

음악제작을 위한 음성합성엔진의 활용과 기술

박병규*

요약

음악제작에 쓰이는 음성합성엔진은 악기 소리와 음색의 합성에 머물던 과거의 신디사이저와는 달리, 인간의 목소리를 각 음소에 따라 샘플화하여 탑재함과 동시에 각 음소의 연결을 주파수 영역 내에서 자연스럽게 처리함으로써 실제 사람이 노래하는 것과 같은 수준까지 도달하게 되었다. 사용자들은 이러한 음성합성엔진을 음악제작에 국한하여 쓰지 않고 캐릭터를 활용한 콘서트, 영상제작, 음반, 모바일 서비스 등 2차 창작물로 새로운 음악의 형태를 창조하며 문화적 패러다임을 바꾸어 나가고 있다. 현재 음성합성엔진 기술은 악보 편집기를 통하여 사용자가 원하는 음과 가사, 그리고 음악적 표현 파라미터를 입력한 뒤, 실제 가성 샘플을 데이터베이스에서 가져와 합성엔진에서 발음들을 조합, 연결하여 노래하는 것을 가능하게 한다. 이러한 컴퓨터음악 기술의 발전으로 인해 파생된 새로운 음악 형태들은 문화적으로 큰 반향을 불러일으키고 있다. 이에 따라 본 논문은 구체적 활용 사례를 살펴보고 합성기술을 탐색함으로써, 사용자들이 음성합성엔진을 이해하고 습득하는 데 기여함과 동시에 그들의 다양한 음악제작에 도움이 되고자 한다.

Application and Technology of Voice Synthesis Engine for Music Production

Byung-kyu Park*

Abstract

Differently from instruments which synthesized sounds and tones in the past, voice synthesis engine for music production has reached to the level of creating music as if actual artists were singing. It uses the samples of human voices naturally connected to the different levels of phoneme within the frequency range. Voice synthesis engine is not simply limited to the music production but it is changing cultural paradigm through the second creations of new music type including character music concerts, media productions, albums, and mobile services. Currently, voice synthesis engine technology makes it possible that users input pitch, lyrics, and musical expression parameters through the score editor and they mix and connect voice samples brought from the database to sing. New music types derived from such a development of computer music has sparked a big impact culturally. Accordingly, this paper attempts to examine the specific case studies and the synthesis technologies for users to understand the voice synthesis engine more easily, and it will contribute to their variety of music production.

Keywords : Voice synthesis engine, Music production, Digital sound

1. 서론

최근 일본에서는 ‘하츠네 미쿠(初音ミク)’라는

가상캐릭터가 등장하여 큰 반향을 일으키고 있다. 일본 최대 포털 사이트 야후 재팬(Yahoo Japan)의 ‘2007년 많이 찾은 검색어’ 8위로 부상한 동영상 사이트 니코니코동화(www.nicovideo.jp)에서는 하츠네 미쿠라는 캐릭터가 등장하면서 네티즌들에게 큰 관심을 불러일으키게 되었다 [1]. 이러한 현상은 각종 UCC 동영상과 만화책, 피규어²⁾는 물론 게임으로 개발하는 등 관련 상

※ 제일저자(First Author) : 박병규
접수일:2010년 05월 30일, 수정일:2010년 06월 14일,
완료일:2010년 06월 26일
* 경기대학교 전자디지털음악학과 조교수
bkmusic@kgu.ac.kr

품으로써 폭넓게 퍼져 나가고 있으며 홀로그램³⁾ 콘서트까지 이어지고 있다.

하츠네 미쿠는 원래 일본 음원 회사인 크립톤 퓨처 미디어(Crypton Future Media)가 야마하(Yamaha)의 음성합성엔진인 '보컬로이드(VOCALOID) 2'[2]를 채택하여 개발한 보컬 음원 소프트웨어 신디사이저의 마케팅 캐릭터 상품이다.

이것을 활용하여 유저들이 만든 곡이 UCC 사이트에서 널리 알려지게 되었는데, 이는 디지털 기술을 바탕으로 생성된 하나의 콘텐츠가 이례적으로 급속도로 파생되어 문화적으로 큰 반향을 일으킨 대표적인 경우이다. 따라서 본 연구에서는 하츠네 미쿠를 탄생시킨 보컬로이드의 활용 현황 및 파급 효과에 대해 알아보고, 그 기술적인 면을 연구하여 사용자가 음악제작에 있어 효과적으로 음성합성엔진을 사용할 수 있도록 기여함과 동시에 향후 기대효과를 전망해 보고자 한다.

