확률적 은닉 성분 분석에 기반한 드럼 Onset 검출 방법

한병준 [1], 김연주 [2], 이장우 [3], 김민제 [4], 이교구 [3]

[1] 고려대학교 전기전자전파공학부
[2] 한국과학기술원(KAIST) 바이오및뇌공학과
[3] 서울대학교 융합과학기술대학원 디지털정보융합학과
[4] 한국전자통신연구원(ETRI)

hbj1147@korea.ac.kr, yunjooya@kaist.ac.kr, leejw41@snu.ac.kr,
mkim@etri.re.kr, kglee@snu.ac.kr

A Drum Onset Detection Scheme Based on Probabilistic Latent Component Analysis

Byeong-jun Han[1], Yunjoo Kim[2], Jangwoo Lee[3], Minje Kim[4], Kyogu Lee[3]

[1]School of Electrical Engineering, Korea University.
[2]Department of Bio and Brain Engineering, KAIST.
[3]Department of Digital Contents Convergence, Seoul National University.
[4]Electronics and Telecommunications Research Institute (ETRI).

요 약

특정 시간에 동시 연주된 다수 음원의 onset 을 검출하기 위해서는 음원 분리 문제가 선결되어야 한다. 특히, 드럼과 같은 조음(噪音) 악기 신호 검출 문제를 해결하기 위해서는 음원 분리 방법의 성능이 중요하다. 이에 본 연구에서는 효과적인 음원 분리 방법으로 알려진 확률적 은닉 성분 분석(PLCA) 방법에 기반한 주요 악기 신호의 onset 검출 방법을 제안한다. 효과적인 onset 검출을 위해, 첫째, 확률적 은닉 성분 분석으로 훈련 된 비음수 주파수 성분 중 최적의 성분을 선택하는 방법을 적용하고, 둘째, 드럼 악기 신호의 정확한 onset 검출을 위해 고안된 비음수 시계열 신호 threshold 방법을 적용한다. 실험에서는 제시된 방법을 이용하여 드럼의 주요 악기 신호 onset 검출 성능이 향상됨을 보인다.

1. 서론

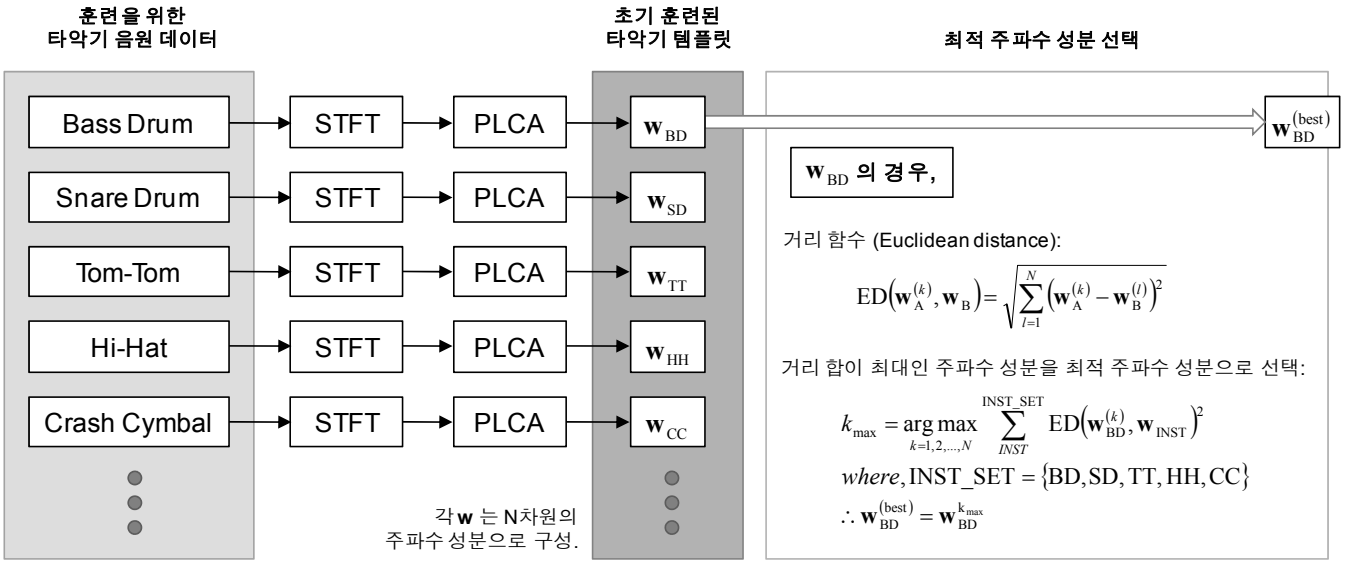
효율적인 음악 정보 처리를 위한 음악 정보 검색(MIR) 연구 분야에서 최근 음악의 리듬 특성을 주도하는 타악기의 인식 및 변환에 대한 연구가 각광받고 있다. Yoshii *et al.*[1]는 베이스 드럼, 스네어 드럼, 그리고 하이햇 등의 주파수 템플릿을 생성하여 onset 을 인식하는 시스템을 제안하였으며, Ravelli *et al.*[2]는 MIDI 로의 변환 없이 드럼 루프에서 리듬을 변환할 수 있는 방법을 제안하였다. 또한, Elias Pampalk *et al.*[3]는 청각 이미지에 기반한 드럼 신호 간 유사도 계산 모델을 제시하였다. Gillet *et al.*[4]는 음원 분리를 통하여 다음(多音) 환경에서 드럼 신호 기보 및 분리를 위한 방법을 제안하였다. 마지막으로, Yoo *et al.*[5]은 비음수 행렬 분해(NMF)를 변형한 비음수 행렬의 부분적 공동 분해(NMPCF)를 제안하여 원 음원으로부터 타악기 음원을 분리하고자 하였다.

