를 ‘지나치게 파란’ 사건으로 경험하며, 바로 이 지점에서 정동의 본질이 드러난다는 것이다. 사건은 늘 예상 불가능한 여분의 강도와 함께 발생하며, 바로 이 초과적 요소가 사건의 특이성을 결정한다. 마수미가 제시하는 “지나치게 파란” 개념은 물리적 파장으로 설명되지 않는 색채적 과잉, 즉 언어적·과학적 규칙으로 포착되지 않는 감각적 특이성에 대한 은유이다.¹⁵⁾

13) 마수미는 이를 설명하기 위해 신경학적 연구를 인용한다. 즉, 어떤 자극이 주어졌을 때 신체는 약 0.2초 만에 반응을 시작하지만, 우리가 그것을 ‘기쁨’, ‘슬픔’과 같은 감정으로 인식하고 언어화하는 데는 0.5초가 걸린다. 이 미세한 시간차에서 드러나듯, 정동은 의식적 감정이 발생하기 전에 이미 신체와 세계 사이에서 흐르고 있는 강도의 사건이다. Massumi, *Parables for the Virtual: Movement, Affect, Sensation*, 28-31.

14) 그는 이를 “너무 빠르게 일어나서 사실상 일어났다고 할 수 없는 사건”의 차원으로 묘사하며, 신체가 곧바로 감각하지만 의식적으로는 나중에야 등록되는 영역으로 이해한다. Massumi, *Parables for the Virtual: Movement, Affect, Sensation*, 30-33, 43.

15) Massumi, *Parables for the Virtual: Movement, Affect, Sensation*, 263-265.

그는 나아가 모든 사건은 ‘필수 조건’(requisites)만으로 환원될 수 없으며, 언제나 우연적 초과(contingent excess), 즉 ‘수반’(accompaniment)을 동반한다고 강조한다.¹⁶⁾ 시나트라의 노래는 정교한 기교나 완벽한 해석만으로 설명되지 않으며, 그의 목소리는 호흡의 길고 짧음, 박자를 미세하게 앞서거나 뒤로 끄는 순간들에서 발생하는 특징을 보여준다. 다시 말해, 시나트라의 노래는 곡의 의미를 잘 전달한다는 차원보다, 호흡과 타이밍의 미묘한 어긋남이 청중의 몸에 강도의 사건을 만들어내는 방식으로 작동한다. 마수미는 이를 통해 음악 수행이 단순히 텍스트와 음정에 충실한 재현이 아니라, 언제나 초과와 변이 속에서 발생하는 정동적 사건임을 드러낸다.¹⁷⁾

이와 같은 정동 개념은 AI 보컬과 비트박스의 발화를 분석할 때도 유용하다. 각각의 사례에서 청자가 경험하는 긴장과 낯센은 단순히 기술적 산출이 아니라, 몸에 작용하는 강도의 사건으로 이해될 수 있기 때문이다. AI 보컬은 데이터와 알고리즘이라는 “필수 조건” 위에서 합성되지만, 청취자가 경험하는 낯센과 긴장은 단순히 합성 과정으로 환원되지 않는다. 오히려 존재하지 않는 신체가 남긴 빈자리가 불안정성과 긴장의 초과를 낳으며, 지나치게 매끈하고 자연스러운 발화는 청취자에게 “너무 인간적인”(too-human) 합성음이라는 역설적 정동을 경험하게 한다. 반대로, 비트박스는 신체라는 필수 조건에 기반하지만, 단일한 발화 기관에서 쏟아져 나오는 다성적 전자 음향은 언제나 초과적이며, “너무 기계적인”(too-mechanical) 특이성을 남긴다. 목소리가 악기와 기계의 경계를 끊임없이 횡단하는 이 경험은 사건으로서의 목소리가 지닌 정동적 과잉을 여실히 드러낸다.

결국, 목소리와 몸은 언제나 현실적 조건을 넘어서는 과잉의 흔적을 남긴다. 포스트휴먼 시대의 목소리는 AI 보컬의 지나치게 인간적인 합성음과 비트박스의 지나치게 기계적인 발화를 통해, 늘 경계를 넘어서는 정동적 사건으로 경험된다. 목소리는 필수 조건과 초과의 얽힘 속에서만 성립하며, 따라서 항상 ‘지나치게 무언가’로 경험되는 특이한 정동으로 나타난다. 이러한 관점에서 목소리는 더 이상 인간 주체의 소유물이 아니라, 정동의 흐름 속에서 잠재성과 현실성이 교차하는 순간마다 새롭게 특이화되는(singularized) 존재론적 현상

16) Massumi, *Parables for the Virtual: Movement, Affect, Sensation*, 260-263.

17) Massumi, *Parables for the Virtual: Movement, Affect, Sensation*, 266-269.

이다. AI 보컬은 신체 부재에서 비롯된 잠재적 긴장을, 비트박스는 신체 과잉에서 발생하는 강도의 폭발을 드러낸다. 두 경우 모두 목소리는 주체의 정체성을 드러내는 안정적 매개가 아니라 청취자의 몸과 감각을 흔드는 사건으로 존재한다.

3. 신체, 기술, 문화, 공간의 얽힘, 그리고 보컬: 파스트와 티아이넨, 월슨

20세기 후반 이후 기술과 신체의 관계를 새롭게 규정하려는 시도는 인문학 전반에서 하나의 흐름을 이루었다. 신유물론(new materialism)과 퍼포먼스 연구의 맥락에서, 신체는 더 이상 고정된 생물학적 실체가 아니라, 기술적 매개와 얽히며 끊임없이 재구성되는 과정적 존재로 이해되었다. 캐런 바라드(Karen Barad)는 이를 ‘에이전셜 리얼리즘’(agential realism)이라 개념화하며, 존재는 독립된 개체들이 사후적으로 상호 작용(interaction)하는 것이 아니라, 관계 속에서 함께 발생하는 ‘인트라액션’(intra-action)의 산물이라고 설명한다.¹⁸⁾ 이는 신체와 기술이 분리된 실체가 아니라 수행적 관계 속에서 서로의 존재 조건을 구성하며, 목소리를 포함한 신체적 현상이 언제나 기술적 매개와 얽혀 있음을 이해하는 토대를 제공한다.

이러한 관점은 목소리를 단순히 인간 내부에서 생성되는 자연적 소산으로 보아온 전통적 성악 이론과는 뚜렷이 구별된다. 파스트와 티아이넨은 목소리를 단순히 인간의 후두, 성대, 호흡 기관에서 발생하는 물리적 산물로 환원하지 않고, 신체, 기술, 문화, 공간이 얽힌 인트라액션¹⁹⁾ 속에서 발생하는 과정

18) 바라드는 양자물리학과 페미니즘 이론을 접목해, 존재와 의미가 ‘인트라액션’을 통해 발생한다는 점을 설명한다. 그녀의 이론은 신체와 기술의 관계를 고정된 실체가 아니라 과정적·관계적 얽힘으로 이해하는 토대를 제공한다. 관련 내용은 제2장 4절 “Agential Realism: How Material-Discursive Practices Matter”에서 자세히 서술된다. Karen Barad, *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning* (Durham: Duke University Press, 2007), 132-185.

19) 여기서 그들은 캐런 바라드의 “인트라액션” 개념을 원용한다. 이는 개체가 먼저 존재한 뒤 상호 작용하는 것이 아니라, 관계 속에서 존재가 함께 발생한다는 점을 강조하는 개념이다. Karen Barad, “Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter,” *Signs: Journal of Women in Culture and Society* 28/3 (2003), 815-817.

적이며(processual) 열린(open-ended) 존재로 규정한다.²⁰⁾ 신유물론적 시각에서 목소리는 발화 순간부터 이미 마이크, 스피커, 녹음 장치, 청취 환경과 결합되며, 따라서 본질적으로 관계적이고 하이브리드적이다. 예컨대 성악 공연에서 청중이 듣는 목소리는 단순히 성대에서 직접 발생한 소리가 아니라, 마이크의 감도, 스피커의 주파수 응답, 공연장의 반향과 같은 기술적 장치들이 개입해 구성된 결과라는 것이다.