2. 음성합성엔진의 활용

2.1 관련 소프트웨어

보컬로이드는 야마하에서 컴퓨터 음악 제작을 위한 음성합성기술 및 그것을 응용한 어플리케이션 소프트웨어로서 보컬(Vocal)과 안드로이드(Android)의 합성어이다. 사용자가 가사와 멜로디를 입력하면 미리 저장된 라이브러리를 통해 인공적으로 사람의 목소리로 된 노래를 합성해 낸다.

2004년에는 VOCALOID 1이, 2007년에는 VOCALOID 2가 개발되었다. 2009년에는 네트워크상의 서버로 동작하는 서비스 'NetVOCALOID'를 개발해 인터넷이나 휴대전화의 서비스 제공 전용으로 개시하였으며, 같은 해 일본 '산업기술 종합 연구소'[3]로부터 기술적인 허가를 받아 보컬로이드의 데이터 입력 지원에 응용 가능한 기술 'VocalListener(보컬리스너, 약칭「보카리스」)'를 공동 개발해 내었다.

야마하는 기술과 소프트웨어를 다른 회사들에

게 공식적으로 허가하고, 허가된 회사들은 자신들이 보유한 가수 라이브러리를 보컬로이드 소프트웨어와 함께 제작하고 있다.

VOCALOID 1의 음원으로는 'LEON', 'LOLA', 'MIRIAM', 'MEIKO', 'KAITO', 'SciArt'가 있으며, SciArt는 VOCALOID 1을 기반으로 하여 한국에서 개발된 가상의 여가수이다. VOCALOID 2의 음원으로는 '하츠네 미쿠(初音ミク)', '카가미네 린&렌(鏡音リン・レン)', '메구리네 루카(巡音ルカ)', 'GACKPOID(がくっぽいど)', 'MEGPOID(メグッポイド)', 'PRIMA', 'SONIKA', 'Sweet Ann', 'Big Al', 'SF-A2 개발코드 miki', '카아이 유키(歌愛ユキ)', '듀파모'가 있다.

2.2 새로운 음악의 형태와 창조

보컬로이드에서 대표되는 캐릭터는 하츠네 미쿠이다. 하츠네 미쿠는 컴퓨터 음악 소프트웨어 VOCALOID2 CV(Character Vocal)의 첫 번째 음원으로서 보컬로이드 2 엔진을 차용하여 제작되었으며, 데이터베이스는 여자 성우의 목소리를 기초로 했다.

하츠네 미쿠는 Otomania라는 아마추어 작곡가에 의해 제작된 'Ievan Polkka' 라는 곡이 일본의 동영상 사이트인 니코니코 동화(ニコニコ動画)에 등장하면서 큰 반응을 얻기 시작한다[4]. 이후 이를 변형한 2차 창작물이 다른 아마추어 작곡가들에 의해 만들어짐으로써 많은 동영상인 니코니코 동화와 유튜브(YouTube) 등을 통해 일본 및 여러 나라에 소개되었으며, 크립톤 퓨처 미디어사에서 저작권 가이드라인 제시로 무한 배포가 가능해졌다. 그리고, 공식 팬 사이트 '피아프로(ピアプロ)' (그림 1)를 개설함으로써 음악 소프트웨어로서는 보기 드물게 캐릭터를 마케팅에 활용한 성공적인 사례가 되었다.

2) 영화·만화·게임 등에 나오는 캐릭터들을 축소해 거의 완벽한 형태로 재현한 인형

3) 실물과 동일하게 입체적으로 보이는 3차원의 사진으로, 홀로그래피의 원리를 이용하여 입체상을 재현하는 간섭 줄무늬를 기록한 매체



(그림 1) 피아프로 홈페이지(http://piapro.jp)

이러한 네티즌의 큰 관심에 하츠네 미쿠의 음악은 점차 뛰어난 실력을 지닌 작곡가들이 참여하면서 메이지 음반 회사를 통해 음반으로 제작된다. 특히 ‘Supercell’이라는 팀의 음악이 니코니코 동화에서 큰 반응을 이끌어 소니 뮤직(Sony Music)을 통해 정식 앨범으로 제작되었다. 2009년 8월에는 하츠네 미쿠 2주년을 기념하여 그동안의 수록 곡들을 모은 앨범인 ‘Memories’와 ‘Impact’가 제작되었다(그림 2).