한편, 최근 효과적인 음원 신호 분리 방법인 PLCA[6]가 제안되었다. 그러나 드럼 음원 분리 및 onset 검출, 기보 등을 위해서는 PLCA 가 활용된 예가 전무하다. 이에 본 연구에서는 PLCA 에 기반한 주요 드럼 악기 신호의 onset 검출 방법을 제안하고 그 성능을 평가한다. 효과적인 onset 검출을 위하여 PLCA 의 비음수 주파수 성분 중 최적의 성분을 선택하는 방법, 그리고 비음수 시계열 신호에 대한 threshold 방법을 제시한다. 이후, 실험에서는 제시된 방법으로 드럼의 주요 악기 신호 onset 검출 성능이 향상됨을 보인다.

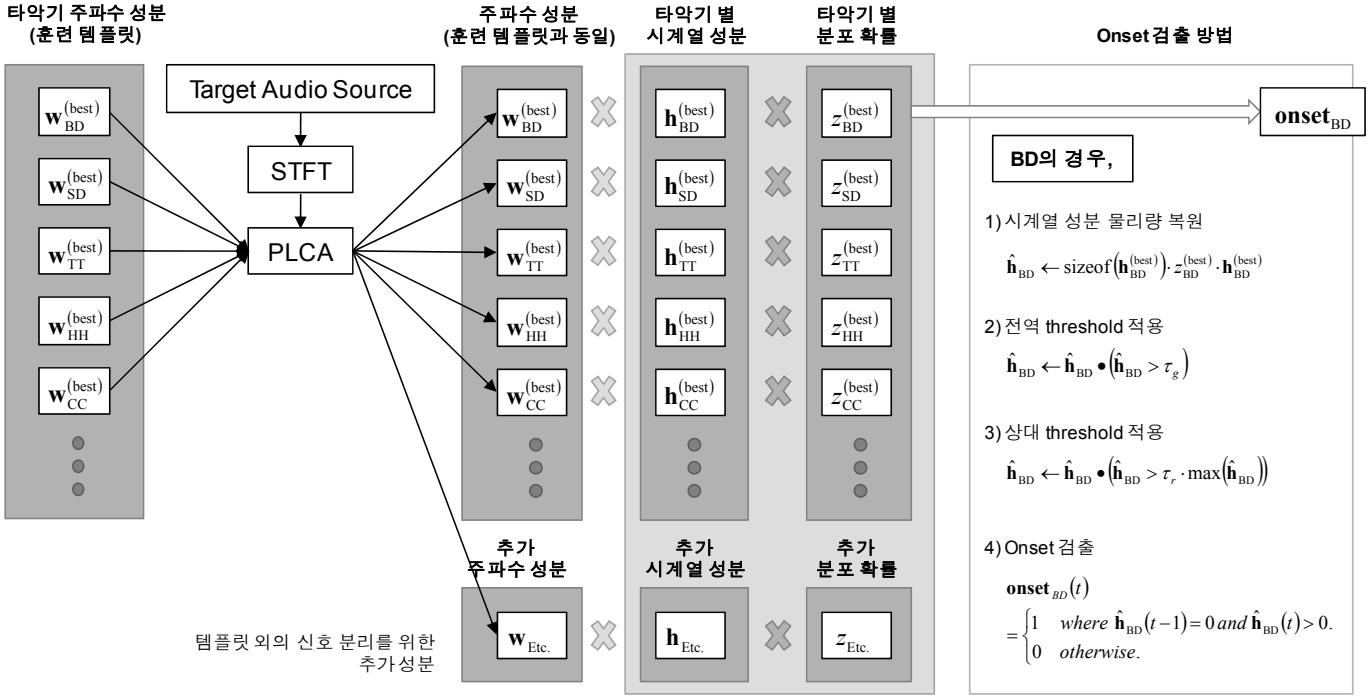
2. 확률적 은닉 성분 분석 (PLCA)

