DDSP 를 활용한 현대 악기 샘플을 기반 전통 악기의 음색 표현
Representing the Timbre of Traditional Musical Instruments Based On Contemporary Instrumental Samples Using DDSP

권유상
Yousang Kwon
울산과학기술원
UNIST
yk7244@unist.ac.kr

변종훈
Jonghoon Byun
울산과학기술원
UNIST
sjsh0925@unist.ac.kr

김선욱
Seonuk Kim
울산과학기술원
UNIST
d02reams@unist.ac.kr

고태영
Taeyoung Ko
울산과학기술원
UNIST
tyk0506@unist.ac.kr

윤주혁
Juhyeok Yoon
울산과학기술원
UNIST
heok95@unist.ac.kr

이경호
Kyungho Lee
울산과학기술원
UNIST
kyungho@unist.ac.kr

요약문

2020 년 구글의 마젠타 팀은 기존의 디지털 신호처리과정(DSP)을 미분가능한 단위로 해석하고 표현해 고품질 오디오 분석 및 합성이 가능하게 하는 Differential DSP (DDSP)[1] 라이브러리를 발표하였다. 본 연구는 DDSP 를 이용해 서로 비슷한 음색을 가진 서양악기들이 아닌, 서양악기 샘플을 입력으로, 우리 전통악기 중 하나인 거문고 소리의 합성값을 출력으로 얼마나 잘 생성할 수 있는지에 관한 탐색적 연구를 진행하였다. 먼저, 2 시간 분량의 국립국악원 거문고의 음조와 음색 데이터를 DDSP 를 통해 학습시켰다. 이후, 39 가지 서양악기를 이용해 국악에서 널리 쓰이는 5 음계인 중,임,무,황,태 샘플들을 생성했다. 합성결과의 음악적 특성이 얼마나 원음과 유사한지 비교하기 위해 서울대학교 예술과학센터에서 실제 거문고 소리를 녹음해 만든 5 음계 샘플과 DDSP 를 통해 생성한 샘플 간의 품질 차이를 분석하였다. 비교를 위해 MFCC 를 특징집합으로 한 39 개 서양악기의 5 음계 샘플과 거문고 원음의 5 음계샘플 간의 유클리드 거리를 계산하였다. 이를 통해 DDSP 가 거문고와 같은 전통악기의 음색 합성을 사용할 수 있는지, 가능하다면 어떤 조건에서 더 좋은 결과가 생성되는지를 분석하고 시사점을 논의하였다.

주제어

전통악기, 오디오 음색 변환, 신호 처리, 머신러닝

1. 서 론

국악기는 오랜 시간 우리 고유의 문화의 중요한 요소로 자리 잡아왔으나 현대에 들어서 교육기회의 제한 및 악기의 접근성 문제로 이전 세대보다는 널리 연주되거나 교육되고 있지 못하다 [4]. 그러나 거문고나 가야금과 같은 악기들은 여전히 한국적인 이미지를 표상하고, 정서를 환기시킨다는 측면에서 영화, 광고 등 많은 미디어에서 사용되고 있다.

전통악기는 악기로 연주하는 음악 뿐 아니라 그 악기의 발전과정도 전통 문화와 예술면에 있어 큰 의미를 가지고 있기 때문에, 이러한 전통악기와 음악의 보존과 사용은 중요한 의미를 가진다. 이에, 본 연구에서는 전통악기의 음조와 음색을 인공지능 기반 디지털신호합성 방법 중 하나인 DDSP 를 이용해 서로 비슷한 음색을 가진 서양악기 간의 오디오 음색의 변환이나 합성을 넘어서서, 서양악기 샘플을 입력으로, 우리 전통악기 중 하나인 거문고 소리의 합성값을 출력으로 지정했을 때, 그 음조와 음색이 얼마나 잘 합성되고 재현될 수 있을지에 대한 탐색적 연구를 진행하였다.

본 연구의 의의는 첫째, DDSP 를 활용해 현재에도 왕성하게 연주되는 국악기 중 하나인 거문고의 음조 및 음색합성 과정을 진행하고, MFCC 특징집합으로, 유클리드 거리분석을 바탕으로 한 정량적 비교방법을 제시하였다. 둘째, 거문고의 경우에 한정해 실험을 진행하였으나 어떠한 서양 악기군과 주법이 거문고의 음조와 음색 합성에 적합한지에 대해 총 7605 개의 전방위 적인 비교분석을 실시하였다. 셋째, 연구자들이 이해하기로 DDSP 를 이용한 국악기 음색 분석 및 합성과 관련된 최초의 연구이므로, 미래 관련 연구 및 연구자들에게 베이스라인 결과로 사용될 수 있다.