(그림 2) Impact(좌)와 Memories(우)
(Sony Music)

대중의 관심을 끌기 시작한 하츠네 미쿠는 각종 게임 및 애니메이션에도 출연하는 등 다방면에서 활동을 나타내기 시작한다. 특히 세가(SEGA)에서는 플레이스테이션 포터블(PSP) 전용으로 「하츠네 미쿠 -Project DIVA-」라는 제목의 게임을 제작하기에 이른다.

하츠네 미쿠의 파생 현상은 피규어 업계에도 나타나기 시작한다. 피규어 업체들은 다수의 하츠네 미쿠 피규어를 제작하게 되는데, 이 중 굿스마일 컴퍼니(Good Smile Company)에서는 하츠네 미쿠 피규어(그림 3)가 수차례 재판(再版)되면서 ‘녹색악마’라는 별명이 붙기도 하였다.



(그림 3) 하츠네 미쿠 피규어
(Good Smile Company)

출판계에서는 보컬로이드를 이용해 노래를 만드는 작곡기법 및 신디사이저 사용법 등의 기초 지식을 다루거나 하츠네 미쿠 뿐만이 아니라 다른 보컬로이드 캐릭터들을 만날 수 있는 「VOCALOID」 잡지도 발행되고 있으며, 공연계에서는 앞서 언급한 바와 같이 하츠네 미쿠를 주인공으로 한 홀로그램 콘서트까지 이루어지고 있다.

이러한 유행을 타고 네트워크를 통해 하츠네 미쿠의 목소리로 자신이 입력한 가사를 통해 노래를 들을 수 있는 서비스도 제공되고 있다. Net VOCALOID는 VOCALOID 엔진을 서버에서 실행하는 것인데, 사용자가 미리 지정된 곡 중에서 원하는 곡을 선택한 후 그에 맞는 가사를 입력하여 합성요청을 하면 완료되었다는 소식이 전해지면서 완성된 곡을 들을 수 있다. 즉, 지정된 멜로디에 맞추어 가사를 입력하면 휴대전화에서 사용자가 원하는 가사를 가지고 하츠네 미쿠가 노래하는 서비스이다.

기존의 보컬로이드는 노래에 특화되어 있었기 때문에 일반 언어로서 자연스럽게 말하는 합성은 어려웠다. 하지만 ‘VOCALOID - flex’[5]의 등장으로 음운과 운율의 세밀한 편집이 가능함에 따라 노래가 아닌 일상적인 언어도 표현할 수 있게 되었다. 구체적으로 보컬로이드에서 불가능했던 모음의 무성화나 탈락화가 표현 가능해져 자음의 길이, 소리의 높이 및 세기를 세밀하게 편집할 수 있게 되었고, 그로 인해 보다 사람의 목소리에 가까운 표현을 할 수 있게 되어 세세한 뉘앙스나 다양한 방언 등의 액센트와 인토네이션을 붙이는 것이 가능해졌다. 더욱이 VOCALOID-flex로 사용하는 음성 라이브러리는 새로운 음원을 추가하지 않고 그 활용이 가능해졌

다.

한편, CEATEC JAPAN 2009[6]에서는 일본의 '산업 기술 종합 연구소'에서 개발한 인간형 로봇 「HRP-4 C」가 하츠네 미쿠의 의상을 입고 '멜트(Melt)', 'World is Mine' 등의 대표곡을 부르기도 하였다.

3. 음성합성엔진 기술

3.1 과거의 음성합성 기술

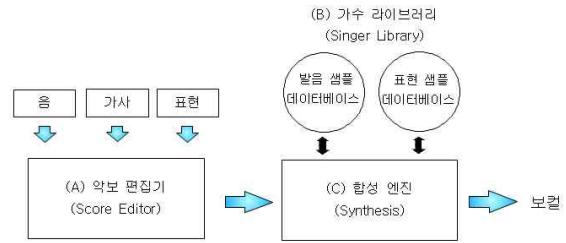
노래가 아닌 일반 언어로서 우리가 생각할 수 있는 음성합성의 방법에는 단어 혹은 그보다 작은 음절, 음소 단위로 음성을 녹음해 놓고 문장 생성시 이것을 연결하는 단순 연결합성 방식이 있다[7]. 그러나, 이 연결합성 방식은 인간의 음성기관의 움직임을 모델링한 조음합성 방식보다 음질은 좋지만 실제적으로 우리가 듣기에 자연스러운 합성음을 얻기에는 부족함이 있었다. 왜냐하면 같은 단어라도 각 음운구조의 길이, 음고(音高) 등 파라미터 값들이 위치에 따라 다양하게 변화하기 때문에 녹음된 발음들의 연결은 각기 일정한 값을 가진 음운들을 이어붙인 것 지나지 않았기 때문이다. 이를 해결하기 위해 고안된 것이 과형처리에 의한 음성합성 방법으로서 TD-PSOLA(Time Domain-Pitch Synchronous Overlap and Add) 방식이 있다[7][8]. 음성 과형을 그대로 이용하는 이 방식은 과형에서 음고 단위로 음성을 분해하고, 생성된 기본 주파수에 따라 분해된 음소각들을 재배열함으로써 음높이 조절이 이루어지며, 지속시간은 과형을 생략하거나 복제함으로써 조절된다. 이 방식은 'TTS(Text-to-Speech) 시스템'4)에서 활용되어 왔다.

