INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY
AND
HONDA RESEARCH INSTITUTE

Translating motion sequences to Vocab Sequences using 3D Convolutional Networks

by

Pranav Sankhe

Application Assignment: Group No. 27 to

Professor Vikram Gadre
Department of Electrical Engineering

July 2018

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY
AND
HONDA RESEARCH INSTITUTE

Abstract

Professor Vikram Gadre
Department of Electrical Engineering

by Pranav Sankhe

Contents

Abbreviations

ICA	**I**ndependent **C**omponent **A**nanlysis
STFT	**S**hort **T**ime **F**ourier **T**ransform
PICA	**P**robabilistic **I**ndependent **C**omponent **A**nanlysis

Chapter 1

Introduction

It has been shown that oscillatory neural activity is sensitive to accented tones in a rhythmic sequence rhythmic sequence. Our entire analysis is based on one assumption that the neural activity associated with music perception is synchronized with the actual recordings. For example, if there's a onset in the recording, there would be an onset in he perceived music as well. We wish to conduct this analysis on the public domain *OpenMIIR dataset* of EEG recordings taken during music perception and imagination. The biggest challenge involved in this task is the extremely noisy nature of the EEG signal. We would be applying several de-noising techniques like , **ICA** (Blind Source Seperation), SSP based filtering, Wavelet transform denoising,Wavelet Packet Transform, Non-Linear Adaptive filtering,etc. This raises the question whether Music Information Retrieval techniques originally developed to detect beats and extract the tempo from music recordings could also be used for the analysis of corresponding EEG signals. One could argue that as the brain processes the perceived music, it generates a transformed representation which is captured by the EEG electrodes. Hence, the recorded EEG signal could in principle be seen as a mid-level representation of the original music piece that has been heavily distorted by two consecutive black-box filtersthe brain and the EEG equipment.

1.1 STFT

When we compute the fourier transform of a signal x(t), it is multiplied with an exponential term, at some certain frequency "f" , and then integrated over **all times**.

The information provided by the integral, corresponds to all time instances, since the integration is from minus infinity to plus infinity over time. It follows that no matter

where in time the component with frequency "f" appears, it will affect the result of the integration equally as well. In other words, whether the frequency component "f" appears at time t1 or t2, it will have the same effect on the integration. This is why Fourier transform is not suitable if the signal has time varying frequency, i.e., the signal is non-stationary. If only, the signal has the frequency component "f" at all times (for all "f" values), then the result obtained by the Fourier transform makes sense.

To introduce a sense of time in our signal analysis, we assume that some portion of a non-stationary signal is stationary and evaluate the fourier transform of the signal in that window. Dennis Gabor (1946) Used STFT to analyze only a small section of the signal at a time – a technique called Windowing the Signal. The segment of signal is assumed *stationary*. **STFT** is a 3D transform. It is a function of time and frequency.

$$STFT_x(t^{'}, f) = \int x(t) * \omega(t - t^{'}) * exp(-j2\pi ft)dt$$

1.2 Dataset

In this study, we use a subset of the OpenMIIR dataset a public domain dataset of EEG recordings taken during music perception and imagination.These stimuli were selected from well-known pieces of different genres. They span several musical dimensions such as meter, tempo, instrumentation (ranging from piano to orchestra) and the presence of lyrics (singing or no singing present). All stimuli were normalized in volume and kept similar in length, while ensuring that they all contained complete musical phrases starting from the beginning of the piece. The EEG recording sessions consisted of five trials $t \epsilon T := \{1, ..., 5\}$ in which all stimuli $s \epsilon S := \{01, 02, 03, 04, 11, 12, 13, 14, 21, 22, 23, 24\}$ were presented in randomize order. This results in total of 300 trials for 5 (example) participants, 60 trials per particpant 25 trials per stimulus. EEG was recorded with a BioSemi Active-Two system using 64+2 EEG channels at 512 Hz. Horizontal and vertical electrooculography (EOG) channels were used to record eye movements.