2. 문헌연구

문헌연구 과정을 통해 음색은 무엇인지, 전통악기 음색의 디지털 분석 방법과 지난 연구에 대해서

논의하고자 한다. 또, 디지털 신호처리의 새로운 방식 중 하나인 DDSP 에 대해서도 고찰한다.

2.1 음색 (Timbre)

음색은 소리의 특징 중 하나로 소리의 높이(진동수), 크기(진폭)과 함께 소리의 3 요소라 불린다. 음색이란 어떤 악기가 연주되기 시작하는 순간에 들리는 소리의 청각적인 신호에 대한 인식이며, 같은 음높이를 가진 악기의 경우 소리의 특징이나 질적인 차이를 나타내는 지표이기도 하다. 오랫동안 음색은 인간의 인지에 의해 정의되어 왔고 주관적인 평가로 음색을 구분해 왔다. 그러나 Jiang et al. 와 Virtanen et al.의 연구[5][6] 등에서 보여지듯 전자공학과 컴퓨터공학의 발전으로 음색을 정량적으로 정의하고 분석, 평가하려는 시도들이 이루어져 음악의 해석 뿐 아니라 신호를 디지털화 해 소리와 음악이 가지는 다양한 인문학적, 예술적 가치와 함의를 더 넓은 영역에 적용할 수 있는 단초를 마련하였다.

2.2 전통악기의 디지털 분석

지난 많은 연구에서 전통악기를 디지털 기술을 활용해 정량적으로 분석하려는 시도를 하였다. 기존의 연구들은 많은 부분 서양음악의 악기적 특성과 장르적 특성 등에 대해 주목하였지만, 최근 들어서는 클래식음악이라 불리는 전통적인 성야음악에서 제 3 세계 음악이나 월드 뮤직에 대한 연구도 활발히 이루어지고 있다. 예를 들면, Nurahmad et al.[7]의 연구에서는 인도네시안 다성음악에서 MFCC 를 활용해 어떤 악기가 쓰였는지 알아내는 시도를 하였고, Shete et al. 의 연구에서는[8] MFCC 를 특징 집함으로 활용한 오디오 분류기를 만들어 북부 인도 음악과 남부 인도 음악이 가지는 음색의 고유한 특징을 가려내는 방법을 고안했다.

2.3 DDSP

머신러닝 혹은 딥러닝 기술을 활용하여 소리를 합성하거나 변환하는 기술은 여러 방법으로 시도되었다. Cifka et al., Chang et al., Lyu et al., Oord et al., Lu et al., [9-13] 등의 연구에서 CNN 모델을 활용해 소리 또는 음악의 스타일 바꾸려는 시도를 해왔다. 그러나, 기존의 딥러닝 기술을 응용한 음악 분석 및 생성에는 많은 데이터를 요구하고 데이터가 가진 세부 특성에 따라 과적합 되거나 바이어스를 가지게 되어 음색의 합성과 생성에는 어려움이 있음이 확인되었다. DDSP 는[1] 마젠타에서 개한 신호처리 라이브러리로 전통적인 신호처리 방법(DSP)과 딥러닝 방법을 합쳐서 직관적이면서 편리한 사용법을 제공하면서 인풋으로서의 원음이 가진 특성의 큰 손실 없이 이용할 수 있다. 이를 통해 일반적으로 많이 쓰이는 DAW 프로그램에서 간편하게 플러그인으로 사용이 가능하며, 간단한 학습 과정을 거쳐 다양하게 확장이 가능하다. 또한 DDSP 플러그인 뿐 아니라 다른 DAW 이펙터나 플러그인들과 조합해서 원하는 소리로 보강할 수 있다. DDSP 는 자체적으로 이펙트의 정도를 조절할 수 있는 파라미터 조절 GUI 를 제공해 기존의 음악제작들이 쉽게 접근할 수 있다.