3.2 가창(歌唱)을 위한 음성합성 기술

가창을 위한 음성합성엔진인 보컬로이드는 단순한 신디사이저가 아니라 사용자가 노래하는 목소리를 직접 제어해서 그 목소리를 음악제작에 사용할 수 있도록 통합적인 환경을 제공하는 기술이다.

보컬로이드는 (그림 4)와 같이 (A)악보 편집기(Score Editor), (B)가수 라이브러리(Singer lib

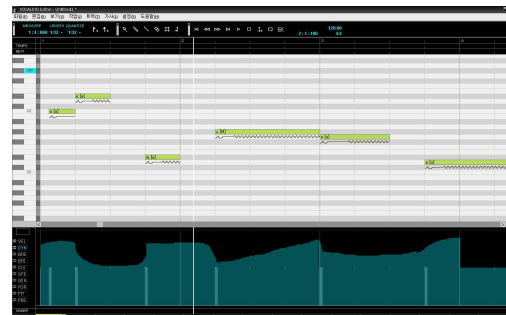
rary), (C)합성엔진(Synthesis Engine)의 세 가지 부분으로 구성되어 있다.



(그림 4) 보컬로이드의 시스템 개요도

3.2.1 악보 편집기

악보 편집기는 사용자가 본인이 원하는 음, 가사, 그리고 미묘한 감정표현까지 입력할 수 있도록 해준다. 이 악보 편집기는 보컬로이드에서 지원되는 가사를 일상적인 글로 입력할 수 있고 이를 기존에 내장되어 있는 발음사전을 통해 발음기호로도 입력할 수 있다. 만약 단어가 2개나 2개 이상의 음절을 구성하고 있을 때 편집기는 자동으로 음절들로 분해시킨다. 이렇게 입력된 데이터에 사용자는 비브라토(Vibrato)5)를 쉽게 넣고 취향에 따라 조절할 수도 있다. (그림 5)는 악보편집기 실행화면을 보이고 있다.



(그림 5) 악보 편집기

3.2.2 가수 라이브러리

가수 라이브러리는 실제 사람의 노래에서 추출한 데이터베이스의 샘플들이다. 이 샘플들은 발음되길 원하는 가사를 노래하기 위해 모든 가능한 언어의 음소의 조합을 포함해야만 한다[9]. 영어를 예로 들면, 모든 가능한 경우의 수인 '자

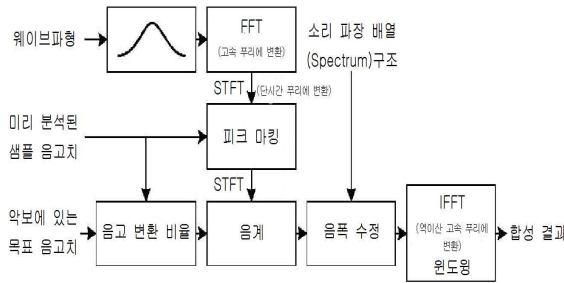
4) 기계와 인간이 서로 대화할 수 있도록 하기 위한 기초기술, 문장을 입력하면 자연스러운 인간의 음성으로 출력

5) 음악 연주에서 악기나 목소리의 소리를 떨리게 하는 방법

음 - 모음', '모음 - 자음', '모음 - 모음'이 녹음 되어 라이브러리에 저장되어야 한다. 사용자는 두 개 이상의 음소에 다음 글자를 선택하여 더할 수 있다. 모음들도 역시 라이브러리에 넣어 지는데, 모음들은 합성에 필수적인 통일된 모음들의 움직임을 복사하여 합성하는데 쓰인다. 이러한 샘플들은 1음 당 약 2000개 정도가 있다. 라이브러리는 3개의 음높이 범위의 샘플을 포함해서 사용자가 제시한 노래에 사용한다.