1.3 Data Preprocessing

EEG pre-processing comprised the removal and interpolation of bad channels as well as the reduction of artifacts using techniques like Independent component analysis to remove ocular artifacts, etc. The following pre-processing techniques have been tried out:

- There are in total 69 channels out of which 64 are EEG channels and the remaining 5 channels are **EOG** channels which record the ocular activity. The ocular activity introduces noisy artifacts in the EEG signal. We cross-correlated the 5 **EOG** channels with each 64 EEG channels and filtered out the bad EEG channels which displayed a high correlation value.

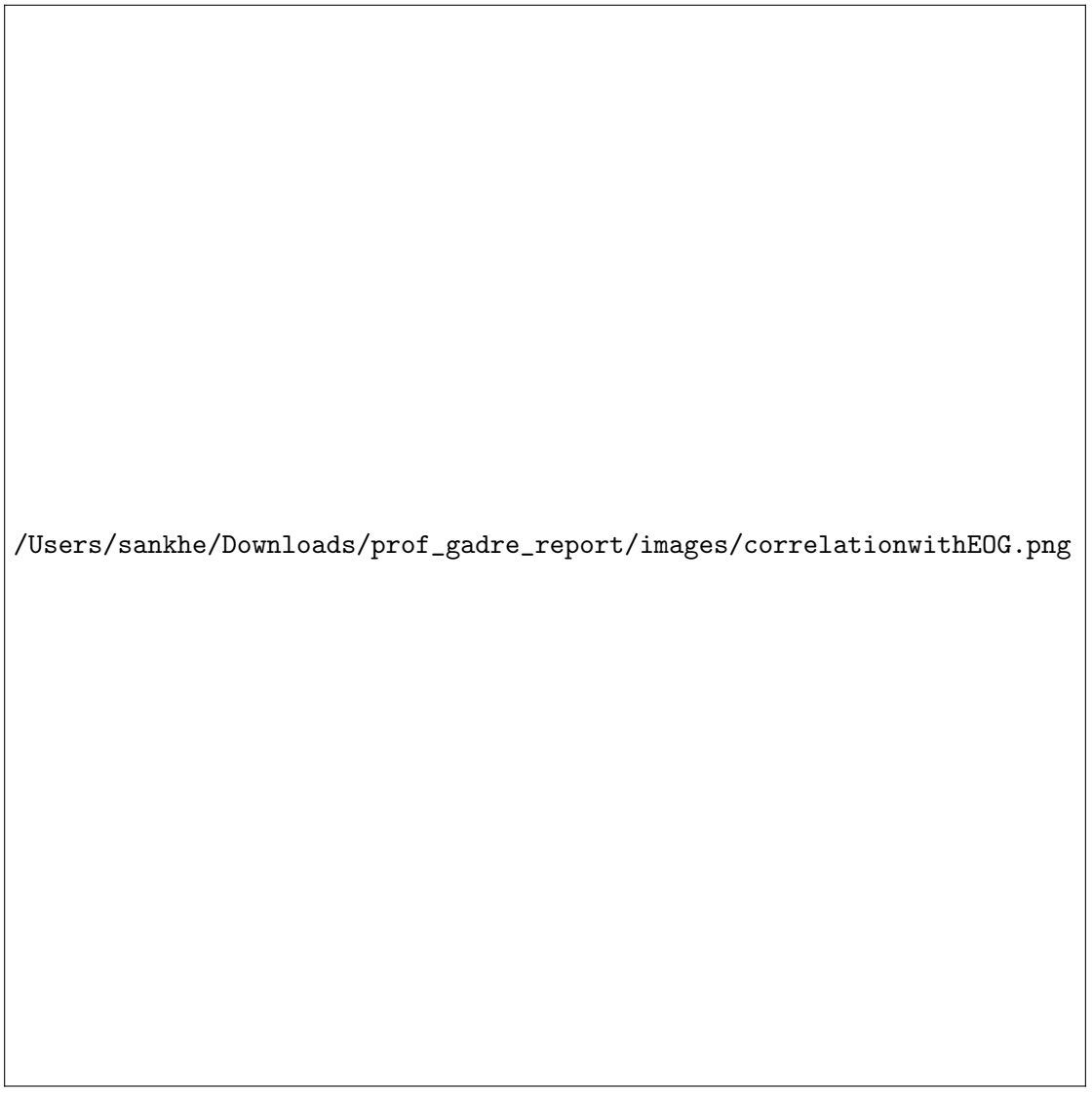

FIGURE 1.1: Correlation of EEG channels with **EOG** channels.