이러한 신유물론적 시각은 월슨의 ‘보철/인공기관’(prosthesis) 개념으로 더욱 구체화된다. 전통적으로 보철은 상실된 신체를 대체하는 장치로 이해되었지만, 줄리언 존슨(Julian Johnson)의 “악기는 인간 신체의 보철적 확장”²¹⁾이라는 견해와 캐서린 헤일스(Katherine Hayles)의 “신체 자체가 원래부터 보철적이며, 다른 보철은 이미 진행 중인 과정의 연속”²²⁾이라는 견해를 바탕으로, 월슨은 보철을 신체 외부의 대체물이 아니라, 신체와 장치가 결합하여 새로운 기능과 행위성을 발휘하는 공동의 행위자로 이해한다.²³⁾ 여기서 보철은 결핍을 보완하는 보조물이 아니라, 신체가 기술과 얽히며 스스로를 초과하고 재구성하는 과정을 드러내는 것으로 개념이 확장된다.

보철의 관점에서 살펴보면, 마이크와 스피커는 단순히 보조 장치가 아니라 신체로 낼 수 있는 목소리 이상의 음향을 산출하는 핵심적 행위자이다. 특히, 비트박스에서 성문 파열음(glottal stop)이나 호흡음, 입술 파열음을 마이크로 증폭해 베이스 드럼 소리를 재현하는 것은, 신체와 장치가 결합해 새로운 악기로 재구성되는 대표적 사례다.

월슨은 이어서 “확장된 목소리”(Extended voices) 챗터에서 목소리가 단지 신체 내부에서 울려 나오는 것이 아니라, 몸을 넘어 확장되며 신체 자체를 울리는 악기처럼 작동한다고 설명한다.²⁴⁾ 브라이언 퍼니호(Brian John Peter

20) Heidi Fast and Milla Tiainen, “Voice,” *Matter: Journal of New Materialist Research* 9 (2024), 2.

21) Julian Johnson, *Out of Time: Music and the Making of Modernity* (Oxford: Oxford University Press, 2015), 47-52.

22) N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (Chicago: University of Chicago Press, 1999), 3-5.

23) Samuel Wilson, “The Composition of Posthuman Bodies,” *International Journal of Performance Arts and Digital Media* 13/2 (2017), 138-140.

Ferneyhough, 1943-)의 《시간과 움직임 연구 III》(*Time and Motion Study III*, 1974)를 예시로, 그는 전자 음향을 통해 목소리를 재생산하는 것이 단순한 신체적 확장이 아니라 인간과 기계의 대립과 결합을 동시에 드러내는 장이라고 지적한다. 16성부 합창과 타악기, 전자음을 동반한 이 작품에서 가수는 더 이상 통합된 주체가 아니라 기술적 매개 속에서 분리되고 재조립될 수 있는 신체 부분들의 집합으로 재구성된다. 페리호가 의도적으로 만든 '사운드 카펫'(sound carpet)은 각 가수의 목소리를 하나의 개별적인 소리로 드러내기보다, 모두를 거대한 집단적 포효 속에 흡수시킨다. 이는 프로이트(Sigmund Freud, 1856-1939)와 이후에 라캉(Jacques Lacan, 1901-1981)이 말한 '해양적'(oceanic)이고 '전언어적'(prelinguistic) 상태, 즉 자아와 세계의 경계가 사라지고, 언어화 이전의 감각적 융합이 이루어지는 상태를 연상시킨다. 월슨은 이를, 목소리가 기술적 매개 속에서 해체되고 다시 조립되면서 드러나는 포스트휴먼적 신체의 전형적인 모습으로 해석한다.²⁴⁾ 다시 말해, 목소리는 더 이상 자연스럽고 통일된 '자아의 목소리'가 아니라, 기술과 얽혀 끊임없이 분해되고 재구성되는 하이브리드한 사건으로 존재한다.

이와 같은 월슨의 분석은 AI 보컬 사례와도 직접 연결된다. 알고리즘이 생성하는 목소리는 물리적 성대가 아닌, 데이터셋과 보코더(vocoder), 신경망(neural network), 스트리밍 플랫폼까지 포함하는 새로운 신체를 형성한다. 데이터셋과 알고리즘은 발화 기관으로, 스피커와 스트리밍 서비스는 공명 공간으로 작동한다. 다시 말해, AI 보컬 역시 목소리가 인간 신체를 넘어 기술적 장치망 속에서 해체·재조립되어 드러나는 확장된 목소리의 또 다른 양상이다. 따라서 AI 보컬은 결핍의 보충물이 아니라, 기술적 장치망이 만들어낸 독립적 신체의 발화라 할 수 있다.

이 지점에서 브뤼노 라투르(Bruno Latour)의 행위자-네트워크 이론(Actor-Network Theory/ANT)은 보철 개념을 넘어 신체와 기술의 관계를 이해하는

24) 그는 정신분석학적 맥락에서 유아기의 어머니 목소리는 신체적 접촉이 부재한 상황에서 아이를 감싸고 위로하며, 분리와 합일을 매개하는 확장적 기능을 수행하며, 성인의 삶에서도 목소리는 음악 공연을 통해 연주자와 청중의 몸을 함께 감싸거나, 녹음과 전화 통화를 통해 신체적 확장의 새로운 가능성을 드러낸다고 강조한다. Wilson, "The Composition of Posthuman Bodies," 147-150.

25) Wilson, "The Composition of Posthuman Bodies," 149.

방식을 한층 더 확장한다. 전통적으로 ‘행위자’(actor)는 의식과 의도를 가진 인간으로 한정되었지만, ANT에서 행위자는 인간으로 한정되지 않으며, 어떤 상황에서 변화를 일으킬 수 있는 모든 것으로 규정된다. 다시 말해, 데이터나 알고리즘, 기계, 사운드 장치 같은 비인간 요소들 역시 행위의 주체로 간주될 수 있다. 라투르는 사물이 단순히 수동적으로 사용되는 도구가 아니라, 행위 과정에서 능동적으로 역할을 한다고 본다. 그의 표현에 따르면 사물은 “승인하고(authorize), 허락하며(allow), 여유를 주고(afford), 장려하고(encourage), 허용하고(permit), 제안하고(suggest), 영향을 미치고(influence), 차단하고(block), 가능하게 하고(render possible), 금지한다(forbid).”²⁶⁾ 이 말은 사물이 단순히 인간의 명령을 수행하는 매개체가 아니라 스스로 그 어떤 가능성을 열거나 제한하는 적극적 행위임을 의미한다. 이 관점을 목소리에 적용하면, 마이크, 알고리즘, 스피커, 플랫폼은 목소리 산출의 주변적 보조물이 아니라, 발화를 가능하게 하거나 변형시키는 능동적 참여자이다. 따라서 목소리는 더 이상 인간 주체의 단일한 소유물이 아니라, 인간과 기술, 환경이 얹혀 만들어 내는 행위의 결과로 존재한다. 즉, 목소리는 언제나 여러 주체와 매개가 함께 작동하는 네트워크 속에서 발생하며, 그 관계적 얹힘을 떠나서는 이해될 수 없다.

비트박스는 신체와 기술의 얹힘 개념을 보여주는 장르이다. 비트박스 공연에서 마이크는 단순한 소리 증폭기가 아니라 악기의 일부가 된다. 마이크를 입 가까이 밀착시켜 팝 노이즈를 증폭하거나, 손으로 마이크 헤드를 감싸 저역을 강조하는 등의 방식은, 신체와 장치가 결합하여 새로운 음향적 가능성을 여는 순간이다. 이때, 목소리는 인간 발성 기관과 마이크, PA 시스템이 얹힌 확장된 신체에서만 존재할 수 있는 사운드가 된다. 청취자에게 비트박서의 목소리는 더 이상 자연적 발화가 아닌 드럼, 트럼펫, 전자 사운드를 모사하는 다층적 사건으로 경험된다.

26) Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (Oxford: Oxford University Press, 2005), 72.

4. 실재 없는 발화: 장 보드리야르의 시뮬라크르

장 보드리야르는 『시뮬라크르 에 시뮬라시옹』(*Simulacres et Simulation*, 1981)에서 시뮬라크르(simulacre) 개념을 통해 현대 사회의 기호 체계가 어떻게 실재(reality)와의 관계를 전도시키는지를 논의한다. 그는 이미지가 단순히 실재를 반영하거나 왜곡하는 수준을 넘어, 원본 없는 복제가 독자적으로 작용하는 하이퍼리얼리티(hyperreality) 단계를 제시한다.²⁷⁾ 이 단계에서 기호는 실재의 부재를 감추는 것이 아니라 스스로가 실재처럼 기능한다.