3.2.3 합성엔진

합성엔진은 제시된 데이터를 받아 라이브러리에서 필요한 샘플을 선택하여 연속적으로 연결시킨다. 아래 (그림 6)에서는 합성엔진의 진행과정을 보여준다.

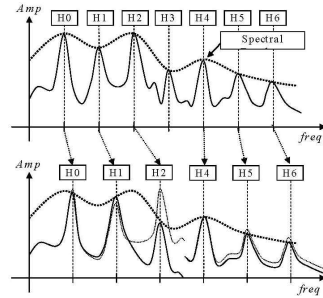


(그림 6) 합성엔진의 진행 다이어그램

미리 녹음된 샘플들은 다른 음의 높이와 다른 음성 문맥으로 녹음되어 있어 연속적으로 연결시키기 어려운 문제점이 존재한다. 이를 자연스럽게 연결하기 위해서는 음의 높이를 반드시 동일한 것으로 바꿔주어야 하며, 음색 또한 샘플의 연결로 부드럽게 해야 한다[10].

합성엔진에서는 음의 진동 범위에서 음의 높이 변화와 음색 조절을 한다. 음의 높이 변화는 음계 스펙트럼에서 일어나는데, Wave형태의 샘플인 STFT(Short Time Fourier Transform)를 받은 후, 스펙트럼상의 각 영역은 크기가 변경되어 음계요소가 음의 높이 변화와 일치하도록 조정된다(그림 7). 각 배음⁶⁾ 근처의 스펙트럼 모양은 그대로 유지된다. 음색 조절은 각각의 배음의 진폭의 변화로 만들어진다.

6) 발음체의 진동수가 근음(根音)의 2배·3배 등 정수배로 되는 음



(그림 7) 음높이와 음색 변환 처리[11]

한편, 음의 높이가 바뀌게 되면 그 위상도 바뀌어야 한다. 따라서 합성엔진은 완벽한 배음을 추정했을 경우, 다음의 보상하는 값을 i번째 배음의 위상에 더해서 산출한다.

$$\Delta \varphi_i = 2\pi f_0(i+1)(T-1)\Delta t \quad [11]$$

$T = f_0t$ (변환 후 값)과 f_0 (오리지널 값)의 음높이 변환 비율
 Δt = 지속 시간

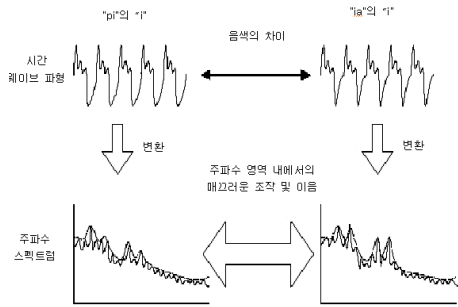
일정한 모음의 음색을 유지하기 위해서는 주변 샘플들의 스펙트럼 엔벨로프에 삽입하면 된다. 예를 들어, 단어 'get'의 g-e, e, e-t의 연속적으로 연결시키려면 [g-e] 끝의 [e]와 [e-t] 앞의 [e]를 스펙트럼 엔벨로프에 삽입하여 각 시간대마다 계속되는 모음 [e]의 일정한 음색을 만들 수 있다. 이러한 과정을 거치게 되면 촘촘히 음색을 연결하여 유지 할 수 있다. 자동적으로 처리되는 샘플 타이밍에 있어 음절 모음의 시작은 항상 음표가 입력된 상태에 있어야 한다.

3.2.4 주파수 영역 사용하는 방법

합성엔진은 발음 표현을 주파수 영역의 처리로 연결한다. 노래하기 위한 발음이 포함된 라이브러리들은 전문 가수들에 의해 만들어진 자료를 토대로 수집하여 데이터화 하고 이를 주파수 영역의 전환을 통해서 새롭게 만들 수가 있다.

원하는 발음을 위해 기존에 입력된 데이터에서 목소리 분절들을 검색하고 합성엔진에서 음높이의 전환이 적용된 다음, 자연스럽게 연결하는 과정을 거친 후 사용자가 발음되기 원하는 노래의 형태로 만들 수 있는 것이다. 이 과정은 주파수 영역에서 이루어지고 주어진 멜로디에 의하여 음높이는 쉽게 변화 될 수 있다. 또한,

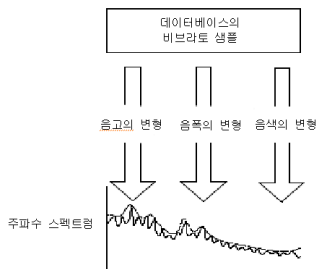
분절된 소리들은 자연스럽게 연결되어 재결합될 수 있다. 예를 들어 “piano”의 “pia”는 “pi”와 “ia”의 두 잘라진 음절들로 만들어질 수 있다. “i”와 “ia”의 모음 음색이 서로 다르기 때문에 단순히 소리들만 연결한다면 매끄럽게 들리지 않을 수가 있다. 이러한 부자연스러운 문제를 해결하기 위해 (그림 8)과 같이 주파수 영역을 이용한 연결 장치로 부드럽게 연결되는 과정이 이루어진다.