본 연구에 적용한 확률적 은닉 성분 분석(PLCA: Probabilistic Latent Component Analysis)은 확률적 은닉 의미론 연동(PLSI: Probabilistic Latent Semantic Indexing)의 확장 형태이다[5]. 확률적 은닉 성분 분석은 원 음원을 비음수 주파수 성분 및 비음수 시계열 성분, 그리고 이들의 분포 확률(또는 가중치)로 분리한다.

확률적 은닉 성분 분석을 위한 음원 모델은 다음과



(그림 1) 타악기 템플릿 생성 방법.



(그림 2) 드럼 onset 검출 방법.

정의된다[7]:

$$S_{f,t} = \sum_{z} \mathrm{P}(z)\mathrm{P}(f|z)\mathrm{P}(t|z) \qquad (1)$$

여기서, $S_{f,t}$는 주파수 f와 시간 t로 표현되는 스펙트로그램이며, z는 은닉 성분 구분자이다. 각 확률 분포 P(f|z) 및 P(t|z)는 각각 시간에 따라 발생하는 주파수 및 시계열 성분, 그리고 P(z)는 은닉 인자의 분포 확률이다.

스펙트로그램 $S_{f,t}$는 P(z), P(f|z) 및 P(t|z)를 분해함으로써 표현될 수 있다. 이 때 각 성분 및 확률을 계산하기 위해 EM (Expectation-Maximization) 알고리즘을 적용한다.

이때, E 알고리즘은 이하 수식 (2)와 같이 정의된다:

$$\mathrm{P}(z|f,t) \leftarrow \frac{\mathrm{P}(z)\mathrm{P}(f|z)\mathrm{P}(t|z)}{\sum_{z}\mathrm{P}(z)\mathrm{P}(f|z)\mathrm{P}(t|z)} \qquad (2)$$

또한, M 알고리즘은 다음 수식 (3~5)로 정의된다:

$$\mathrm{P}(f|z) \leftarrow \frac{\sum_{t} S_{f,t}\,\mathrm{P}(z|f,t)}{\sum_{f}\sum_{t} S_{f,t}\,\mathrm{P}(z|f,t)} \qquad (3)$$

$$\mathrm{P}(t|z) \leftarrow \frac{\sum_{f} S_{f,t}\,\mathrm{P}(z|f,t)}{\sum_{t}\sum_{f} S_{f,t}\,\mathrm{P}(z|f,t)} \qquad (4)$$

$$\mathrm{P}(z) \leftarrow \frac{\sum_{f}\sum_{t} S_{f,t}\,\mathrm{P}(z|f,t)}{\sum_{z}\sum_{f}\sum_{t} S_{f,t}\,\mathrm{P}(z|f,t)} \qquad (5)$$

3. 드럼 Onset 검출 방법

본 장에서 제안하는 드럼 onset 검출 방법은 크게 타악기 템플릿 생성(그림 1) 및 생성된 템플릿을 적용한 드럼 onset 검출 방법(그림 2)으로 나뉜다.

3.1. 타악기 템플릿 생성 방법

타악기 템플릿을 생성하는 방법은 다음과 같다. 우선 각 악기의 실제 샘플 음원을 수집한다. 이후, 각 악기 음원에 대해 STFT 를 적용하여 스펙트로그램을 구한다. 다음으로 PLCA 에 스펙트로그램을 적용하여 각 악기 음원으로부터 N 차원의 비음수 주파수 성분을 추출한다. 여기서, 추출된 주파수 성분들은 각 악기 음원을 대표할 가능성이 있는 성분들이다.