ICA

We could view a measurement as an estimate of a single source corrupted by some random fluctuations (e.g. additive white noise). Instead, we assert that a measurement can be a combination of many distinct sources each different from random noise. The broad topic of separating mixed sources has a name -blind source separation (BSS). As of todays writing, solving an arbitrary BSS problem is often intractable. However, a small subset of these types of problem have been solved only as recently as the last two decades this is the provenance of independent

component analysis (ICA).

Solving blind source separation using ICA has two related interpretations filtering and dimensional reduction. If each source can be identified, a practitioner might choose to selectively delete or retain a single source (e.g. a persons voice, above). This is a filtering operation in the sense that some aspect of the data is selectively removed or retained. A filtering operation is equivalent to projecting out some aspect (or dimension) of the data in other words a prescription for dimensional reduction. Filtering data based on ICA has found many applications including the analysis of photographic images, medical signals (e.g. EEG, MEG, MRI, etc.), biological assays (e.g. micro-arrays, gene chips, etc.) and most notably audio signal processing.

Quick Summary of ICA:

- Subtract off the mean of the data in each dimension.
- Whiten the data by calculating the eigenvectors of the covariance of the data.
- Identify final rotation matrix that optimizes statistical independence

- ICA has been extensively used as an effective filtering technique for EEG channels. We applied ICA and we got some interesting results which we want to analyse further. The following plots are STFT of the signals. Since music information is embedded in the frequency content, we applied STFT. We are reading up on PICA, a variant of ICA and planning to implement it on the raw data.
- We calculated the moving average by specifying a window-size and subtracted it from the EEG signal. When we take average using window, we allow some impulses in the signal due to which renders the STFT of the signal meaningless. Also you can see repeating patterns in the STFT plot which is due to the sampling at different harmonics of a fundamental frequency decided by the window size.

PICA

One of the drawbacks of ICA is that it does not come (like PCA does) with a well-defined method to select the most important components. In the original formulation, the number of independent components is equal to the dimension of the variables, so that the decomposition is achieved without dimensional reduction. This leads to computational and overfitting issues when dealing with high-dimensional data and small sample sizes, and a lack of interpretability of the obtained results.

(a) Without ICA (b) With ICA

FIGURE 1.2: Sampled at 512 Hz

(a) Without ICA (b) With ICA

FIGURE 1.3: Sampled at 256 Hz

Probabilistic ICA (alternatively called noisy ICA, or independent factor analysis, although we will reserve the latter term to a more specific method in which factors are Gaussian mixtures) assumes a small number of indepen- dent components, with a residual term which is modeled as Gaussian noise. The explicit model is therefore given by

$$x_i = As_i + \mu + \eta_i$$

(a) Window Size = 5 (b) Window Size = 10

FIGURE 1.4: Sampled at 512 Hz without applying ICA

Here,

- x_i denotes the p-dimensional column vector of individual measurements at voxel location i
- s_i denotes the q-dimensional column vector of non-Gaussian source signals contained in the data
- η_i denotes Gaussian noise $\eta_i N(0, \sigma^2 \Sigma_i)$

We assume that $q < p$, i.e. that there are fewer source processes than observations in time.

Chapter 2

Wavelet Transform

2.1 Introduction to Wavelet Transform

What is a wavelet? - *A small wave, a wave like oscillation*

Wavelet Transforms

- Converts a signal into a series of wavelets
- Provides a way for analyzing waveforms, bounded in both frequency and duration
- Allow signals to be stored more efficiently than by Fourier transform
- Able to better approximate real-world signals

To make a real long story short, we pass the time-domain signal from various highpass and low pass filters, which filters out either high frequency or low frequency portions of the signal. This procedure is repeated, every time some portion of the signal corresponding to some frequencies being removed from the signal.