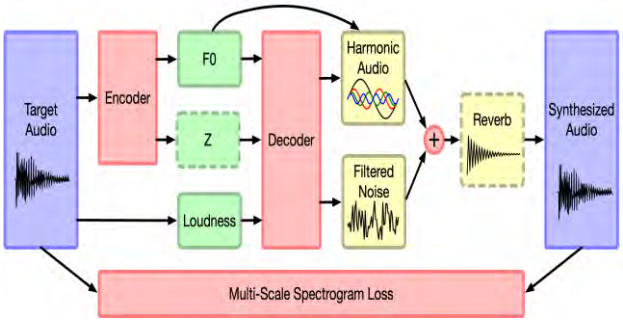

그림 1. DDSP 의 구성요소 ([1]에서 차용)

2.4 MFCC

MFCC 는 Mel Frequency Cepstral Coefficient 의 약자로 오디오 신호 혹은 소리데이터를 특징벡터화 하는데 비선형 mel 스케일 주파수를 소리 신호의 단기 로그 파워 스펙트럼의 선형 코사인 변형으로 표현한 것이다. 더불어 MFCC 는 인간의 청각적 한계와 가청 영역을 고려한 분석기법으로 소리의 정성적 특성을 정량화 시킬 때 많이 사용되어 왔다. Lalitha et al, 의 연구에서는[14] MFCC 로 음성에 실려 있는 감정을 측정했고, Mckinney et al.의 연구에서는[15] MFCC 를 활용해 음악을 분류하는 방법을 제시했다. 또한 Suksri et al., Muda et al.의 [16][17] 연구에서 사람 목소리를 분석하는데 MFCC 를 사용했다

MFCC 의 측정은 다음의 과정을 거친다.

1. 소리 신호의 푸리에 변환을 수행한다.
2. 삼각 중첩 또는 코사인 중첩을 이용해 1.의 스펙트럼 전력을 멜 스케일에 매핑한다.
3. 각 멜 주파수의 거듭제곱에 로그를 취한다
4. 멜 로그 거듭제곱들을 이산 코사인 변환(DCT)을 수행한다.
5. MFCC 는 최종 스펙트럼의 진폭이다.

본 연구에서 Python Librosa[3] 라이브러리를 이용해 MFCC 를 계산하였다.

3. 실험

DDSP 가 거문고의 음색을 변환하는데 어느 정도의 유사성을 보이는지 정량적으로 평가하고 그 결과를 토대로 어떤 악기를 변화했을 때 더 많은 유사성을 보이는지 그리고 그 음색들의 특징을 분석하고자 했다. 실험은 DDSP 학습 데이터 수집 – 모델 학습 – 샘플 생성 – 비교 연구를 위한 기준표 생성 – 음색비교 – 결과 분석의 단계로 이루어져 있다.

3.1 DDSP 학습 데이터 수집 및 학습

DDSP 를 개발한 Engel et al.의 연구에서 DDSP 를 학습시키기 위해서는 화성이나 다성 음악 보다는 단음 위주의 데이터를 최소 10 분 이상 수집하여 학습시키기를 권장하고 있다. 따라서, DDSP 의 학습데이터를 수집하기 위해 국립국악원 악기연구소의 국악 디지털 음원에서 정악거문고의 단음 음원을 수집했다.

크기는 약 2 시간 분량의 2GB 용량이었고 파일 모두 2304kbps wav 형식의 고음질 음원이다. Magenta 에서 DDSP 를 Google Colab 환경에서 간단하게 학습시킬 수 있도록 모델 학습 방법을 제공하고 있다. 구글드라이브에 학습 데이터를 저장한 후 Colab 환경에서 셀을 실행시키면 최종적으로 학습된 모델이 생성된다.

학습에 이용된 환경은 Google Colab 에서 Nvidia T4 GPU 를 이용했고 학습에는 약 30 분이 소요되었다.

3.2 샘플 생성

DAW 에서 DDSP 학습 모델은 설치된 가상악기에 플러그인으로서 작동하는 것이기 때문에 입력 가상악기에 따라 결과물이 다르게 나올 것이라는 가설을 세우고, 이를 위해 서로 다른 5 개 군의 18 개의 다른 악기 그리고 주법에 따른 변환 결과를 확인하기 위해 총 39 개 서로 다른 가상악기를 활용했다. 가상악기는 Fl Studiio 에서 기본 제공하는 음색을 이용했고 39 의 가상악기에 대해 국악기 대표 5 음(중, 임, 무, 황, 태) 각각 생성해 총 195 개의 샘플을 생성했다. 모든 샘플들은 120bpm 4 분음표로 100 퍼센트의 Velocity 로 입력했고, 320kbps mp3 파일로 추출했다. 비교를 위한 샘플은 서울대학교 예술과학센터의 거문고 가상악기를 같은 방식으로 입력해 각 음 별로 총 5 개의 파일을 생성했다.

