인공지능 시대의 음악 기술

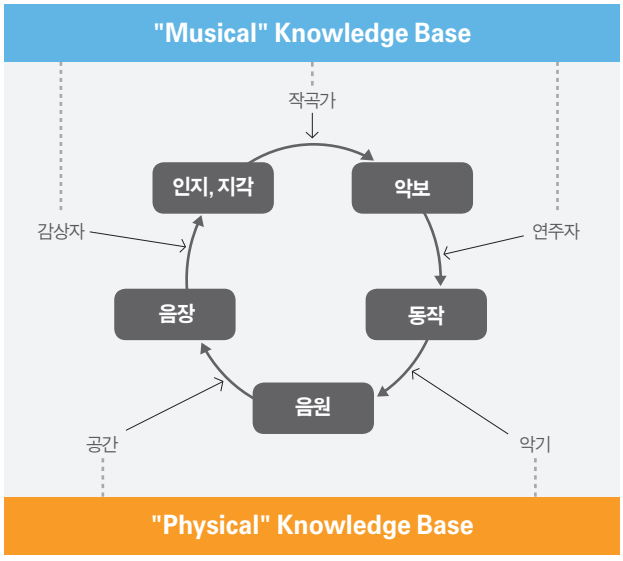
남주한 KAIST 문화기술대학원 교수

컴퓨터 음악을 전공한 나에게 몇 가지 각별한 책이 있다. 그 중 하나가 컴퓨터 음악가 리차드 무어(F. Richard Moore)가 저술한 〈컴퓨터 음악의 요소(Elements of Computer Music)〉다. 도서관에서는 음악 도서로 분류되지만, 내용은 공대에서 배우는 디지털 신호처리 과목에서 볼 수 있는 그림과 설명으로 가득 차 있어서 공학도에게 더 친근해 보이는 책이기도 하다. 그 책의 첫 번째 장을 펼쳐보면 유일하게 공학적이지 않은 내용이 있는데, 거기에는 아래와 같은 도식이 있다. 이 그림에서 리차드 무어는 작곡가, 연주자, 감상자 등 음악적 행위 주체와 악기, 공간 등 음악 소리의 발생과 전파에 관련된 물리적 객체를 어떤 입력 데이터 받고 새로운 형태의 출력 데이터를 발생하는 '프로세스'로 정의하고 있다. 예를 들어, 연주자는 악보를 입력 데이터로 취하여 동작 데이터를 발생하는 프로세스이고, 악기는 그 동작 데이터를 입력으로 취하여 음원 데이터를 발생하는 프로세스이며, 공간은 여러 악기에서 발생한 음원을 혼합하고 잔향 효과를 만들어 최종적으로 귀에 들리는 음장을 만드는 프로세스에 해당한다는 것이다. 이러한 관계는 음악이 존재했던 인류의 아주 먼 과거로부터 현재까지 여전히 유효한 인간과 음악 소리와의 보편적인 관계로 볼 수 있다. 하지만, 기술이 진화하면서 음악 데이터와 프로세스는 점차 다른 형태로 진화되어 왔다.

예를 들어, 우리가 귀로 듣는 소리인 공기중의 진동은 녹음 기술이 발명된 이후 LP, 테이프 등 물리적 매체에 기록되는 파형으로 표현되었고, 현재 우리가 감상하는 MP3나 스트리밍 음악에서는 0과 1로 이루어진 디지털 스트리밍 데이터로 표현되고 있다. 그 소리가 귀에 도달하기 전까지 여러 악기 소리를 섞고 잔향을 더하는 공간의 역할도 멀티 트랙 레코더를 거쳐서, 디지털 오디오 워크스테이션(DAW: Digital Audio Workstation)이라는 컴퓨터 기반의 사운드 편집 시스템으로 대체되어 음악 제작의 필수 도구로 자리 잡았다. 어쿠스틱 악기 또한 실제 악기를 매우 충실히 재현하는 샘플 기반 디지털 악기나 추상적인 알고리즘으로 소리를 합성하는 전자악기 신디사이저(Synthesizer)로 대체되어 왔고, 이는 주로 미디(MIDI: Musical Instrument Digital Interface)라는 연주 데이터를 통해 소리를 발생하고 있다.

