

허밍을 이용한 고품질 음악 생성 Humming based High Quality Music Creation

이윤재† ‡ · 김선민*
Yoonjae Lee and Sunmin Kim

Key Words : Humming (허밍), Melody Transcription (멜로디 전사), Automatic Music Creation (자동 음악 생성)

ABSTRACT

In this paper, humming based automatic music creation method is described. It is difficult for the general public which does not have music theory to compose the music in general. However, almost people can make the main melody by a humming. With this motivation, a melody and chord sequences are estimated by the humming analysis. In this paper, humming is generated without a metronome. Then based on the estimated chord sequence, accompaniment is generated using the MIDI template matched to each chord. The 5 Genre is supported in the music creation. The melody transcription is evaluated in terms of onset and pitch estimation accuracy and MOS evaluation is used for created music evaluation.

1. 서 론

작곡은 음악적 지식을 가지고 있는 전문가가 하는 것으로 주로 인식이 되어 왔으며 실제로 음악적 지식이 없는 일반인들에게는 매우 어려운 영역이라 할 수 있다. 하지만 음악적 지식이 없는 일반인들도 음악을 만들고자 하는 욕구는 가지고 있다. 이러한 욕구를 만족시키기 위하여 최소한의 정보를 이용하여 음악을 자동으로 생성시켜주는 연구가 최근 많이 이루어 지고 있으며 스마트폰 어플리케이션 형태의 상용화 제품도 나오고 있다⁽¹⁻⁴⁾. 음악적 지식이 없는 일반인이 쉽게 사용할 수 있는 작곡의 도구로 허밍을 예로 들 수 있다. 허밍은 누구나 쉽게 할 수 있으며 허밍을 통해 주 멜로디를 만드는 것은 일반인도 쉽게 할 수 있는 방법이다. 이미 상용화 된 허밍 기반의 어플리케이션의 경우, 설정한 Beats Per Minute(BPM)에 따라 메트로놈이 생성되며 사용자는 메트로놈에 따라 허밍을 하는 경우가 대부분이

며 발생한 허밍이 그대로 음악에 믹싱되는 경우가 대부분이다⁽⁵⁾. 이 경우 화성적으로 어울리지 않는 멜로디가 발생할 수도 있다. 본 논문에서는 메트로놈 없이 자유롭게 허밍한 경우에 대해 악보를 추정하며 추정된 정보를 이용하여 사용자 허밍에 어울리는 음악을 생성하는 것에 대해 기술한다. 자유로운 허밍으로 인한 박자에 어긋나는 멜로디와 음악적 화성에 어울리지 않는 멜로디는 음악 완성도를 높이기 위해 자동으로 보정되는 알고리즘도 적용하였다. 2 장에서는 알고리즘에 대해 설명하고 3 장에서는 실험 결과를 기술하며 4 장에서 결론을 언급한다

2. 고품질 음악 생성

2.1 음악 생성 과정

허밍 기반 음악생성 과정은 그림 1과 같다. 입력된 허밍을 분석하여 악보 추정(Melody Transcription) 및 최소 음표 길이 (Minimum Note Length)를 추정하며 입력된 장르 (genre) 기반으로 마디(Bar)와 Chord 를 추정한다. 추정된 chord 를 이용하여 반주음(Accompaniment) 를 생성하며 입력된 허밍 멜로디가 반주음과 더욱 잘 어울리도록 하기 위해 음길이 및 음높이 보정을 하여 반주음에

† 교신저자; 이윤재, 삼성전자 DMC 연구소
E-mail : yj0604.lee@samsung.com
Tel : 031-279-9442

‡ 발표자; 삼성전자 DMC 연구소

* 공동저자 소속 : 삼성전자 DMC 연구소

믹싱하여 완성된 음악을 생성한다.