만일 PLCA 에 의해 추출된 N 차원의 비음수 주파수 성분을 각 타악기의 대표 성분으로 사용할 경우, 이들 성분들로부터 추출되는 비음수 시계열 성분은 정확하지 않을 수 있다. 왜냐하면 타악기는 음고(音高, pitch)를 정의할 수 없는 조음 악기(噪音樂器, unpitched instrument)이며 다른 악기와 차별화되는 주파수 영역 및 하모닉 패턴이 일정하지 않기 때문이다. 따라서 추출된 성분 중 다른 악기와 가장 차별화가 되는 성분을 구할 필요가 있다.

본 방법에서는 비음수 주파수 성분 간의 유클리디안 거리(Euclidean distance)를 다른 악기와의 차별화 정도 측정을 위한 거리 함수로 사용한다. 특정 비음수 주파수 성분과 다른 악기의 전체 비음수 주파수 성분들 간 정의된 거리 함수는 다음과 같다:

$$\mathrm{ED}\left(\mathbf{w}_{\mathrm{A}}^{(k)}, \mathbf{w}_{\mathrm{B}}\right) = \sqrt{\sum_{l=1}^{N}\left(\mathbf{w}_{\mathrm{A}}^{(k)} - \mathbf{w}_{\mathrm{B}}^{(l)}\right)^2} \tag{6}$$

여기서 $\mathbf{w}$ 는 비음수 주파수 성분이며, $\mathbf{w}_{\mathrm{A}}$, $\mathbf{w}_{\mathrm{B}}$ 등은 서로 다른 악기 A, B 로부터 추출된 성분, k 및 l 은 은닉 성분 구분자, 그리고 N 은 각 악기로부터 추출된 은닉 성분 수이다.

수식 (6)의 거리 함수를 사용하여 최적 주파수 성분을 선택하기 위해 특정 악기의 특정 비음수 주파수 성분과 다른 모든 악기의 비음수 주파수 성분의 거리를 측정하여 더한다:

$$k_{\max} = \arg\max_{k=1,2,\ldots,N} \sum_{INST}^{\mathrm{INST\_SET}} \mathrm{ED}\left(\mathbf{w}_{\mathrm{A}}^{(k)}, \mathbf{w}_{\mathrm{INST}}\right)^2 \tag{7}$$

여기서 $k_{\max}$ 는 가장 차별화가 된 주파수 성분의 은닉 성분 번호이며, INST_SET 은 악기의 인덱스 집합이다. 또한, A 는 INST_SET 의 원소이다.

최종적으로, 악기 A 의 최적의 비음수 주파수 성분 벡터 $\mathbf{w}_{\mathrm{A}}^{(\mathrm{best})}$는 $\mathbf{w}_{\mathrm{A}}^{k_{\max}}$가 된다.

3.2. 시계열 성분 Threshold 방법

그림 1 에서 구한 최적의 비음수 주파수 성분들은 각 악기의 비음수 시계열 성분을 계산하기 위한 재료가 된다. 그림 2 와 같이 최적 비음수 주파수 성분들을 PLCA 의 비음수 주파수 성분으로 고정시키고, 드럼 신호 검출 대상 음원에 대해 PLCA 를 적용한다. 그 결과로, 고정된 주파수 성분들에 대한 악기 별 시계열 성분, 악기 별 분포 확률, 그리고 추가 성분 차원에 대한 비음수 주파수/시계열 성분 및 분포 확률이 계산된다.

<표 1> 타악기 템플릿 생성 음원 샘플 데이터셋 정보

악기 종류	개수	총 길이(s)	f_s [Hz]	bits
베이스 드럼 (BD)	41	20.17	44,100	16
스네어 드럼 (SD)	74	49.66	44,100	16
탐탐 (TT)	58	158.16	44,100	16
하이 햇 (HH)	36	12.84	44,100	16
크래쉬 심벌 (CC)	40	306.02	44,100	16
총계	249	546.85		

비음수 시계열 성분을 사용하여 시간대 별 해당 비음수 주파수 성분의 존재량을 어느 정도 추정할 수 있다. 그러나 추출된 시계열 성분은 여전히 정확도가 낮으므로 각 악기의 onset 을 검출하기 위해서는 추가 작업이 필요하다. 이를 위하여 본 연구에서는 onset 을 검출하기 위한 threshold 방법을 제안한다.