요약을 하자면, 에디슨의 녹음 기술 발명 이후 현재까지 음악 기술은 컴퓨터가 악기와 방이라는 프로세스의 역할을 하고, 소리와 연주 데이터가 디지털 형태로 표현되는 등 아래 도식의 하단에 해당하는 물리적인 영역을 중심으로 진화하고 발전되어 왔다. 하지만, 21세기 들어서 인터넷의 발전으로 음악 데이터가 급격하게 증가하고 최근 인공 지능 기술이 발전하면서 컴퓨터의 역할이 점차 아래 도식의 상단에 있는 작곡자, 연주자, 감상자 등 인간의 영역으로 상위 이동하고 있다.

먼저 감상자의 영역을 살펴보자. 미국에서 시작한 '판도라(Pandora)'라는 인터넷 기반 라디오는 원하는 음악 채널을 개인이 만들면 그와 유사한 음악이 라디오처럼 연속으로 재생되어 나



음악적 행위의 주체와 객체를 음악 데이터(네모로 표시)와 프로세스(화살표로 표시)로 설명

출처 : 'Elements of Computer Music (F. Richard Moore 저)'

녹음, 믹싱, 마스터링 등 음악 제작을 위한 소프트웨어인 디지털 오디오 워크스테이션 (Digital Audio Workstation)

출처 : Wikipedia

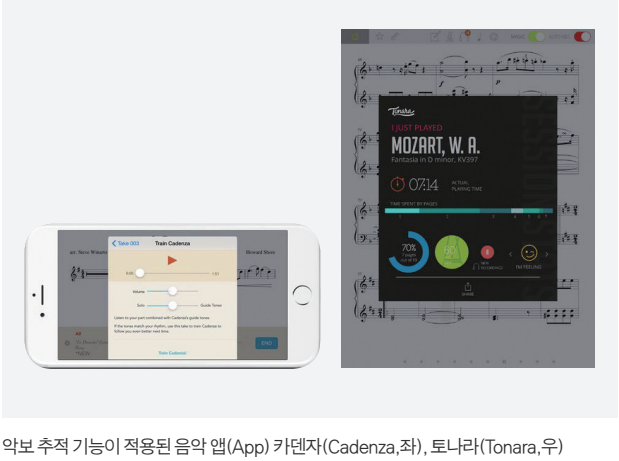
오고, '좋아요'와 같은 반응을 남기면 그 선호도에 맞추어서 개인화된 선곡을 해주는 서비스이다. 예를 들어, '쇼팽(Chopin)'이라는 채널을 만들었는데 '전주곡 마단조' 곡이 흘러나오고, 여기 '좋아요'라는 반응을 남기면 그 다음에 나오는 곡은 피아노곡이거나 단조의 곡이 될 가능성이 더 커지는 식이다. 이러한 개인화된 선곡은 판도라가 회사 설립 이전에 시작한 음악 게놈 프로젝트(Music Genome Project)를 바탕으로 하고 있는데, 이는 음악을 묘사하는 약 450개의 음악적 특성(예: 장르, 악기, 가수 음색 묘사 등)을 정의하고 모든 음원에 대하여 각각의 음악적 특성이 해당되는지 분석하는 작업이었다. 이러한 분석 결과를 이용하면 음원들 간의 음악적 유사성을 계산하여 쉽게 재생 목록을 만들 수 있고, 선호도에 따른 선곡 알고리즘도 음악적인 특성에 맞게 적용할 수 있다.