그는 시뮬라크르의 전개를 네 가지 단계로 구분한다. 첫째, 기표가 실재를 반영하는 단계에서는 전통적 상징과 재현의 질서가 유지되어, 기표가 실재를 충실히 반영한다. 둘째, 기표가 실재를 왜곡하는 단계에서는 모방이나 패러디를 통해 실재가 변형되어 나타난다. 셋째, 기표가 실재의 부재를 은폐하는 단계에서는 허구적 이미지가 실재를 대체하며, 기표는 더 이상 참조할 원본을 갖지 않는다. 마지막으로, 기표가 실재와 완전히 분리되어 독자적으로 작동하는 단계에서는 기표가 원본 없이 순환하며, 이 상태가 바로 시뮬라크르의 영역이다.²⁸⁾

AI 보컬은 바로 네 번째 단계 ‘시뮬라크르’에 해당한다. AI 보컬은 특정 가수의 발화를 단순히 모방하지 않고, 실제로는 존재하지 않는 “이상적 목소리”를 만들어낸다. 성대의 떨림, 호흡과 발음의 불완전성, 미세한 긴장과 같은 물리적 과정을 제거한 채, 매끄럽고 통제된 소리가 만들어진다. 청취자는 AI 보컬을 듣고 단순히 “부족한 모방”이라고 느끼지 않으며, 그것이 인공적으로 생산된 것임을 알면서도 때로는 인간보다 더 정밀하고 감정적으로 설득력 있는 발화로 경험하기도 한다. 이는 시뮬라크르가 단순히 결핍된 복제가 아니라, 새로운 차원의 ‘더 실재적인’ 발화로 기능한다는 점을 보여준다.

여기서 중요한 문제는 현전(presence)의 붕괴이다. 전술한 바와 같이 옹의 구술성 이론에서 목소리는 언제나 발화 주체의 현전을 전제한다. 그러나 AI 보컬은 원본 가수를 필요로 하지 않으며, 주체 없는 발화, 곧 현전 없는 현전

27) Jean Baudrillard, *Simulacra and Simulation*, trans. Sheila Faria Glaser (Ann Arbor: University of Michigan Press, 1994), 1-2.

28) Baudrillard, *Simulacra and Simulation*, 6-7.

을 구현한다. 물론 데이터셋은 실제 인간의 목소리에서 추출되지만, 산출된 목소리는 특정 개인의 발화를 재현하지 않고, 존재하지 않는 발화자의 목소리로 청취된다. 청취자는 “이 목소리가 누구의 것인가?”라는 질문에 답할 수 없으면서도 동시에 그 목소리를 실재처럼 경험하게 되는 것이다. 이는 시뮬라크르적 발화가 주체성의 부재 속에서도 현전의 효과를 극대화한다는 역설을 보여준다. 다시 말해, AI 보컬은 보드리야르가 말한 실재보다 더 실재적인 목소리의 사레이자 원본 없는 복제가 독자적으로 작동하는 순간이며, 이때 우리는 하이퍼리얼리티의 장 안에 서 있게 된다.

AI 보컬은 감정 표현의 층위에서도 시뮬라크르적 성격을 드러낸다. 전통적으로 성악가는 호흡, 비브라토, 강세, 다이내믹 등을 통해 감정을 표현했지만, AI의 합성 보컬은 알고리즘적 패턴을 통해 감정의 시뮬레이션을 구현한다. 청취자는 그것이 실제 주체의 감정이 아님을 알면서도, 오히려 더 설득력 있는 감정으로 받아들이는 모순적 경험을 하게 된다. 이는 결핍 없는 과잉, 곧 감정의 하이퍼리얼리티이다.

AI 보컬이 신체 없는 발화로서 시뮬라크르적 성격을 지닌다면, 비트박스는 반대로 신체의 과잉된 발화를 통해 목소리를 악기로 전환한다. 그러나 두 사례는 공통적으로 목소리의 존재 조건이 인간 신체에 고정되어 있지 않음을 드러낸다. AI 보컬은 신체 없는 발화가 가능함을, 비트박스는 신체가 기술처럼 작동할 수 있음을 보여준다. 이 양극단은 서로 반대처럼 보이지만, 보드리야르의 관점에서는 모두 실재와 허구, 인간과 기계의 경계를 교란하는 사건으로 연결된다.

보드리야르의 시뮬라크르 이론을 통해 본다면, AI 보컬은 단순히 기술적 성과물이 아니라 존재론적 전환의 징후다. 목소리는 더 이상 주체의 신체와 일대일로 대응하지 않으며, 원본 없는 발화가 오히려 더 실재적으로 경험되는 하이퍼리얼리티의 장으로 옮겨졌다. 이 과정에서 목소리는 주체성, 실재성, 감정의 전달이라는 전통적 범주를 벗어나, 스스로 작동하는 기호이자 사건으로 변모한다. 보드리야르의 시뮬라크르 이론은 이러한 변화를 설명하는 강력한 개념적 틀이며, 이를 통해 목소리가 어떻게 새로운 존재 조건을 획득하는지를 파악할 수 있다.

5. 아상블라주로서의 목소리: 이론의 통합

아상블라주(assemblage)는 질 들뢰즈(Gilles Deleuze, 1925-1995)와 펠릭스 가타리(Pierre-Félix Guattari, 1930-1992)가 『안티 오이디푸스』(*L'Anti-Edipe*, 1972)와 『천 개의 고원』(*Mille Plateaux*, 1980)에서 제시한 개념으로, 사물, 신체, 기호, 정동, 기술 등이 특정 맥락에서 임시적으로 결합하여 형성되는 배치를 의미한다. 이들에게 아상블라주는 고정된 실체라기보다 언제나 과정적이고 관계적인 사건이며, 끊임없이 형성되고 해체되는 운동을 통해 드러난다. 특히 그들은 아상블라주를 내용(content)과 표현(expression)의 축, 그리고 영토화(territorialization)와 탈영토화(deterritorialization), 재영토화(reterritorialization)의 운동으로 설명하였다.²⁹⁾

마누엘 데란다(Manuel DeLanda)는 들뢰즈와 가타리의 아상블라주 개념을 계승하면서, 이를 사회적·문화적 분석틀로 확장하였다. 그는 아상블라주를 “외적 관계 속에서 결합하는 부분들의 전체”³⁰⁾로 정의하였다. 이 정의는 부분이 전체에 내재적으로 종속되지 않고, 다른 배치로 이동하거나 재조합될 수 있다는 점을 강조한다. 데란다는 들뢰즈와 가타리가 이미 제시했던 영토화-탈영토화-재영토화의 축을 그대로 계승하면서, 여기에 코딩(coding)과 디코딩(decoding)이라는 세 번째 축을 추가하였다. 첫 번째 축인 ‘내용-표현’ 축은 물질적 요소와 기호적 요소의 상호 작용을 가리키며, 두 번째 축인 ‘영토화-탈영토화-재영토화’ 개념은 배치가 안정화, 이탈, 재구성되는 운동을 포착한다. 마지막으로 ‘코딩-디코딩’ 축은 언어, 유전자와 같은 특수한 표현 매체가 아상블라주의 정체성을 공고히 하거나 유연하게 만드는 과정을 설명한다.³¹⁾

목소리를 단일한 기관의 산물이 아니라, 신체, 기술, 공간, 청취자가 얹혀 발생하는 사건으로 분석하기 위해서는 데란다가 제시한 세 축이 특히 유용하다. 내용-표현의 측면에서 볼 때, AI 보컬은 데이터셋, 알고리즘, 보코더, 서

29) Gilles Deleuze and Félix Guattari, *A Thousand Plateaus: Capitalism and Schizophrenia*, trans. Brian Massumi (Minneapolis: University of Minnesota Press, 1987), 88-89.

30) Manuel DeLanda, *A New Philosophy of Society: Assemblage Theory and Social Complexity* (London: Continuum, 2006), 24.

31) DeLanda, *A New Philosophy of Society*, 9, 21, 24.