시계열 성분의 경우 전체 합이 1 인 확률 분포값이므로, 다음 방법을 사용하여 실제의 물리량에 근접하는 값으로 복원할 필요가 있다.

$$\hat{\mathbf{h}} \leftarrow \mathrm{sizeof}\left(\mathbf{h}\right) \cdot z \cdot \mathbf{h} \tag{8}$$

여기서 $\hat{\mathbf{h}}$ 은 변동되는 시계열 성분 벡터이며, sizeof 는 벡터의 길이를 구하는 함수이다.

이후, 무음인 시계열 성분을 필터링하기 위하여 다음과 같이 전역 threshold τ_{g} 를 적용한다.

$$\hat{\mathbf{h}} \leftarrow \hat{\mathbf{h}} \bullet \left(\hat{\mathbf{h}} > \tau_g\right) \tag{9}$$

악기의 연주자에 및 악기의 특성에 따라 악기의 최대 강도가 다를 수 있다. 이를 고려하여, 다음과 같이 시계열 성분의 최대값에 대한 상대 threshold τ_{r} 을 적용한다.

$$\hat{\mathbf{h}} \leftarrow \hat{\mathbf{h}} \bullet \left(\hat{\mathbf{h}} > \tau_r \cdot \max\left(\hat{\mathbf{h}}\right)\right) \tag{10}$$

마지막으로, onset 은 $\hat{\mathbf{h}}$ 에 0 이 아닌 비음수 값이 국소 최초로 출현하는 지점으로 정의한다. 따라서 다음의 수식(11)을 이용하여 검출한다.

$$\mathbf{onset}(t) = \begin{cases} 1 & where\ \hat{\mathbf{h}}(t-1) = 0\ and\ \hat{\mathbf{h}}(t) > 0. \\ 0 & otherwise. \end{cases} \tag{11}$$

4. 실험 결과 및 분석

4.1. 실험 환경 및 데이터

전체 드럼 onset 검출의 구현은 MATLAB 환경에서 이루어졌다. 이 때 구현 및 실험에 사용된 시스템은 인텔 i7-930 쿼드 코어 2.8GHz CPU, 8GB 메모리, 1TB HDD 사양의 시스템이었다.

타악기 템플릿을 생성하기 위해 사용한 음원 데이터셋은 The Freesound Project [8]에서 확보하였다. 이 때 확보한 음원의 정보는 표 1 에 제시되어 있다. 흥미로운 것은, 크래쉬 심벌(CC) 및 탐탐(TT)의 음원 길이가 베이스 드럼(BD), 스네어 드럼(SD), 하이햇(HH) 보다 상대적으로 더 긴 점이다. 이유는, 심벌은 드럼에 비해 지속 시간(decaying time)이 길며, 탐탐은 하이탐(HT), 미들 탐(MT), 로우 탐(LT), 플로어 탐(FT)과

같이 다양한 주파수 분포를 가지는 악기로 구성되기 때문이다.

타악기 템플릿 생성 시 은닉 성분 수 N 은 5 로 설정하였다. 또한, 드럼 onset 검출을 위한 전역 및 상대 threshold τ_g 및 τ_r으로 각각 15 및 50%를 설정하였다. 마지막으로, 제안된 드럼 onset 검출 방법의 평가를 위해 작곡가로부터 제공받은 한국 가요의 실제 드럼 음원 5 종에 대해 실험을 진행하였다.

4.2. 실험 결과

제안된 방법의 효과를 검증하기 위한 실험을 위하여 드럼 음원으로부터 BD 및 SD 의 onset 을 추출하였다.

표 2 및 표 3 은 제안된 최적 성분 검색 방법 사용 유무 및 제안된 threshold 방법 사용 유무에 따른 베이스 드럼 및 스네어 드럼의 onset 검출 방법 효과를 검증하기 위한 실험 결과이다.

실험 결과, 최적 성분 검색 방법과 threshold 를 사용한 onset 검출 방법을 모두 함께 사용한 경우 베이스 드럼 및 스네어 드럼의 onset 검출 정확도가 각각 95.98%, 95.48%으로, 양쪽 방법 모두를 사용하지 않은 경우의 정확도(89.04%, 80.10%)를 상회하는 것으로 나타났다.