그런데, 약 450개나 되는 음악적 특성을 음원 서비스들이 제공하는 수백만 곡 이상에 대해서 어떻게 분석했을까? 음악 게놈 프로젝트는 이를 위해 전문 음악 '감상자'의 힘을 빌렸다. 즉, 많은 음악 전문가를 고용하여 한 곡 한 곡 유심히 들으면서 450개에 각각에 대한 항목을 일일이 체크하도록 했다고 한다. 평균적으로 한 곡당 20-30분의 시간이 걸린다고 하는데, 이를 통해 정확하고 양질의 음악 분석은 할 수 있겠지만, 하루에도 수 많은 신곡이 쏟아져 나오는 음원을 일일이 분석하기에는 시간과 비용이 많이 발생할 수 밖에 없다.

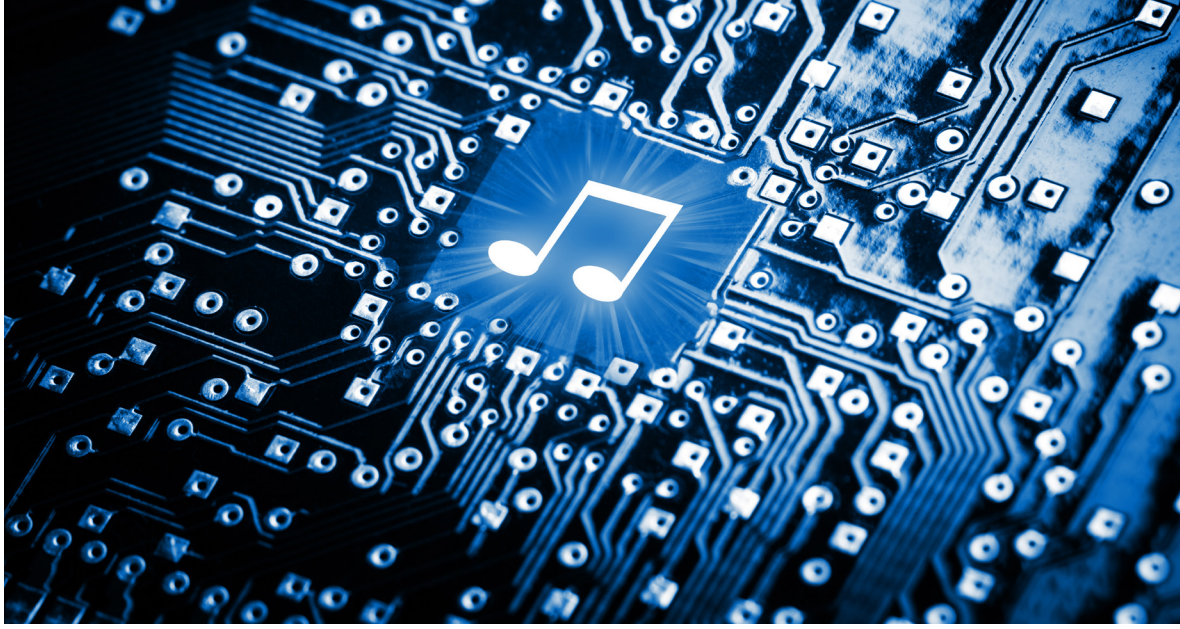
이에 대한 대안으로 가능한 것이 컴퓨터에 의한 음악 자동 태그 분석(Music Auto-Tagging)이다. 즉, 컴퓨터로 하여금 음악 감상자로서 음악 신호를 입력으로 취하고 음악 장르나 무드, 악기, 기타 곡의 특징 등에 대한 해당 여부를 직접 예측하도록 학습을 시키는 것이다. 최근 이미지 인식, 음성 인식, 자동 번역 등 인공 지능과 관련된 기술의 핵심이 딥러닝(Deep Learning)이라는 기계 학습 알고리즘인데, 딥러닝은 이러한 음악 자동 태그 분석에도 적용되어 최근 몇 년간 활발하게 연구 되어 왔다. 그 결과로 '구글 음악(Google Music)'이나 '스포티파이(Spotify)'와 같은 음원 서비스는 컴퓨터에 의한 자동 태그 분석을 음원 검색이나 추천 알고리즘으로 사용하기 시작하였고, 이를 통해 곡의 인기도나 특정 아티스트에 편중되지 않는 다양한 음악을 감상자들의 취향에 맞게 선곡해주고 있다. 컴퓨터에 의한 음악적 특성 분석은 무엇보다 계산 속도가 매우 빠르기 때문에, 상업 음원 뿐만 아니라 아마추어 뮤지션들이 제작하여 업로드 하는 사운드클라우드(soundcloud.com)나 밴드캠프(bandcamp.com)의 음원에도 쉽게 적용 가능하다. 물론 정확한 예측을 위해서는 음악 게놈 프로젝트와 같이 전문가가 양질의 특성 분석을 한 데이터를 많이 확보하고 이에 맞는 알고리즘을 계속 발전시켜야 하는데, 이는 음악 정보 검색(Music Information Retrieval) 이라는 분야에서 계속 활발히 연구 되고 있다.