표 1. 샘플 생성에 사용한 악기

악기군	피아노	목관	금관	현악기	기타	베이스기타
악기	Bosendorfer, Yamaha, Steinway	Bassoon Clarinet	Hron Trombone Trumpet Tuba	Violin Viola Cello Contrabass	Electric Nylon Steel	Bass 1 Bass2 Bass3

19 의 악기 준 들 중 악기에 따라 복수의 주법을 사용할 수 있는 악기들은 주법에 따라 샘플을 생성했다.

표 2. 복수의 주법을 사용한 악기

악기군	악기	주법
목관	Bassoon	Sustain
	Clarinet	Staccato
금관	Horn	Sustain
	Trombone	Staccato
	Trumpet	
	Tuba	
현악기	Violin	Sustain
	Viola	Staccato
	Cello	Pizzicato
	Contrabass	
베이스 기타	Bass 1	Fingered
	Bass 2	Picked
	Bass 3	Slap

3.3 샘플 특징 추출

모든 샘플들의 MFCC 를 추출하기 위해 Python Librosa 라이브러리를 이용해 MFCC 를 추출했다.

3.4 비교기준표 생성

거문고의 기준 샘플과 DDSP 를 통과한 비교 샘플들의 MFCC 를 구한 후 그 Euclidean Distance 의 의미를 파악하기 위해 Gonzalez et al., Wan, X. [19][20] 등의 연구를 참고하여 비교 연구 실험을 진행하였다.

MFCC 의 Euclidean Distance 가 현실의 상용되는 악기들 사이의 거리와 어느 정도 차이가 있는지 확인하고자 했다. DDSP 를 통과하기 전 39 의 샘플들끼리 MFCC 의 Euclidean Distance 를 측정했고 그 결과 5 개음 39*39 개 총 7605 개의 비교 값을 생성했다.

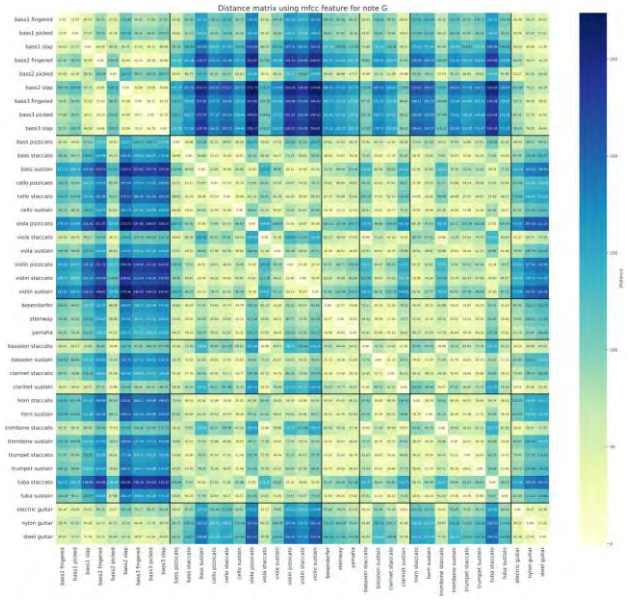

그림 2. 중(G)음의 MFCC 비교기준표
(푸른색에 가까울수록 유사도 낮고, 노란색에 가까울수록 유사도가 높음)

3.5 유클리드 거리를 이용한 샘플 비교

DDSP 를 통과한 각 음 별 39 개의 샘플과 서울대학교 거문고와 비교를 진행하였다. 비교할 두 샘플들의 MFCC 데이터를 추출한 후 거문고 원음과 합성된 결과의 Euclidean Distance 를 측정해 분석하였다.

4. 결과

서울대학교에서 만든 샘플링 된 거문고 가상악기와 DDSP 를 통과한 가상악기의 샘플들의 MFCC 의 Euclidean Distance 를 측정한 결과는 표 3 과 같다.