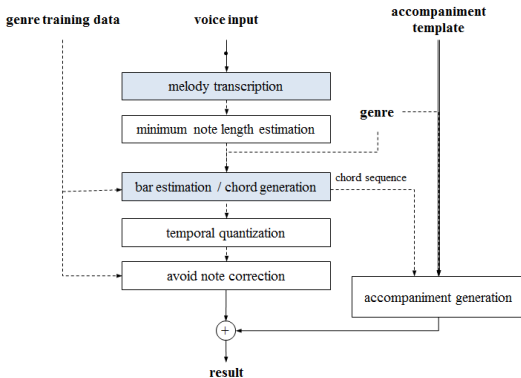


FIGURE 1 Music Creation Block Diagram

2.2 악보 추정 기술

허밍 입력 신호의 멜로디 추정을 위해 Musical Instrument Digital Interface (MIDI) 악보 추정을 한다. 멜로디의 주요 정보는 Onset, Offset, Pitch 이다.

허밍시 멜로디 음이 바뀔 때 음높이가 바뀌어 Pitch가 바뀌거나 발성이 끊어지는 구간이 생기게 된다. 이 경우 음성 스펙트럼 상에서 주요 하모닉 값이 바뀌게 되거나 주요 에너지가 끊기게 되는 구간이 생기게 되어 이전 프레임들의 주파수 대역의 Magnitude 값과 비교했을 때 크게 변하게 된다. 이러한 특성을 이용하여 본 논문에서는 주파수 신호에서의 하모닉 peak 위치 변화량을 이용하여 변화량이 threshold 이상 발생하는 부분에서 Onset, Offset 설정을 하였다. Onset, Offset 이 정해지면 Onset, Offset 사이 구간에서 Pitch 를 추정하게 되며 본 논문에서는 Cepstrum 기반의 Pitch 를 추정하였다. Cpestrum 기반의 Pitch 는 간단한 방법으로 입력 신호를 Fourier Transform 을 통해 주파수 축 신호로 변환 한 후 Magnitude 에 Logarithm 을 취한다. 이 신호를 다시 Fourier Transform 을 취하여 Cepstrum 영역으로 신호를 변환한 후 Magnitude 값이 최대가 되는 위치가 Pitch 를 결정하게 된다.

2.3 음악 생성 기술

(1) 최소 음길이 추정

음악을 생성하기 위해서는 각 Note 의 음표, 마디, 템포, Chord 등의 정보가 필요하며 이를 위해 가장 먼저 최소 음표를 설정해야 된다. 음의 길이는 Offset 위치와 Onset 위치 차이로 정의되

며 전체 허밍 입력 신호로부터 음의 길이에 대한 Histogram 을 구한다. 이상적인 음표는 최소음 길이의 정수배의 음표로 구성되어 있기 때문에 Histogram 에서 Correlation 을 이용하여 가장 큰 주기성을 가지게 되는 구간을 찾아 최소음표를 추정한다. 그림 2는 알고리즘 흐름에 따른 결과를 나타낸 것이며 Histogram에 Hamming Window 를 적용하여 Smoothing 을 한 후, Correlation 을 구하였다.

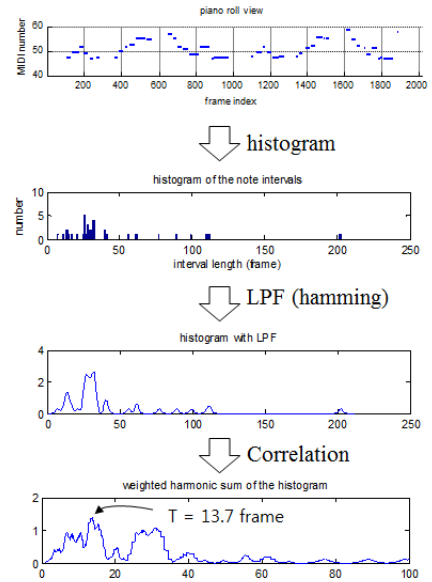


FIGURE 2 Minimum Note Length Estimation

(2) Chord Sequence, 마디, 템포, Key 추정

4/4박자 기준으로 최소음표를 16/8/4분음표로 후보 설정을 한다. 최소 음길이의 정수배 되는 위치들 중에서 추정된 Onset과 가까운 위치로 Onset을 이동 시킨 후, 모든 음표의 경우의 수에 대해 마디 위치를 후보로 찾는다. Chord 추정을 위해 Hidden Markov Model (HMM) 을 이용하며 장르별로 24개(Major 12개, Minor 12개) State로 구성된 HMM 을 이용하여 모든 음표 및 모든 마디에 대해 가장 높은 확률을 가지는 Chord Sequence 를 찾는다. HMM 인식에 사용되는 특징은 한 마디안에 있는 Note Number들의 누적량이 된다. 또한 인식된 결과 Chord Sequence를 가지는 마디, 최소 음표, 템포를 최적의 값으로 결정한다.