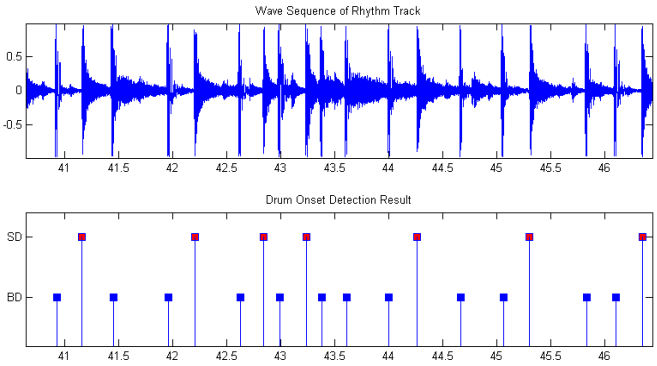
그림 3 은 음원 #1 에서 베이스 드럼 및 스네어 드럼의 onset 을 검출하는 예시로, BD 및 SD 가 올바르게 검출되는 것을 확인할 수 있다.

<표 2> 제안 방법의 사용 유무 비교 결과 (BD)

최적성분 검색방법	Threshold 검출방법	#1	#2	#3	#4	#5	총계
미사용	미사용	402	86	125	561	199	1373
미사용	사용	416	86	125	599	203	1429
사용	미사용	429	86	127	611	210	1463
사용	사용	432	86	127	623	212	1480
Ground-truth		436	86	129	677	214	1542

<표 3> 제안 방법의 사용 유무 비교 결과 (SD)

최적성분 검색방법	Threshold 검출방법	#1	#2	#3	#4	#5	총계
미사용	미사용	133	51	71	239	162	656
미사용	사용	156	46	74	245	179	700
사용	미사용	165	57	77	261	189	749
사용	사용	174	60	77	270	201	782
Ground-truth		180	62	87	281	209	819



(그림 3) 드럼 onset 검출 예시.

5. 결론

본 연구에서는 효과적인 음원 분리 방법으로 알려진 확률적 은닉 성분 분석(PLCA) 방법에 기반한 주요 악기 신호의 onset 검출 방법을 제안하였다. 효과적인 onset 검출을 위해, 첫째, 확률적 은닉 성분 분석으로 훈련 된 비음수 주파수 성분 중 최적의 성분을 선택하는 방법을 적용하고, 둘째, 드럼 악기 신호의 정확한 onset 검출을 위해 고안된 비음수 시계열 신호 threshold 방법을 적용하였다. 실험에서는 제시된 방법을 이용하여 드럼의 주요 악기 신호 onset 검출 성능이 향상되었음을 보였다.

향후 연구는 베이스 드럼, 스네어 드럼 뿐만 아니라 서스테인 검출, 악기의 특성화, 음원 분리가 어려운 심벌 등의 타악기 전반 onset 검출로 확장해야 할 것이다.

Acknowledgements

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2010 년도 문화콘텐츠산업기술지원사업의 연구결과로 수행되었음.

참고문헌

[1] Kazuyoshi Yoshii, Masataka Goto, Hiroshi G. Okuno, “Drum sound recognition for polyphonic audio signals by adaptation and matching of spectrogram templates with harmonic structure suppression,” IEEE Trans. on ASLP, vol.15, no.1, pp.333-345, Jan. 2007.

[2] Emmanuel Ravelli, Juan P. Bello, and Mark Sandler, “Automatic rhythm modification of drum loops,” IEEE SPL, vol.14, no.4, pp.228-231, Apr. 2007.

[3] Elias Pampalk, Perfecto Herrera, and Masataka Goto, “Computational models of similarity for drum samples,” IEEE Trans. On ASLP, vol.16, no.2, pp.408-423, Feb. 2008.

[4] Olivier Gillet and Gaël Richard, “Transcription and separation of drum signals from polyphonic music,” IEEE Trans. On ASLP, vol.16, no.3, pp.529-540, Mar. 2008.

[5] Jiho Yoo, Minje Kim, Kyeongok Kang and Seungjin Choi, “onnegative matrix partial co-factorization for drum source separation,” Proc. of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010), pp.1942-1945, Mar. 2010.

[6] Paris Smaragdis, Bhiksha Raj, and Madhusudana Shashanka, “A probabilistic latent variable model for acoustic modeling,” Advances in models for acoustic processing workshop, Neural Information Processing Systems (NIPS), Dec. 2006.

[7] Juhan Nam, Gautham J. Mysore, Joachim Ganseman, Kyogu Lee, Jonathan S. Abel, “A super-resolution spectrogram using coupled PLCA,” INTERSPEECH 2010, Sept. 2010.

[8] The Freesound Project, http://www.freesound.org/ .