두 번째는 연주자의 영역이다. 2010년경 필자는 UC 버클리 대학의 음악홀에서 재즈 기타 연주자인 팻 매스니(Pat Metheny)의 오케스트리온(Orchestrion)이라는 공연을 본 적이 있다. 일종의 프로젝트성 공연인데, 재미있는 점은 기타를 제외한 모든 악기가 솔레노이드(Solenoid)로 제어되는 로봇 팔에 의해 연주 되어 로봇이 인간 연주자를 대신한다는 것이다. 이처럼 인간과 로봇의 협연 또는 로봇들 만의 공연이 최근 몇몇 실험 연주 단체의 의해서 선보이고 있고, 인공 지능이 이제 악기도 인간처럼 연주한다고 언론에 소개되기도 한다. 하지만, 대부분의 경우 미리 입력된 연주 데이터를 그대로 재현하는 하는 수준이기 때문에 실제 연주자의 수준과는 거리가 멀다. 로봇이 '진정한 연주자'가 되기 위해서는 작곡가의 곡을 해석하여 자신만의 스타일을 만들거나, 다른 연주자와 협연 시 다른 악기 연주의 템포와 세기에 맞게 나의 악기 타이밍과 세기를 상황에 맞게 조절할 수 있어야 한다.

연주자 자신만의 스타일을 만드는 것은 사실 창조의 영역이다. 소수의 대가를 중심으로 언급되기 때문에, 인공 지능으로 대가의 스타일을 모방하거나 이를 뛰어 넘어서 새로운 연주 스타일을 합성하는 연구는 아직 초기 연구 단계에 있다. 하지만 컴퓨터가 악기 연주를 듣고 음표 정보를 예측하거나 현재 악보상의 위치를 추적하여 이에 맞게 연주를 하는 수준의 능력은 오래전부터 연구 되어 왔다. 악기 소리에서 음표 정보를 예측하는 것을 자동 채보 기



악보 추적 기능이 적용된 음악 앱(App) 카덴자(Cadenza,좌), 토나라(Tonara,우)
출처 : 구글 이미지 검색



술(Automatic Music Transcription), 주어진 악보에 대해서 위치를 파악하는 것은 악보 추적 기술(Score Following) 이라고 하는데, 역시 앞서 언급했던 딥러닝 기반의 음표 정보 예측이 학계에서는 활발히 연구되고 있다. 악보 추적 관련해서 최근 눈에 띄는 제품 중에는 소네이션(Sonation)이라는 회사의 카덴자(Cadenza)라는 앱이 있다. 이는 피아노나 오케스트라 반주 시 연주자가 자신만의 템포로 바이올린을 연주하면 그에 맞게 타이밍을 학습하였다가 다음 연주를 할 때는 바이올린 연주자의 맞추어서 자동으로 반주의 연주 속도와 타이밍을 조절한다. 또 토나라(Tonara)라는 앱은 내가 연주하고 있는 지점을 악보 상에서 정확히 파악하여 자동으로 악보를 넘겨주는 기능을 가지고 있다.