HMM 파라미터 훈련에 사용된 Database는 wikifonia.org 에서 제공하는 악보 데이터를 사용

하였다. Music XML 포맷의 파일로 chord 와 Note number 가 모두 라벨링 되어 있는 장르별 로 1000개의 Database 를 사용하였다.

그림 3은 마디와 Chord Sequence 추정 개요도 를 나타낸다.

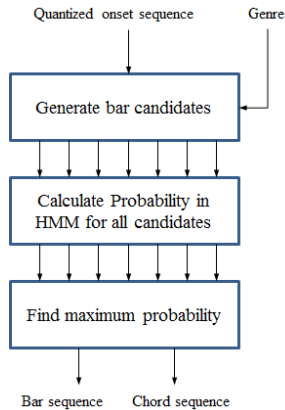


FIGURE 3 Chord/Bar Estimation

(3) 양자화

입력 허밍을 그대로 악보화 하여 반주 생성을 하고 생성된 반주에 멜로디를 믹싱하면 음악적으로 어색할 수 있다. 멜로디 길이가 정확히 정수배의 길이에 맞지 않아서 어색하거나 화성에 어울리지 않는 음에 의해 반주와 화성적으로 어울리지 않는 허밍을 예로 들 수가 있다. 이를 보정하여 보다 완성도 높은 음악을 만들기 위해 멜로디 길이, 위치 및 음 높이 자동 보정 (Avoid Note Correction)을 한다.

멜로디 음 위치는 최소 음길이의 정수배 정보를 이용하여 가까운 위치로 보정하며 음 길이 역시 반올림을 통해 가까운 정수배 위치까지 음 길이를 증가시키거나 감소시킨다.

멜로디 음 높이는 추정된 Chord 를 기반으로 HMM 에 사용된 Gaussian Model 을 사용한다. 입력음 높이를 기준으로 보정 허용치 구간 내에서 가장 확률이 높은 음을 선택하여 주어진 Chord 에 가장 잘 어울리는 음높이로 보정한다. 본 논문에서는 Note Number 2개 차이를 허용구간으로 설정하였다. 그림 4는 음높이 보정에 대한 전체 개요도를 표시한 것이며 그림 5는 음 길이, 높이 양자화 결과를 그린 것이다. 그림 5의 위 그림은 양자화 하기 전의 입력 Note 를 나타내며, 아래 그림에서 파란색은 길이 양자화 된 Note이며, 빨간색은 길이 양자화 후 음 높이가 보정까지 된 Note 를

나타낸다.

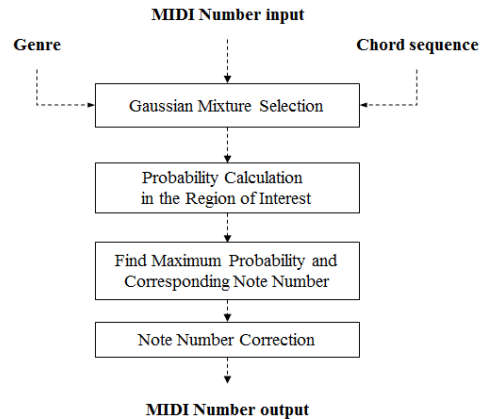


FIGURE 4 Avoid Note Correction

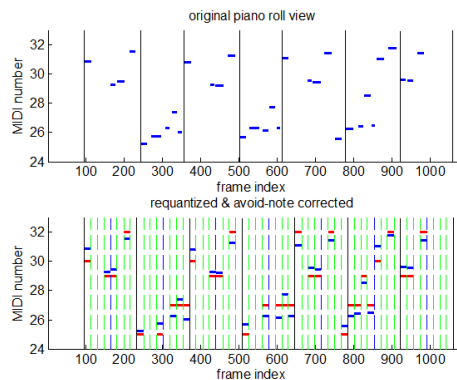


FIGURE 5 Quantization Result

(3) MIDI 템플릿을 이용한 음악 생성

추정된 Chord Sequence 을 이용하여 MIDI 템플릿을 조합하여 원하는 길이 만큼의 음악 생성을 한다. 원하는 길이에 따라 일정 횟수까지는 반복되는 횟수에 따라 스타일이 미리 설정되어 있으며 일정 횟수가 넘어가면 다시 반복되는 구조로 음악을 생성한다.