버, 스피커 같은 물질적 장치들의 작동을 통해 성립하고(내용), “인간처럼 들려야 한다”는 규범, 원본 없는 발화를 실제로 경험하는 청취자의 인식, 하이퍼리얼리티적 감각 속에서 의미를 획득한다(표현). 비트박스 또한 후두, 혀, 입술 등 발화 기관과 마이크, PA 시스템, 공연장의 음향 조건이라는 물질적 기반(내용) 위에서, 공연 맥락, 관객 반응, 목소리를 악기로 인식하게 하는 문화적 코드(표현)가 결합한다. 이 두 사례는 목소리를 신체 내부의 자연적 발화로 한정할 수 없으며, 물질과 기호가 얹힌 배치 속에서만 이해할 수 있음을 보여준다.

두 번째 축을 대입해보면, AI 보컬은 목소리를 신체와 동일시해온 전통적 이해를 탈영토화한다. 목소리가 신체 없이도 성립할 수 있다는 사실은, 발화를 인간 주체의 현전과 결부시켜온 음악학적 전제를 해체한다. 그러나 동시에 AI 보컬은 “자연스러운 발음”, “정서적 프레이징”과 같은 기술적 표준에 의해 재영토화된다. 비트박스는 발화 기관이 언어적, 성악적 기능을 넘어 다성적, 타악적 소리를 산출하는 탈영토화의 실천이다. 그러나 공연 관습, 마이크 테크닉, 장르적 규범 속에서 다시 영토화된다. 이처럼 목소리는 언제나 탈영토화와 재영토화의 긴장 속에서 잠정적 정체성을 획득한다.

마지막 코딩-디코딩 축을 적용해보면, AI 보컬은 음향 데이터를 정규화하고 음색, 발화 양식을 규격화하는 알고리즘적 코딩의 산물로 볼 수 있다. 동시에 새로운 맥락 속에서 활용될 때는 기존 규범을 벗어나 디코딩의 과정을 겪게 된다. 비트박스 역시 장르적 코드에 의해 일정 부분 규격화되지만, 개인 퍼포머의 스타일과 창의적 변주는 그 코드를 디코딩하면서 새로운 발화 방식을 열어감을 알 수 있다.

이러한 관점은 음악학의 전통적 분류 체계에 변화를 요구한다. 첫째, 작품과 연주자의 구분은 상대화된다. AI 보컬은 주체 없는 발화라는 새로운 층위를 형성하여 작품과 연주자라는 이분법을 흔든다. 둘째, 성악과 기악의 구분은 해체된다. 목소리가 기술과 얹힌 음향적 배치의 산물일 때, 성악은 더 이상 인간 신체의 특권이 아니라 악기적 사운드와 연속체를 이룬다. 셋째, 실연과 가상의 구분도 재구성된다. 비트박스 공연과 AI 보컬은 만들어지는 경로와 작동 방식에서 서로 다르지만, 두 경우 모두 목소리라는 가상블라주를 구성하며, 듣는 사람에게는 고유한 현실감을 주고 실제로 존재하는 소리처럼 경

힘된다.

따라서 목소리는 단일한 실체가 아니라, 물질과 기호, 규범과 기술이 얹히는 과정 속에서 형성되는 아상블라주적 현상으로 볼 수 있다. AI 보컬과 비트 박스는 이 가능성을 양극단에서 보여주는 사례로, 목소리를 성악/기악, 인간/기계, 실연/가상이라는 전통적 범주를 넘어 사유할 수 있는 길을 제시한다. 이는 디지털 시대, 포스트휴먼 시대의 음악 존재론을 새롭게 구성하는 데 핵심적인 이론적 틀을 제공한다.

III. 사례 분석: AI 보컬과 비트박스

1. 수노 AI의 노래 《몸은 없는데 숨을 쉬네》

수노 AI는 노래를 자동으로 생성할 수 있는 대표적인 인공지능 음악 생성 플랫폼으로, 가사 작성, 멜로디 생성, 편곡 및 보컬 합성을 통합적으로 처리한다. 사용자는 단순히 가사 또는 짧은 지시문이나 키워드를 입력하는 것만으로 완결된 노래를 얻을 수 있으며, 장르, 스타일, 분위기 등을 선택할 수 있다.³²⁾

본 연구자는 자동 가사 생성 기능을 사용하지 않고, 연구 주제를 응축한 내용의 5연 4행 구조의 가사와 제목 ‘몸은 없는데 숨을 쉬네’라는 제목을 직접 입력하였다. 희망 장르는 ‘Hip-Hop’, 스타일은 ‘Lyric’을 선택하였는데, 이는 리듬과 발화 억양에 가까운 운율적 특성을 살리기 위함이었다. 수노 AI가 산출한 결과물은 1분 38초 길이의 남성 솔로곡으로,³³⁾ 노래는 전주-1연-2연-

32) 본 연구에서 사용한 버전은 수노 AI 3.5 무료판으로, 이는 2024년 공개된 버전3과 2024년 말의 버전4 이후 현재까지 유지되는 체계이다. 버전4가 “라디오 품질에 준하는 완곡을 수초 내 생성”과 “더 선명한 가사, 더 역동적인 곡 구조”를 강조했다면, 2025년 현재의 무료 이용 구간(버전3.5)은 제한된 크레딧 내에서 텍스트 기반 곡 생성을 지원한다. 이러한 구조적 특징은 사용자가 후렴, 형식, 보컬 톤을 세세히 지정하지 않아도 모델이 반복 구절과 hooks를 자동으로 조직하는 경향과 맞닿아 있다. 다만 수노는 핵심 아키텍처와 훈련 데이터셋을 공개하지 않으므로, 본 연구에서 기술적 언급은 공개 자료 범위 안으로 한정한다.

33) 해당 음원은 다음 링크를 접속하면 감상할 수 있으며, 논문 말미에 QR코드로도 접속할 수 있다. <https://m.blog.naver.com/PostView.naver?blogId=tnmusiclab&>

간주-3연-4연-5연-5연-후주의 구조로 이루어졌다. 가사 반복을 지정하지 않았음에도 AI는 ‘우린 정말 존재할까 / 아니면 코드의 환상(거짓말)일까 / 목소리와 몸을 나눌 수 있나 / 데이터 숲에 답이 있을까’의 네 가지 질문으로 구성된 5연(2연)의 가사를 후렴으로 배치하여 노래의 핵심 메시지를 청취자가 기억할 수 있도록 하였다. 박자는 4비트, 템포는 여유 있으며, 한 행이 하나의 악구로 구성되었고, 선율의 억양과 강세는 실제 구어체와 매우 흡사하여 가사의 의미가 자연스럽게 전달되었다. 흥미로운 점은 지시하지 않았음에도 전주, 간주, 후주에서 성악 선율과는 독립된 플루트 솔로 선율이 리프레이너처럼 반복되어 곡 전체에 통일감을 부여한다는 점이다. 특히 후주에서는 플루트 선율과 함께 AI 보컬의 보컬리제가 삽입되는데, 여기에 기계음을 연상시키는 이펙터가 결합되어 곡 전체의 주제인 ‘가상적 소리’(virtual sound)를 환기하며, 인간과 비인간의 경계를 소리로 표현한다. 발음은 전반적으로 한국어와 영어 모두 원어인 수준의 자연스러움을 보였으나, 4연 2행의 ‘프로그램은 꿈결일까’에서 ‘프로그램은’ 부분의 음절 구분이 다소 어색하게 들렸다.

수노 AI가 산출한 결과물 《몸은 없는데 숨을 쉬네》는 단순히 기계가 생성한 “노래 같은 소리”가 아니라, 인간적 발화와 음악적 구조를 모방하면서도 동시에 새로운 표현 방식을 창출한 산물이라 할 수 있다. 이러한 현상을 이해하기 위해서는 목소리가 어떠한 과정을 거쳐 생성되는가를 면밀히 살펴볼 필요가 있다. 물론 수노 AI의 내부 구조는 공식적으로 공개되지 않았기에 구체적인 알고리즘을 직접 확인할 수는 없다. 그러나 지금까지 보고된 일반적 싱잉 보이스 신테시스(Singing Voice Synthesis, ‘SVS’)의 파이프라인을 통해 추론적으로 설명하는 작업은 여전히 주목할 만하다. 이는 단순히 기술적 과정에 대한 가설을 제시하는 데 그치지 않고, 왜 청취자가 그것을 “실재하는 목소리”로 경험하는지, 또 신체가 부재함에도 불구하고 인간적 호흡, 프레이징, 억양을 감각할 수 있는지를 이해하는 핵심 단서를 제공하기 때문이다. 다시 말해, 수노 AI의 목소리 합성 과정을 추정해보는 일 자체가 목소리의 존재론적 의미를 밝히는 데 중요한 연구 가치가 있다.