표 3. 샘플링된 거문고 가상악기와 DDSP 를 통과한 악기 샘플의 비교 결과

악기	G	A	C	D	E
bass_pizzicato	47.08	18.21	57.59	94.03	180.50
bass_staccato	90.28	55.02	102.93	119.90	171.01
bass_sustain	114.40	63.31	56.40	77.34	120.49
bass1_fingered	60.11	104.55	28.60	30.19	22.65
bass1_picked	85.24	99.65	38.66	22.12	34.65
bass1_slap	97.54	106.45	100.58	81.58	61.04
bass2_fingered	110.54	160.31	92.19	78.72	60.16
bass2_picked	85.07	91.67	22.89	41.42	16.54
bass2_slap	122.86	159.97	71.39	54.67	29.72
bass3_fingered	115.74	155.69	101.20	78.72	58.27
bass3_picked	123.38	139.96	94.45	81.09	61.65
bass3_slap	124.41	158.32	24.86	24.12	19.89
bassoon_staccato	33.65	35.34	87.31	71.53	98.42
bassoon_sustain	83.28	43.93	65.43	91.85	118.84
bosendorfer	78.44	61.43	71.18	89.46	100.82
cello_pizzicato	90.28	65.58	85.42	108.71	97.08
cello_staccato	88.87	56.55	35.98	74.66	110.27
cello_sustain	53.37	91.99	143.60	134.13	204.33
clarinet_staccato	82.30	43.80	82.40	116.81	126.96
clarinet_sustain	22.27	57.22	15.36	26.50	54.32
electric_guitar	21.41	47.54	9.93	36.93	45.29
horn_staccato	92.25	60.57	69.64	107.15	110.36
horn_sustain	69.81	22.12	33.62	47.33	78.53
nylon_guitar	40.83	65.23	36.79	46.77	64.11
steel_guitar	79.71	129.79	102.41	47.22	17.59
steinway	75.47	60.76	77.19	99.88	124.56
trombone_staccato	32.57	27.79	38.64	74.61	94.01
trombone_sustain	89.19	36.51	49.97	53.34	84.06
trumpet_staccato	127.08	55.04	87.52	120.96	159.30
trumpet_sustain	72.46	33.28	48.87	80.79	170.72
tuba_staccato	88.54	69.56	98.63	123.14	146.70
tuba_sustain	91.72	50.04	82.45	80.82	77.32
viola_pizzicato	158.37	104.32	99.83	123.99	92.80
viola_staccato	33.70	25.13	75.53	107.94	64.34
viola_sustain	79.51	79.49	91.84	114.97	148.62
violin_pizzicato	126.84	50.06	82.29	110.67	137.81
violin_staccato	170.18	143.58	152.44	166.75	198.06
violin_sustain	144.27	65.02	109.17	138.14	189.96
yamaha	93.24	50.83	94.38	114.73	131.86
Average	87.08	75.53	72.30	84.45	99.58
Minimum	21.41	18.21	9.93	22.12	16.54
Maximum	170.18	160.31	152.44	166.75	204.33

5. 결과 및 시사점

DDSP 와 서울대학교 국악기 가상악기를 이용하여 대중적으로 많이 쓰이는 서양악기가 국악기의 음조와 음색을 얼마나 잘 표현할 수 있는지 생성된 샘플들의 MFCC 의 Euclidean Distance 를 측정하여 탐색적으로 연구하였다.

5 개의 음조의 최소값은 표 3 에서 붉은색으로 표시된 것처럼 각각 21.41, 18.21, 9.93, 22.12, 16.54 로 나타났다. 5 개 음조의 최대값은 푸른색으로 나타냈으며 각각 170.18, 160.31, 162.44, 166.75, 204.33 으로 나타났다. 비교 기준표에서 같은 악기군들끼리의 비교 값과 유사하게 나옴을 확인했다. 악기 기준표에서 30 미만의 상대적으로 작은값(원음과 유사한 값)들은 악기 군들은 같으나 주법만(pizzicato, staccato, sustain 등) 다른 샘플에서 나타나는 경우가 많았다.