3. 실험 결과

허밍 입력의 악보 추정 성능 평가를 위해 남자 6 명, 여자 3명이 총 22곡의 동요를 노래한 DB 를 구축하였다. 성능 평가는 Onset 위치, pitch 로 추정한 MIDI Note Number 의 정확도로 평가하였으며 Onset 과 Note Number 는 DB를 직접 들으면

서 직접 Notation 한 것을 Reference 로 설정하였다. Onset 은 3명의 Notator 결과의 평균을 사용하였으며 50ms 허용 구간을 두어 추정 결과가 허용 구간안에 위치하면 맞게 인식한 것으로 판단하였다. Pitch 추정의 경우 Note Number가 일치할 경우 가중치 1, 반음(Note Number 1 차이) 차이가 날 경우 가중치를 0.5로 인식 성능 %를 구하였으며 Onset 을 제대로 추정한 Note에 대해 성능평가를 진행하였다. Onset 결과에서 Precision 은 추정된 Onset 개수 중에서 맞게 추정한 개수의 비율을 의미하며 Recall 은 실제 Onset 개수 중에서 맞게 추정한 개수의 비율을 의미한다. F-measure 는 $2 * Precision * Recall / (Precision + Recall)$ 로 정의되는 평균값을 의미한다. 본 논문에서는 91.8%의 F-measure 성능을 확보하였다.

Table 1 Onset Result

	Precision	Recall	F-measure
Onset	89.4%	94.4%	91.8%

Table 2 Pitch Result

	F-measure
Pitch	95.4%

생성된 음악에 대한 평가는 정량적 평가가 불가능하여 MOS 평가로 진행하였으며 10명의 실험자가 멜로디 및 반주가 얼마나 자연스럽게 음악적으로 잘 생성되었는지 평가하였다. MOS 점수는 5: excellent, 4: good, 3: fair, 2: poor, 1: bad 로 설정하여 평가하였으며 평균 4.0의 결과를 얻었다.

Table 3 Music Creation MOS Result

	평균
MOS	4.0

4. 결 론

본 논문에서는 음악적 지식이 없는 일반인도 허밍으로 쉽게 작곡을 할 수 있는 방법을 제시하였으며 기존 방법과 다르게 메트로놈 없이 허밍을 해도 음악적으로 완성도 높은 음악을 생성할 수 있는 방법을 제시하였다. 메트로놈 없는 자유 허밍에 대한 악보 추정을 위해 HMM 모델에서 가장 Chord Sequence 확률이 높은 마디 및 최소 음표를 추정하였으며 추정된 Chord Sequence를 이용하여

MIDI 템플릿을 조합하여 음악을 생성하였다. 악보 추정에 중요한 요소인 Onset 과 Pitch 는 각각 91.8%, 95.4% 인식 결과를 각각 얻었으며 생성된 음악에 대한 MOS 평가는 평균 4.0의 결과를 얻었다.

참 고 문 헌

- (1) Bello, J.P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M. and Sandler, M.B., 2005, A Tutorial on Onset Detection in Music Signals, *IEEE Transactions on Speech and Audio Processing* 13(5), pp. 1035~1047.
- (2) Schroeder, M. R., 1968, Period Histogram and Product Spectrum: New Methods for Fundamental-Frequency Measurement, *J. Acoust. Soc. Am.* Vol. 43.
- (3) Heo, H., Sung, D. Y. and Lee, K.G., 2013, Note Onset Detection based on Harmonic Cepstrum Regularity, *IEEE International Conference on Multimedia & Expo*
- (4) Vicar, M. M., Santos-Rodriguez, R., Ni, Y. and Bie, T. D., 2014, Automatic Chord Estimation from Audio: A Review of the State of the Art, *IEEE /ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No.2
- (5) Simon, I., Morris, D., and Basu, S., 2008, MySong: Automatic Accompaniment Generation for Vocal Melodies, *Proceedings of ACM CHI 2008*, pp. 725~724