수노 AI의 목소리 생성 과정 내부 구조는 공식적으로 공개되지 않았으나, SVS의 일반적 구조인 텍스트-투-싱잉(Text-to-Singing, 'TtSing') 파이프라인을 따랐을 가능성이 크다. TtSing은 텍스트, 멜로디 계획과 보컬 합성, 후처리의 단계를 거친다. 먼저, 입력된 가사는 음소 단위로 변환되어 발음 패턴이 설정된다. 이어서 장르와 스타일, 무드 정보가 텍스트 임베딩과 함께 처리되며, 이를 기반으로 피치 곡선(f0), 길이(duration), 악센트(accent) 등 노래 특유의 운율과 멜로디 구조가 생성된다. 이때 후렴구나 반복 구절은 사용자가 사전에 지정하지 않더라도, 모델이 곡의 서사와 음악적 흐름을 분석하여 자동으로 반복 패턴을 설계한다. 이어서 가사와 멜로디 계획을 바탕으로 멜 스펙트로그램이 생성되고, HiFi-GAN과 같은 고해상도 신경 보코더를 거쳐 실제 파형(waveform)으로 변환된다. 이 과정에서 비브라토, 호흡, 발성 강약 등 세밀한 표현 요소가 반영되어, 인간 보컬리스트의 발화를 연상시키는 자연스러운 목소리가 구현된다. 여기에 학습 데이터에서 추출한 발성, 억양, 호흡 패턴이 장르와 분위기에 맞춰 스타일 임베딩으로 적용된다. 이 덕분에 폐, 성대, 구강 등 실제 신체 기관이 전혀 없는 상태에서도 인간 보컬과 유사하거나 때로는 더욱 정교한 발화 표현이 가능하다. 마지막으로 반주 트랙과 보컬 트랙이 믹싱, 마스터링 과정을 거쳐 리버브, 공간감, 압축, EQ 등의 효과가 더해지면서 청취자가 이를 '실재하는 목소리'로 인식하도록 완성된다.

이 과정에서 중요한 점은, 신체적 기관이 없음에도 불구하고 수노 보컬은 마치 실제 폐와 성대가 작동하는 것처럼 호흡, 강세, 프레이징을 재현한다는 점이다. 이는 단순한 “음성 모사”가 아니라, 데이터와 알고리즘이 만들어내는 새로운 목소리의 존재 방식이라 할 수 있다. 실제 결과물에서 호흡 소리는 구절의 흐름과 음악적 호흡에 맞추어 삽입되었으며, 프레이징과 억양은 강세, 속도, 음량의 변화를 섬세하게 조정하여 인간 가창의 표현적 뉘앙스를 구현한다. 발음은 실제 가수와 거의 구분되지 않을 정도로 자연스러우며, 음소 간 전환도 자연스럽게 이어진다. 이러한 청취 경험은 발화 주체가 부재함에도 불구하고 청취자가 “누군가 실제로 노래하고 있다”라고 느끼게 만든다. 특히 스스로 후렴구를 지정하고 이를 반복하거나 기악적 리프레이ンを 삽입한 것은 AI가 단순히 가사를 읽는 것이 아니라 곡 전체의 의미를 이해하고 그것을 스스로 구조화하고 음악으로 설계할 수 있음을 보여준다. 또한, 후주의 기계적 보컬리제

는 인간과 비인간, 현실과 가상이라는 대립 관계를 실제 사운드를 통해 구현함으로써 AI 보컬이 단순한 모사를 넘어 독자적 미학을 구축하고 있음을 보여준다.

이상의 분석을 통해 수노 AI 사례는 본 연구의 서두에서 제기한 질문들에 다음과 같은 답을 제시한다. 첫째, “디지털 시대의 보컬 사운드는 성악과 기악의 경계를 어떻게 전복하는가?” 수노 AI의 보컬은 전통적으로 성악의 고유한 기능으로 여겨져 온 텍스트 전달을 수행하면서 동시에 기악적 기능(플루트 솔로와 결합된 후주에서 나타나는 보칼리제)을 병행한다. 이는 목소리를 단순히 성악과 기악이라는 이분법적 구분 안에서 설명하기 어려운 새로운 하이브리드적 위치에 놓이게 한다. 이는 강서희의 2014년 선행 연구가 지적한 목소리의 다중적 역할을 계승하면서, 그 대상을 인간 가수가 아닌 AI 보컬로 확장한다는 점에서 학문적 의의를 가진다.

둘째, “실재하는 신체 없이 생성된 AI 보컬은 음악적으로 어떤 존재론적 정체성을 가지는가?” 신체의 부재에도 청취자는 호흡, 질감, 억양과 같은 신체적 흔적을 경험하며, AI 보컬을 ‘실체 없는 목소리’이자 동시에, ‘실재하는 사건’으로 체험하게 된다. 이는 곧 AI 보컬이 비가시적 신체성과 사건적 실재성을 동시에 내포하는 이중적 정체성을 지닌다는 사실을 보여준다.

셋째, “목소리라는 개념은 어떻게 재정의될 수 있는가?” 수노 AI의 사례는 목소리가 더 이상 신체와 동일시되지 않음을 보여주며, 인간과 기계, 성악과 기악, 실제와 가상의 경계를 가로지르는 관계적 사건(relational event)으로 새롭게 사유될 필요가 있음을 드러낸다.

결국 수노 AI보컬 사례는 감정, 호흡, 발음, 곡 구조까지 포함한 총체적 창작 수행을 기계가 독자적으로 구현할 수 있음을 입증한다. 이는 발화 주체의 실재성을 전제로 하지 않는 새로운 형태의 목소리이며, 포스트휴먼 시대 목소리 존재론의 전환을 실질적으로 경험하게 하는 장면이다. 더 나아가, 가상 환경 속에서 기술이 스스로 곡 구조와 표현을 설계하는 방식은 향후 음악 창작에서 ‘주체’ 개념이 새롭게 정의될 필요가 있음을 분명히 보여준다.

2. 비트박서 윈(Wing)의 《도파민》(Dopamine)

2025년 2월 18일 공식 발표된 비트박서 퍼포먼스 《도파민》은 전형적인 EDM 사운드를 오직 목소리와 신체, 마이크 등의 보조 장치만으로 구현한 작품이다. 곡의 제목 ‘도파민’이 암시하듯, 실제 연주는 긴장과 해방, 급격한 강도의 상승을 반복적으로 교차시키며 청취자의 생리적 반응을 강하게 자극한다. 원래 전자 악기와 이펙터, 사이드체인 컴프레션이 담당하던 역할이 여기서는 성문, 혀, 입술, 구강의 섬세한 움직임과 마이크, 손, PA라는 보철적 장치의 결합을 통해 재현된다. 이는 신체와 기계가 얹히며 새로운 음악적 장을 형성하는 대표적 사례이다.

《도파민》은 일반적인 댄스 음악 구조³⁴⁾를 압축한 인트로(intro), 빌드업(build-up), 드롭(drop), 재빌드/아웃트로(rebuild/ outro)의 네 부분으로 구성된다. 브레이크(break)나 세컨드 드롭(second drop) 없이 하나의 드롭이 길게 확장되는 것이 특징이다.³⁵⁾ 각 부분에서 비트박서의 목소리는 다성적 직조를 통해 여러 층위의 음색을 동시에 산출한다.

인트로는 트럼본 사운드와 내레이션으로 시작한다. 음고를 점차적으로 끌어올리듯 탐색하다가 순차상행하는 선율을 반복한 후 내레이션 “The brain’s motivation neurotransmitter, dopamine”이 낭송되는데, 이는 곡의 개념적 배경과 청각적 몰입을 동시에 마련하는 장치이다. 빌드업에서는 킥과 스네어의 규칙적 반복, 하이햇과 크래시의 점층적 배치, 서브 스위프(sub sweep)³⁶⁾

34) 일반적으로 댄스 음악은 빌드업(build-up), 드롭(drop), 브레이크(break), 재빌드(rebuild), 세컨드 드롭(second drop)의 구조이다. 빌드업은 킥 드럼, 스네어 롤, 신스 패턴, 이펙트 등을 활용해 점차 긴장감을 고조시키며 드롭을 예고하는 구간이며, 드롭은 빌드업으로 축적된 긴장이 폭발하며 강렬한 베이스와 리듬이 전면화되는 순간으로, EDM의 클라이맥스를 이룬다. 브레이크는 드롭 이후 에너지를 비우고 보컬이나 멜로디가 부각되며 청자가 숨을 고를 수 있도록 하는 전환부이다. 재빌드는 브레이크 이후 다시 긴장을 쌓아 올려 다음 클라이맥스를 준비하는 구간이며, 세컨드 드롭은 변형되거나 반복된 형태로 두 번째 클라이맥스를 형성하여 곡의 후반부 에너지를 이끌어낸다.