비교 결과에서 유사도 수치가 가장 좋은 악기들은 대부분 기타와 베이스 기타로 나타났다. 같은 악기 중에서 주법 및 표현에 있어서는 음을 유지하는 sustain 기법 보다는 현을 뜯거나 짧고 간결하게 연주하는 pizzicato 혹은 picked 기법에서 유사도가 높은 경향을 나타냈다. 이는 L1 Spectrogram Loss 를 사용하는 DDSP 기술적 특성에 기인하며, 학습된 DDSP 모델을 악기가 통과할 때 원래의 악기 소리를 반영하기 때문에 거문고와 비슷한 뜯거나 튕기는 주법의 악기들이 유사한 결과를 보여줄 수 있음을 시사한다.

더불어 결과값에서 평균 이상의 값들은 보여준 악기들은 악기의 재질이 다른 금관악기들과 바이올린처럼 거문고와 음역대가 다른 악기 그리고 주법이 거문고와 상이한 경우였다. 이는 DDSP 국악기의 음색분석 및 합성에 활용하기 위해서는 적절한 입력악기를 선택하고 또한 주법이 비슷한 경우를 선택해야 목표하는 출력값에 가깝게 나올 수 있음을 시사한다.

6. 결론 및 미래연구

본 연구에서는 AI 기술을 활용한 DDSP 모델로 변환된 음색이 목표하고자 하는 음색과 얼마나 유사한지 탐색적으로 연구하였다. 그러나 본 연구는 연구자들의 경험과 전문성을 바탕으로 선택한 39 개의 서양악기를 입력으로 국악기 중 현을 사용한 거문고만을 출력으로 설정한 한계를 지닌다. 또, 정량적 평가와 사람이 인식하는 음색의 유사도의 관계에서는 평가하지 못했다는 한계를 지닌다. 미래연구에서는 이러한 한계점이 보완되어야 할 것이며, MFCC 만이 아닌 보다 다양한 특징집합 생성을 통한 다면적, 다층적 유사도 평가가 이루어져야 할 것이다.

참고 문헌

1. Engel, J. DDSP: Differentiable Digital Signal Processing. In International Conference on Learning Representations. 2020.

2. Center for Arts and Technologies Seoul National University http://en.catsnu.com/Main/Main.aspx 2014.

3. Librosa Library. https://librosa.org/. 2023.

4. Shi, Q. The Study on the Development of Traditional Music in Internet Age. Proceedings of the 5th International Conference on Algorithms, Computing and Systems. 2022.

5. Jang, W. Analysis and Modeling of Timbre Perception Features in Musical Sounds. Applied Sciences. 2020.

6. Virtanen, T. Computational Analysis of Sound Scenes and Events.2018.

7. Nurahmad, A. Identifying traditional music instruments on polyphonic Indonesian folksong using mel-frequency cepstral coefficients (MFCC). Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia - MoMM '12.2012.

8. Shelke, A. An Effective Feature Calculation For Analysis & Classification of Indian Musical Instruments Using Timbre Measurement. Proceedings of the Sixth International Conference on Computer and Communication Technology 2015.2015.

9. Chifka,, O. Groove2Groove: One-Shot Music Style Transfer With Supervision From Synthetic Data. IEEE/ACM Transactions on Audio, Speech and Language Processing.2020.

10. Chang,Y. Semi-supervised Many-to-many Music Timbre Transfer. Proceedings of the 2021 International Conference on Multimedia Retrieval. 2021.

11. Lyu, Y. Convolutional Neural Network based Timbre Classification. Proceedings of the 2020 International Conference on Cyberspace Innovation of Advanced Technologies. 2021.

12. Oord, A. WaveNet: A Generative Model for Raw Audio.2016

13. Lu, C. Play as You Like: Timbre-enhanced Multi-modal Music Style Transfer. 2018.

14. Lalitha, S. Emotion Detection Using MFCC and Cepstrum Features. Procedia Computer Science. 2015.

15. McKinney, M. Features for Audio and Music Classification. 2003.

16. Suksri, S. Speech Recognition using MFCC. Proceedings of the International Conference on Computer Graphics, Simulation and Modeling. 2015.

17. Muda, L. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. Journal of Computing Volume 2, Issue 3. 2010

18. FL Studio. https://www.image-line.com/. 2023

19. Gonzalez, Y. Similarity of Musical Timbres Using FFT-Acoustic Descriptor Analysis and Machine Learning. MDPI. 2023.

20. Wan, X. A Comparative Study of Cross-Lingual Sentiment Classification. Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01. 2012.