(그림 8) 주파수 영역 내의 처리과정

3.2.5 비브라토 과정

합성엔진은 비브라토와 같은 표현 효과를 얻기 위해 다양한 피치와 음색을 주파수 영역에서 전환한다(그림 9). 실제 가수의 목소리로부터 얻어진 비브라토를 데이터베이스화 하여 합성시 적용하게 된다.



(그림 9) 비브라토 처리과정

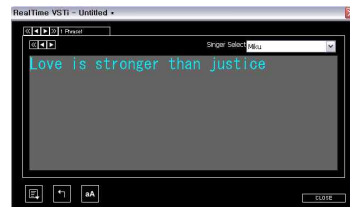
3.2.6 기능 및 표현 범위

보컬로이드에서는 사용자가 자신의 마스터 건반을 이용하여 MIDI 신호를 입력할 수 있다. 그리고, 미리 입력해 둔 가사(그림 10)를 기초로 건반으로 연주할 수 있는 리얼타임 연주기능(그림 11)이 가능함에 따라 음표를 일정시간 계속 누르면 실시간으로 그 음표에 부여되고 있는 음

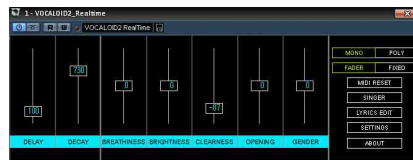
정과 가사로 노래할 수 있다. 또, 음절과 음소의 변경이 가능하고 입력되는 음표마다 정확한 타이밍을 위치시킬 수 있으며 비브라토의 종류를 선택, 적용 위치와 양, 진폭을 변경 가능하다. 최대 16트랙까지 가능하며 코러스나 하모니를 형성할 수도 있다.

세부적으로는 벨로시티(Velocity)⁷⁾, 다이내믹(Dynamic)⁸⁾, 포르타멘토(Portamento)⁹⁾, 레조넌스(Resonance)¹⁰⁾, 하모닉스(Harmonics), 노이즈(Noise), 성적 특성(Gender Factor), 피치 밴드(Pitch Bend), 밝기(Brightness), 호흡(Breathiness) 등을 직관적으로 조절 할 수 있다.

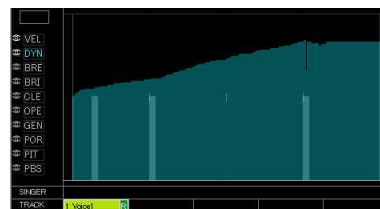
트랙은 투과하여 볼 수 있게 되어 동시에 2개의 파라미터 확인이 가능하다.(그림 12)



(그림 10) 가사 입력창



(그림 11) 리얼타임 기능 화면



(그림 12) 투과 컨트롤 트랙

또한, 트랙마다 실시간 합성과정을 통하여 일반적인 재생을 하는 것인지 이미 합성하여 렌더링 한 것을 재생하는지를 지정할 수 있으며, 편

- 7) 기악·성악에서의 최초의 발음이나 발성
- 8) 음악에서 셈여림을 이르는 말
- 9) 한 음에서 다른 음으로 미끄러지듯이 옮겨가는 것
- 10) 공명, 공진

집이 완료된 트랙을 렌더링 해두는 것으로 재생 시 부하를 줄일 수 있다. 그리고, 사용자의 임의적인 위치에서 라이브러리 가수의 변경이 가능할 뿐만 아니라, Standard MIDI 파일의 읽기 및 ReWire 디바이스로서 ReWire에 대응하는 다른 소프트웨어와 동기화가 가능하다.

3.3 보컬리스너(약칭: 보카리스)

사용자가 보컬로이드를 통해 좀 더 사실적인 인간의 가성을 합성해 내기 위해서는 수차례에 걸쳐 다양한 파라미터들을 다시 조정하고 들어보는 과정을 반복해야 한다. 또한, 사용자가 초기 지정된 가수 라이브러리를 통해 작업한 가성의 데이터를 가지고 다른 가수 라이브러리에 적용하고자 할 경우, 파라미터들을 새로운 값들로 조정해야 하는 어려움이 존재한다. 이러한 배경으로 등장하게 된 것이 보컬리스너(VocaListene r)이다.