35) 공식 뮤직비디오 영상을 기준으로 각 부분의 재생 시간은 인트로(0:00-0:36), 빌드업(0:36-0:54), 재빌드/아웃트로(2:25-2:48)이다. 영상은 다음 링크를 접속하면 감상할 수 있으며, 논문 말미에 첨부된 QR로도 접속할 수 있다. <https://www.youtube.com/watch?v=qlrpeYdm9Ec> [2025년 8월 29일 접속].

36) 서브 스위프는 저주파에서 고주파로, 혹은 그 반대로 음높이와 음압을 점차적으로

과 필터 스네어, 스네어 롤(snare roll)³⁷⁾이 더해져 EDM 특유의 긴장과 상승감을 구현한다. 드롭 부분에서는 드럼, 베이스, 신스 레이어가 전면적으로 가동되며 음향의 밀도를 극대화한다. 빌드업이 주로 리듬 위주의 골격을 세우는 구간이었다면, 드롭은 모든 악기 레이어가 동시에 작동하여 에너지를 최고조로 끌어올린다. 특히 서브 스윙과 업리프터(uplifter)³⁸⁾ 같은 사운드 이펙터가 드롭의 시작과 동시에 결합되어 음향적 공간을 확장하고 압도적인 몰입감을 형성한다.³⁹⁾ 베이스와 신스 계열 사운드는 비트박서의 성대 진동과 구강 공명에 의해 발생한 것으로, 이는 기계적 합성음과 유사한 음향 질감을 만들어내면서 동시에 인간 신체성의 흔적을 드러낸다.

드롭 구간의 충만한 음향은 전자 사운드의 단순 모사가 아니라, 발성 기관의 세밀한 움직임들을 통해 음역대별로 분화된 소리를 만들어내는 방식으로 실현된다. 따라서 비트박서가 어떤 발성 기관과 조음 동작으로 소리를 산출하는지 살펴보는 일은, 신체와 기술의 얽힘이라는 새로운 음악적 층위를 이해하는 것이다. 이를 위해 두번째 사례 연구에서는 비트박서가 소리를 생성하는 원리를 살펴보고 목소리가 어떻게 악기화되는지, 또한 신체와 기계의 관계가 어떻게 변주되는지를 탐색한다.

먼저 저음역에서는 킥과 베이스 사운드가 중심을 잡는다. 이 소리는 양순 파열음 [b]를 변형하거나, 입술을 느슨하게 진동시켜 구강과 후두강을 공명시키는 립롤 베이스(lip roll bass)가 사용된다.⁴⁰⁾ 여기에 성문을 짧게 닫아 압축했다가 열어주는 동작을 더하면 전자 베이스의 어택과 릴리스를 표현할 수

변화시켜 상승감(rising tension)이나 하강감(releasing tension)을 유도하는 전자 음향 효과를 뜻한다. EDM의 빌드업이나 드롭 직전 구간에서 긴장과 기대를 고조시키는 대표적인 장치이다.

- 37) 스네어 드럼의 음을 짧은 간격으로 연속적으로 배치하여 점차 속도와 강도를 높여가는 리듬 패턴을 말한다.
- 38) 업리프터는 주파수, 음량, 필터 컷오프 등을 점차적으로 상승시켜 청취자에게 긴장감과 고조감을 부여하는 전자 음향 효과를 말한다.
- 39) 본 연구의 빌드업 및 드롭 구간의 음향 분석은 윈의 《도파민》을 MIDI로 재현한 커버 영상의 시각적 파형을 참고하였다. 해당 영상은 다음 링크 접속하면 확인할 수 있으며, 논문 말미에 첨부된 QR로도 접속할 수 있다. <https://www.youtube.com/watch?app=desktop&v=H6C7GXxLP1g> [2025년 8월 27일 접속].
- 40) 립롤베이스 튜토리얼은 SkilleR의 유튜브 영상을 참고하였다. <https://www.youtube.com/watch?v=ilGhsSY4crw> [2025년 8월 27일 접속].

있다.⁴¹⁾ 중음역에서는 스네어와 클랩의 변주가 두드러진다. 대표적으로 PF-스네어(PF-snare),⁴²⁾ 인워드 스네어(inward snare),⁴³⁾ K-스네어(K-snare)⁴⁴⁾ 등이 사용되며, 연주자는 발성의 방향(호기/흡기)과 조음 위치(양순/치조/연구개)를 교차적으로 활용하여 다채로운 음색과 질감을 만들어낸다. 치조·입천장 마찰음을 겹치면 클랩과 유사한 잡음 계열의 효과를 낼 수 있다. 고음역에서는 [ts], [s], [ʃ] 같은 마찰음을 흔들어 하이햇의 셔플감을 구현한다. 이 소리들은 질량감은 적지만, 어택의 미묘한 어긋남이 리듬의 그루브를 만들어낸다. 리드나 브라스 계열의 음향은 양순 버즈(bilabial buzz)로 구현된다. 입술의 압력과 구강의 형태, 그리고 볼의 미세한 팽창을 조절해 포먼트의 피크를 이동시키면 실제 트럼펫처럼 피치를 꺾거나 휘어 올릴 수도 있다. 효과음의 경우, 혀의 측면 클릭과 흡착 해제를 통해 물방울 소리를 내거나, 손바닥을 마이크에 덧대어 스크래치와 유사한 마찰음을 형성하며, 성문 파열음과 공기층의 급격한 방출로 이른바 건조한 크랙음과 혀를 이용해 만드는 클릭롤 사운드를 만들어 곡 전반의 액센트와 질감을 풍부하게 한다.⁴⁵⁾

이 다양한 발성 어휘의 공통점은 촉각성에 있다. 소리의 출발점이 성문, 점막, 치조, 입술의 마찰과 접촉에서 비롯되기 때문에, 청취 경험은 점차 듣기(listening)에서 감각하기(sensing)로 옮겨간다. 이는 바르트가 말한 ‘목소리의 결’이 과잉으로 드러나는 순간이며, 듣는 이는 마치 소리를 귀로 듣는 것만이 아니라 몸으로 만지는 듯한 감각을 경험하게 된다.

공간적 차원에서 마이크는 단순한 증폭기가 아니라 악기의 본체와 같다. 다이나믹 마이크의 근접 효과는 저음을 과도하게 부각시켜 베이스의 체감을 강

41) 이 연구에서는 비트박스 연주 기법을 설명하기 위해 seboom 홈페이지에 게시된 튜토리얼 영상 자료를 참고하였다. <https://www.youtube.com/watch?v=XBSPHZp0s60> [2025년 8월 27일 접속].

42) 자음 P와 F의 조합으로, 입술 파열음과 치조 마찰음을 동시에 사용하여 실제 스네어 드럼과 유사한 타격감을 구현한다.

43) 숨을 들이마시면서 내는 스네어 소리로, 강한 압력과 날카로운 효과를 낼 수 있으며 호흡을 아끼는 데에도 유리하다.

44) 혀 뒷부분을 입천장에 붙였다가 강하게 떼며 내는 스네어 소리로, 단단하고 메탈릭한 타격감을 재현한다.

45) 비트박서 윈이 MBC ‘전참시’에 출연하여 직접 소리 내는 방법을 설명하는 영상은 다음 링크에서 확인할 수 있다. <https://youtu.be/EUuSJOkQipw?si=wTMhrSk6iThsFwge&t=102> [2025년 8월 27일 접속].

화한다. 연주자는 손으로 마이크를 감싸는 컵핑(cupping)을 통해 저음을 늘리거나, 손가락의 틈을 조정해 고역을 살리며 아날로그 필터링에 가까운 효과를 구현한다. 이 과정에서 음향 엔지니어는 사실상 연주자의 외부 공명 기관으로 기능하며, 확장된 신체의 일부가 된다.