마지막으로는 작곡가의 영역이다. 컴퓨터에 의한 음악 작곡은 사실 '알고리즘 작곡법'이라는 생성 예술(Generative Art)의 한 분야로 오랫동안 다루어져 왔다. 현재까지 나온 음에 대하여 다음에 어떤 음이 나올지를 모델링하여 순차적으로 음을 계속 발생하는 방식인데, 1980년대 작곡가 데이빗 코프(David Cope)는 바하, 모짜르트 등 특정 작곡가의 스타일을 모방하는 자동 작곡 알고리즘을 개발하여 자동 작곡된 곡을 앨범으로 발매하기도 했다. 최근에는 소니 연구소(Sony CSL)에서는 '플로우 머신(Flow Machines)'이라는 알고리즘을 이용, 비틀즈의 데이터로 스타일을 학습하여 유사한 곡을 자동 작곡한 예를 선보였는데, 정말 비틀즈가 만든 것 같아서 많은 사람들의 감탄을 자아내기도 했다. 최근 구글에서는 '마젠타(Magenta)'라는 이름의 딥러닝 기반 자동 작곡 소프트웨어 라이브러리를 오픈 소스로 배포하여, 음악 데이터만 있으면 쉽게 누구나 쉽게 작곡할 수 있도록 하였다. 하지만, 이러한 스타일 기반의 작곡은 이미 주어진 데이터를 모델링하는 방식이기 때문에 순수한 창작이라기 보다는 모방에 더 가깝다고 할 수 있다. 즉, 창작의 과정에서 창작물에 대한 미적인 평가를 스스로 판단을 해서 작곡하는 것이 아니라, 단순히 이전에 '들은 적 있는 음악(즉 학습 데이터)' 과 유사하게 음을 발생시킨다는 것이다.

자동 작곡 기반으로 배경 음악을 만드는 회사인 쥬크덱(Jukedeck)의 CEO도 최근 인터뷰에서 비슷한 얘기를 한적이 있는데 이러한 점을 보완하기 위해서 뮤지션인 회사 직원들 이 자동 작곡된 곡을 직접 듣고 점수를 매기거나 사용자들의 다운로드 회수 등을 이용하여 간접적으로 음악에 대한 미적 판단을 학습에 적용하고 있다고 한다.

그렇다면 어떤 식으로든 이러한 창작과 관련된 요소를 데이터로 만들어 학습시킬 수 있다면 그때는 인간의 고유의 영역이라고 생각되는 창작마저 인공 지능에게 내어주게 될 것인가? 이에 대한 대답은 지금까지 음악 기술의 역사가 어느 정도 대답을 해준다고 생각한다. 전자 악기가 처음 나오고 기술적으로 점점 발전하면서, 사람들은 나무나 금속으로 만들어진 어쿠스틱 악기들이 수백 가지의 악기 소리를 연주하고 혼자서 오케스트라 반주도 할 수 있는 전자 악기들에 의해 대체될 것이라 걱정을 했다. 물론 어느 정도 영향을 받기는 했지만, 연주자들은 여전히 어쿠스틱 악기를 더 선호하고 있고, 작곡가들은 전자 악기를 어쿠스틱 악기의 대체품이라기 보다 새로운 음악을 만드는 도구로 삼아 이를 활용한 새로운 장르의 음악을 만들어 왔다. 마찬가지로 인공 지능기반의 음악 기술도 인간의 역할을 대체하기 보다는 새로운 창작을 위한 도구가 되거나 일상적인 음악 활동을 도와주는 방식으로 발전할 것으로 생각된다. SF