보컬리스너는 일본 '산업 기술 종합 연구소'에서 개발한 보컬로이드의 파라미터 조정 지원 시스템이다. 이것은 사용자가 노래를 직접 부르며 입력하여 사용하는 것으로, 제시된 가성으로부터 노래하는 방법을 찾아 음높이와 음량 등에 관한 파라미터를 자동으로 추출 및 조정하여 보컬로이드의 라이브러리를 가지고 원음에 가깝도록 합성해 내는 기술이다.

보컬리스너는 원음에서 추출한 파라미터와 사용자가 발음 입력한 가사로부터 가성을 합성해 내고 결과물을 제시한다. 만약 원음을 흉내 내지 못한다면 파라미터를 다시 조정하며 합성한 뒤 결과물을 제시하는데 이러한 과정을 되풀이함으로써 사용자가 원하는 가성을 만들어 내는 것이다.

또한, 제시된 가성이 불완전할 경우 두 가지의 기능을 통하여 보완 할 수 있다. 첫째는 음높이를 변경하는 기능으로, 이것은 불완전한 음정이 존재하는 경우 부분이나 전체에 걸쳐 반음 단위로 음높이를 조정한다. 둘째는 노래하는 스타일을 변경하는 기능으로, 소리의 높이나 크기를 변경하고 비브라토를 자동적으로 추출하여 다른 요소들과 구분지어 가창 스타일을 변경 해주는 기능이다.

보컬리스너는 향후 보컬로이드 뿐만 아니라 가성 지각과 관련된 분야에서도 연구가 이루어

질 것이며, 인간의 가성을 다양한 관점에서 관찰 및 분석하는데 사용될 것이다.

3.4 그 외의 음성합성엔진

야마하의 보컬로이드 이외에 인공적으로 사람의 목소리로 된 노래를 합성해 내는 컴퓨터 음악 제작을 위한 소프트웨어로는 미국의 EAST WEST/Quantum Leap에서 개발한 'Symphonic Choirs[12]'가 있다.

Symphonic Choirs는 보컬로이드와 마찬가지로 사용자가 원하는 음과 가사를 입력하면 미리 저장되어 있는 라이브러리를 통해 인간의 가성으로 합성해 낸다. 이것은 MIDI 시퀀서에서 소프트웨어 가상악기의 형태로 합성엔진을 불러들이고, 사용자가 생성한 트랙에 가사 입력과 발음을 편집할 수 있는 '워드 빌더(Word Builder)'라는 별도의 프로그램에 Rewire 방식으로 연결한다.

솔로 음원으로 구성된 보컬로이드와는 달리 Symphonic Choirs는 합창(코러스) 합성을 목적으로 한다. 미리 저장된 라이브러리는 실제 합창 가수들을 통해 녹음된 데이터이며, 합창의 4성부(소프라노, 알토, 테너, 베이스)로 구분하여 수음(受音)되어 있다.

Symphonic Choirs는 포르타멘토, 레가토(Legato)¹¹⁾, 레퍼티션(Repetition)¹²⁾ 표현이 조절 가능하며, 비브라토의 경우 보컬로이드와 마찬가지로 사용자가 원하는 범위에서 조절 할 수 있다. 입력된 데이터에 대한 여러 가지 파라미터들은 편집기가 별도로 있지 않고 시퀀서에 의해 조절되므로, 보컬로이드처럼 투영방식의 직관적인 확인은 어렵다. 워드 빌더는 영어 단어를 입력하여 노래하도록 되어있지만 지원이 되는 다른 언어로도 표현이 가능하며, 자체적인 발음기호 입력 방식인 'VoTox'[12]를 사용하여 음소에 따른 컨트롤이나 타이밍, 볼륨, 크로스 페이드(Cross Fade)¹³⁾가 조절되어 원하는 발음을 만들어 낼 수 있다.

4. 결론

11) 계속되는 음과 음 사이를 끊지 말고 원활하게 연주하는 것

12) 악절 등의 반복

13) 앞 음악이 점점 사라지며 뒤 음악이 점점 나타남

악기의 소리나 음색 합성 위주에 머물던 기존의 신디사이저 기술은 비로소 인간의 언어를 통해 가성을 만들어내는 수준까지 도달하게 되었다.