《도파민》의 핵심은 자기-보철화된 신체다. 성문, 혀, 입술, 구강의 세밀한 운동이 마이크, 손, PA와 얹히며 전자악기의 가상 밴드를 대체할 뿐 아니라, 기계가 낼 수 없는 차별적 가능성을 만들어낸다. 예를 들어, 숨을 들이마시며 연주하는 스네어 사운드는 전자 샘플로는 구현하기 어려운 시간적 결을 만들어내며, 그루브의 미시적 비대칭을 끊임없이 산출한다. 신체와 장치는 단순한 모방 관계가 아니라 새로운 차이를 낳는 창조적 장치인 것이다.

정동 이론의 관점에서 볼 때, 《도파민》은 감정으로 표출되기 이전 단계에서 강도를 점화한다. 빌드 구간에서는 미세한 진동이 점차 누적되고, 드롭에서는 순간적 고조가 폭발하며, 브레이크에서는 여백이 마련된다. 이 과정은 청자의 호흡, 심박, 근육 긴장과 같은 신체 리듬을 음악적 강도의 변화와 동조시키며, 고조-해소-재고조의 정동적 회로를 완성한다. 특히 곡의 초반에 삽입된 내레이션은 단순히 신경과학적 사실을 전달하려는 설명이라기보다는 오히려 곡 전체의 구조가 어떻게 작동하는지를 상징적으로 지시한다. 즉, 도파민이라는 신경 전달 물질이 쾌락과 보상의 회로를 형성하듯, 이 퍼포먼스 역시 긴장과 해소, 다시금 상승하는 고조의 리듬을 통해 청취자의 몸에 직접 작용한다.

이 퍼포먼스는 또 다른 층위에서 ‘시뮬라크르’의 전복을 보여준다. AI 보컬이 신체 없는 발화로 원본 없는 복제를 구현한다면, 《도파민》은 신체가 기계의 소리를 역으로 모사한다. 이는 단순한 흉내가 아니라, 기계적 문법의 내재화이자 “기계의 생체화”라 부를 만한 사건이다. 청취의 차원에서도 《도파민》은 새로운 규범을 제시한다. 전통적 성악이 깨끗한 발음과 균등한 음질을 선호했다면, 비트박스는 노이즈와 마찰, 파열로 생성된 사운드를 주요 자원으로 사용함으로써 ‘거칠의 세련화’라는 역설을 드러낸다.

비트박스 wings의 퍼포먼스는 서두에서 제기한 첫 번째 질문, “디지털 시대의 보컬 사운드는 전통적인 음악 분류 체계(성악 대 기악)를 어떻게 전복하는가?”에 대해 명확한 답을 제공한다. 비트박스는 인간 신체에서 발생한 소리이지만, 전통적 의미의 ‘노래’도 ‘악기 연주’도 아닌 제3의 실천에 속한다. 이 퍼포먼

스에서 목소리는 더 이상 성악의 범주로만 환원되지 않으며, 동시에 기악의 범주에도 온전히 귀속되지 않는다. 목소리는 성악과 기악의 경계를 해체하며, 신체와 기술이 얹혀 생성되는 다성적 사건으로 드러난다.

세 번째 질문, “인간의 발성 기관으로 디지털 악기 사운드를 구현하는 비트 박스는 기존 음악 장르 체계에 어떤 균열을 일으키는가?”에 대해서도 위의 사례는 중요한 단서를 제공한다. 이 퍼포먼스에서 인간의 입, 성문, 혀, 구강은 드럼 머신이나 신시사이저를 대체하는 하나의 ‘확장된 악기’로 작동한다. 이는 장르적 구분을 전제하는 음악 체계 내부에 균열을 일으킨다. 비트박스는 성악이자 기악, 동시에 퍼커션이자 신시사이저적 수행으로, 단일 장르나 연주 방식으로는 정의될 수 없는 새로운 혼종적 실천을 형성한다.

마지막으로 네 번째 질문, “이러한 사례들은 ‘목소리’라는 개념 자체를 어떻게 재정의하게 하는가?”라는 물음에 대해, 《도파민》은 목소리를 단일 주체의 소유물이나 단순한 의미 전달의 매개체가 아닌, 신체와 기술, 규범과 공간이 결합하여 발생하는 이상블라주적 사건임을 분명히 보여준다. 과잉된 신체의 마찰과 진동은 전자악기의 음향을 재현하면서도, 기계가 낼 수 없는 미시적 어긋남과 특이성을 생산하며, 이는 미디(MIDI)로 구현한 커버 퍼포먼스에서도 확인된다. 이처럼 위의 사례에서 목소리는 ‘인간의 발화’라는 본질적 정의를 벗어나, 신체와 기계, 감각과 규범이 얹히는 현장에서 생성되는 존재론적 사건으로 드러나며, 인간/기계, 성악/기악 그 어느 영역에도 귀속되지 않고 새로운 독자적 층위를 구축한다.

IV. 나가며

포스트휴먼적 조건에서 목소리는 특정한 주체의 소유물로 환원되지 않는다. 오히려 기술과 신체, 환경이 교차하는 자리에서 매번 다른 방식으로 출현하며, 그때마다 새로운 의미와 감각을 만들어낸다. AI 보컬은 신체가 전혀 없는 상태에서도 습결과 억양, 심지어 감정의 흔적까지 재현하고, 비트박스는 과잉된 신체를 통해 기계적 음향을 육화한다. 두 사례에서 보았듯이, 오늘날 목소리를 이해한다는 것은, 그것을 하나의 본질로 고정하지 않고, 신체와 기술, 공간과

규범의 경계를 가로지르며 발생하는 관계적 사건으로 사유하는 일이다.

이러한 변화는 청취 방식과 연주 방식 모두에 중요한 함의를 가진다. 청취자의 입장에서 포스트모던 시대의 목소리는 더 이상 “기계인가 인간인가?”, “실재인가, 모조품인가?”의 이분법적 기준으로 평가되지 않는다. 중요한 것은 그것이 남기는 감각적 흔적과 정동적 울림이다. AI 보컬의 매끈한 발화가 주는 낯센과 불안, 비트박스의 과잉된 신체가 주는 진동은, 그 출처와 방식이 달라도 모두 청취자의 몸을 흔드는 사건으로 경험된다. 청취자는 목소리를 의미나 정체성의 기호로만 듣는 대신, 몸으로 감각하는 사건으로 받아들여야 한다.

반대로 연주자의 입장에서 포스트휴먼 시대의 목소리는 새로운 예술적 과제를 제시한다. 성악가는 더 이상 목소리를 “나의 신체” 안에만 한정하지 않고, 무대, 마이크, 공간, 청취자의 감각과 얹혀 생성되는 관계적 실천으로 인식할 수 있다. 이는 성악가가 단순한 발성 기술을 넘어, 목소리가 어떤 감각적 흔적을 남기고, 어떤 정동을 일으키며, 어떤 울림을 남기는지를 자각하게 한다. 나아가 AI 보컬과 비트박스의 사례는 성악가들에게 기존의 경계 바깥에서 목소리를 확장하고 새롭게 활용할 수 있는 상상력을 제공한다. 이러한 인식은 성악 연구 과제를 새롭게 설정한다. 전통적인 성악 이론은 주로 가수의 발성 기관과 음악적 해석에 초점을 두었지만, 오늘날 목소리를 연구한다는 것은 기술, 공간, 청취의 조건이 얹혀 있는 관계망 속에서 그 존재 방식을 탐구한다는 의미로 확장된다. 목소리를 둘러싼 이러한 조건들을 드러내고 비평하는 것이야말로 현세대의 중요한 과제이다.

포스트휴먼 시대의 목소리는 기술과 생명, 유기체와 비유기체가 교차하며 순간적으로 성립하는 진동이자 사건임을 이해할 때, 우리는 그것이 단일한 주체나 본질에 묶이지 않고, 항상 새로운 맥락 속에서 다른 얼굴을 드러낸다는 사실을 알게 된다. 이에 음악 연구는 목소리가 남긴 흔적을 해석하고 서술할 새로운 언어를 끊임없이 발명할 필요가 있다. 이러한 언어는 단순히 기술적 분석에 머무르지 않고, 청취자의 신체와 감각 속에서 발생하는 감응을 설명할 수 있어야 하며, 더 나아가 목소리를 둘러싼 사회적 규범과 기술적 조건까지 비평할 수 있어야 한다.

이 지점에서 글의 서두에서 던졌던 질문으로 돌아가 보자. “목소리는 어떻게 존재하는가?” 이 물음은 단 하나의 답으로 닫히지 않는다. 목소리는 매년

다른 방식으로 성립하며, 사건처럼 나타났다 사라진다. 따라서 이 질문은 답을 요구하는 동시에, 새로운 사유를 촉발하는 열린 물음으로 남는다. 목소리를 연구한다는 것은 이 열린 물음을 계속 이어가며, 목소리가 만들어내는 수많은 사건의 풍경을 기록하는 일이다.

한글검색어: 포스트휴먼 목소리, AI 보컬, 인공지능 노래, 비트박스, 목소리 존재론, 아상블라주, 시뮬라크르, 정동 이론, 확장된 신체, 보철

영어검색어: Posthuman Voice, AI-generated Singing, Vocal Ontology, AI Vocals, Beatboxing, Human Beatbox, Assemblage Theory, Simulacra (Baudrillard), Affect Theory, Extended Body, Prosthesis

참고문헌

- 강서희. “오페라 속 가사가 사라진 목소리의 의미와 역할.” 『서양음악학』 17/3 (2014): 11-38.
- 서동재. “휴먼 비트박스를 활용한 라이브 루핑 주법.” 『대중음악콘텐츠』 1/2 (2017): 36-44.
- Barad, Karen. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham: Duke University Press, 2007.
- Barthes, Roland. *Image-Music-Text*. Translated by Stephen Heath. London: Fontana Press, 1977.
- Baudrillard, Jean. *Simulacra and Simulation*. Translated by Sheila Faria Glaser. Ann Arbor: University of Michigan Press, 1994.
- Cusick, Suzanne G. “On Musical Performances of Gender and Sex.” In *Audible Traces: Gender, Identity, and Music*. Edited by Elaine Barkin and Lydia Hamessley, 25-48. Zürich & Los Angeles: Carciofoli Verlagshaus, 1999. [Originally presented as a conference paper in 1992].
- DeLanda, Manuel. *A New Philosophy of Society: Assemblage Theory and Social Complexity*. London: Continuum, 2006.
- Deleuze, Gilles and Félix Guattari. *A Thousand Plateaus: Capitalism and Schizophrenia*. Translated by Brian Massumi. Minneapolis: University of Minnesota Press, 1987.
- Fast, Heidi and Milla Tiainen. “Voice.” *Matter: Journal of New Materialist Research* 9 (2024): 1-5.
- Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press, 1999.
- Johnson, Julian. *Out of Time: Music and the Making of Modernity*. Oxford: Oxford University Press, 2015.
- Latour, Bruno. *Reassembling the Social: An Introduction to Actor-Network-*

Theory. Oxford: Oxford University Press, 2005.

Massumi, Brian. *Parables for the Virtual: Movement, Affect, Sensation*. Durham: Duke University Press, 2002.

Ong, Walter J. *Orality and Literacy: The Technologizing of the Word*. London: Routledge, 2002.

Wilson, Samuel. "The Composition of Posthuman Bodies." *International Journal of Performance Arts and Digital Media* 13/2 (2017): 137-152.

인터넷 자료

<https://m.blog.naver.com/PostView.naver?blogId=tnmusiclab&logNo=223981688068&proxyReferer=&noTrackingCode=true> [2025년 8월 27일 접속].

<https://www.youtube.com/watch?v=qlrpeYdm9Ec> [2025년 8월 27일 접속].

<https://www.youtube.com/watch?v=H6C7GXxLPlg> [2025년 8월 27일 접속].

YouTube video, <https://www.youtube.com/watch?v=LgDJFA3ktfI> [2025년 8월 27일 접속].

영상 자료

연번	영상 내용 / 링크	QR
1	수노 AI로 생성한 《몸은 없는데 숨을 쉬네》. https://m.blog.naver.com/PostView.naver?blogId=tnmusiclab&logNo=223981688068&proxyReferer=&noTrackingCode=true [2025년 8월 27일 접속].	
2	Wing. "Dopamine (Official Video)." YouTube video. https://www.youtube.com/watch?v=qlrpeYdm9Ec [2025년 8월 27일 접속].	
3	Space one. WING - Dopamine (DAW Cover) Synth, MIDI Cover Version. YouTube video. https://www.youtube.com/watch?v=H6C7GXxLP1g [2025년 8월 27일 접속].	
4	Seboom. "Beatbox Tutorial." YouTube video. https://www.youtube.com/watch?v=XBSPHZp0s60 [2025년 8월 27일 접속].	
5	SkilleR. BEATBOX TUTORIAL LIPROLL BASS SUB BASS. https://youtu.be/ilGhsSY4crw?si=B1K_ouMcIaNU4DhI [2025년 8월 27일 접속].	
6	MBC. "전지적 참견 시점 - 뽀빠이 출연 영상." YouTube video. https://www.youtube.com/watch?v=LgDJFA3ktfl [2025년 8월 27일 접속].	

국문초록

몸 없는 숨, 과잉된 몸:
AI 보컬과 비트박스로 본 포스트휴먼 시대
목소리의 존재론적 의미

강 서 희

본 연구는 “목소리는 어떻게 존재하는가”라는 물음에서 출발해, 바르트의 ‘결’, 옹의 현전, 보드리야르의 시뮬라크르, 마수미의 정동, 파스트, 티아이넨, 월슨의 기계와 신체의 얽힘/보철 개념을 종합해 포스트휴먼 시대의 목소리를 아상블라주, 즉 관계적 배치로 이해하고자 한다. 이를 위해 수노(Suno) AI로 생성한 《몸은 없는데 숨을 쉬네》와 비트박서 윙(Wing)의 《도파민》(*Dopamine*)을 살펴본다. 전자는 신체 없는 발화임에도 불구하고 실제처럼 들리는 시뮬라크르적 목소리를, 후자는 과잉된 신체를 통해 기술과 결합하며 다성적 사운드를 만들어내는 사례를 보여준다.

두 사례는 성악과 기악, 인간과 기계, 실연과 가상의 경계를 허물며, 목소리를 더 이상 개인의 소유물이 아닌, 관계적이고 사건적인 현상으로 재정의하게 한다. 이로써 이 연구는 청취는 감각과 정동의 층위에서, 연주는 장치와 규범, 공간의 얽힘을 고려하는 확장된 실천으로 나아갈 필요가 있음을 제시한다.

Abstract

Bodiless Breath, Excessive Body: Rethinking Vocal Ontology in the Posthuman Age through AI Vocals and Human Beatboxing

Kang, Seo-Hee

This study asks a fundamental question: “How does the voice exist?” and rethinks the ontology of the voice in the posthuman era. Drawing on Roland Barthes’s notion of the ‘*grain*’, Walter Ong’s theory of ‘presence’, Jean Baudrillard’s ‘simulacra (*simulacre*)’, Brian Massumi’s concept of ‘affect’, and the ideas of Heidi Fast and Milla Tiainen on the ‘entanglement of body and technology’ and Samuel Wilson’s notion of the ‘prosthesis’, the research frames the voice as an assemblage. From this perspective, the voice is not a fixed identity or private property, but a relational event that emerges through the interplay (*intra-action*) of bodies, technologies, environments, and listening practices.

Two case studies ground this discussion: the AI-generated song *Bodiless Breath* (2025) created with Suno AI, and beatboxer Wing’s performance *Dopamine* (2025). The first example illustrates how a voice without physical organs can still generate presence and function as a simulacral utterance. The second shows how an excessive yet extended performance renders the human vocal tract as a multi-layered, hybrid instrument, enacting distributed agency across body, microphone, and sound system. Both cases destabilize conventional binaries — vocal / instrumental, human / machine, live / virtual — demonstrating that the voice is best understood as contingent, processual, and event. In the posthuman condition,

listening shifts toward sensing the affective intensity and tactile grain of the voice, while performance becomes an expanded practice shaped by devices, norms, and spaces. This perspective redefines the ontology of the voice and opens new pathways for twenty-first-century vocal theory and practice.

[논문투고일: 2025. 08. 30]

[논문심사일: 2025. 09. 19]

[게재확정일: 2025. 09. 23]