라이브러리 제작자에 의해 제공되는 음원으로 사용자는 본인이 원하는 음, 가사 등의 음악적 요소들을 악보 편집기를 통해서 입력하고 편집할 수 있게 되었고, 세부적인 억양이나 비브라토, 소리의 질 등을 조절할 수 있게 되었다. 더 나아가 개발자나 음악 창작자가 아닌 일반 사용자도 좀 더 쉽게 사용할 수 있을 정도로 인터페이스가 발달하면서 문화적으로도 큰 반향을 불러일으키는 단계까지 이른 것이다.

하츠네 미쿠로 대표되는 보컬로이드는 인간의 가성을 모방해내는 컴퓨터음악의 기술이 2차 창작물을 생산해냄으로써 문화, 음악, 출판, 게임 등에 큰 파급효과를 불러일으킨 대표적인 사례로 기록될 것이다.

기존의 음악제작 시스템에서는 데모(Demo) 제작, 가수의 섭외 및 녹음에 상당한 시간과 비용이 소요 되었지만, 나날이 발전하는 음성합성의 기술로 인해 이러한 점들은 상당부분 축소될 것으로 보인다. 또한, 보컬로이드 뿐만 아니라 보컬리스너의 등장은 실제 가수가 없어도 완성도 높은 음악 제작이 가능한 시기가 곧 다가올 것임을 미루어 짐작할 수 있다.

비록 완전한 모방의 단계에 이르기까지는 다소 시간이 걸릴 수 있지만, 빠르게 발전하는 기술의 향상은 그 시일을 좀 더 앞당길 것이며, 거의 완벽에 가까운 단계가 머지않았음을 시사해 준다. 이에 따라, 컴퓨터음악 기술의 발전이 미치는 영향은 음악에서 뿐만 아니라 그 영역이 점점 더 확대되어 사회적, 문화적인 패러다임의 중심축으로도 더 넓게 자리 잡을 것으로 예상된다.

참 고 문 헌

[1] Masahiro Hamasaki, Hideaki Takeda, and Takuichi Nishimura, "Network Analysis of Massively Collaborative Creation of Multimedia Contents", uxTV'08 Silicon Valley, 2008.
 [2] <http://www.vocaloid.com/en/index.html>
 [3] Tomoyasu Nakano and Masataka Goto, "VOCALIS

TENER: A SINGING-TO-SINGING SYNTHESIS SYSTEM BASED ON ITERATIVE PARAMETER ESTIMATION", 6th Sound and Music Computing Conference Porto - Portugal, pp. 343-348, 2009.

[4] Masahiro Hamasaki, Hideaki Takeda, Tom Hope, and Takuichi Nishimura, "Network Analysis of an Emergent Massively Collaborative Creation Community", ICWSM Conference, pp. 222-225, 2009.
 [5] <http://www.yamaha.co.jp/news/2010/10022501.html>
 [6] <http://www.ceatec.com/2009/en/index.html>
 [7] F.J. Charpentier and M. G. Stella, "Diphone Synthesis Using An Overlap-Add Technique for Speech Waveforms Concatenation", Proc. ICASSP, Tokyo, pp. 2015-2018, 1986
 [8] E. Moulines, and F. J. Charpentier, "Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones", Speech Communication, vol. 9, no. 5-6, pp.453-467, 1990
 [9] Jordi Janer, Jordi Bonada, and Merlijn Blaauw, "PERFORMANCE-DRIVEN CONTROL FOR SAMPLE-BASED SINGING VOICE SYNTHESIS", DAFX-06 Montreal, pp. 41-44, 2006.
 [10] Peter Q. Pfordresher and James T. Mantell, "Singing as a Form of Vocal Imitation: Mechanisms and Deficits", ESCOM Jyväskylä, Finland, pp. 425-430, 2009.
 [11] Hideki Kenmochi and Hayato Ohshita, "VOCALOID - Commercial singing synthesizer based on sample concatenation", INTERSPEECH 2007, 2007.
 [12] <http://www.soundsonline.com/Symphonic-Choirs-Virtual-Instrument-PLAY-Edition-pr-EW-182.html>



박 병 규

1999년 : 맨해튼음대 작곡과 졸업

2001년 : 뉴욕대(NYU) 대학원 뮤직테크놀로지학과 졸업

현 재: 경기대학교 전자디지털음악학과 조교수
 관심분야 : 컴퓨터음악(Computer Music), 디지털 사운드(Digital Sound), 대중예술음악(Popular Music Composition)