Stationary Signal Signals with frequency content unchanged in time.All frequency components exist at all times.

Non-stationary Signal

Frequency changes in time. One example: the Chirp Signal

Disadvantages of Fourier Transform

- FT Only Gives what frequency components *exist* in the signal
- The Time and frequency information can not be seen at the **same time**
- Time-frequency representation of the signal is needed.

Most of real-world Signals are Non-stationary. We need to know **whether** and also **when** an incident was happened.

One earlier solution in the quest of searching the answer for *when an incident happened*, was the **STFT** which we analyzed earlier.

The drawbacks of STFT

- Unchanged Window
- Dilemma of Resolution
  - Narrow window → poor frequency resolution
  - Wide window → poor time resolution
- *Heisenberg Uncertainty Principle*: Cannot know what frequency exists at what time intervals

In FT, the kernel function, allows us to obtain perfect frequency resolution, because the kernel itself is a window of infinite length. In STFT is window is of finite length, and we no longer have perfect frequency resolution. You may ask, why don't we make the length of the window in the STFT infinite, just like as it is in the FT, to get perfect frequency resolution? Well, than you loose all the time information, you basically end up with the FT instead of STFT. To make a long story real short, we are faced with the following dilemma: If we use a window of infinite length, we get the FT, which gives perfect frequency resolution, but no time information. Furthermore, in order to obtain the stationarity, we have to have a short enough window, in which the signal is stationary.
The narrower we make the window, the better the time resolution, and better the assumption of stationarity, but poorer the frequency resolution:

Wavelet Transform is an alternative approach to the short time Fourier transform invented to overcome the resolution problem It is similar to STFT in the sense that the signal is multiplied with a function. Using wavelet transfrom, we can analyze the signal at different frequencies with different resolutions.
High frequencies → Good time resolution and poor frequency resolution
Low frequencies → Good frequency resolution and poor time resolution
Hence wavelet transform is more suitable for short duration of higher frequency; and longer duration of lower frequency components.

2.2 Definition of Wavelet Transform

The wavelet transform splits up the signal into a bunch of signals representing the same signal, but all corresponding to different frequency bands; only providing what *frequency bands* exists at what *time intervals*.

Mathematically, the following is the equation of the wavelet transform:

$$CWT_x^{\psi}(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t) * \psi^*(\frac{t-\tau}{s})dt$$

where,

τ represents translation, the location of the window

s represents the scale

The term $\psi^*(\frac{t-\tau}{s})$ is the mother wavelet: A prototype for generating the other window functions. All the used windows are its dilated or compressed and shifted versions.

Some insights on scale (s)

- $s > 1$: dilate the signal
- $s < 1$: compress the signal
- Low frequency $\rightarrow$ high $s \rightarrow$ Non-detailed global view of signal $\rightarrow$ span entire signal
- High frequency $\rightarrow$ low $s \rightarrow$ detailed view last in short time
- Only limited interval of scales is necessary

2.3 Computation of Wavelet Transform

- Step 1: The wavelet is placed at the beginning of the signal, and set s=1 (the most compressed wavelet);
- Step 2: The wavelet function at scale 1 is multiplied by the signal, and integrated over all times; then multiplied by ;
- Step 3: Shift the wavelet to $t = \tau$, and get the transform value at $t = \tau$ and s=1;
- Step 4: Repeat the procedure until the wavelet reaches the end of the signal;
- Step 5: Scale s is increased by a sufficiently small value, the above procedure is repeated for all s;
- Step 6: Each computation for a given s fills the single row of the time-scale plane;
- Step 7: CWT is obtained if all s are calculated.

We plotted the wavelet transforms using the `python packages` like `pywt` and `numpy`. The y-axis represents the scales which map to frequency.

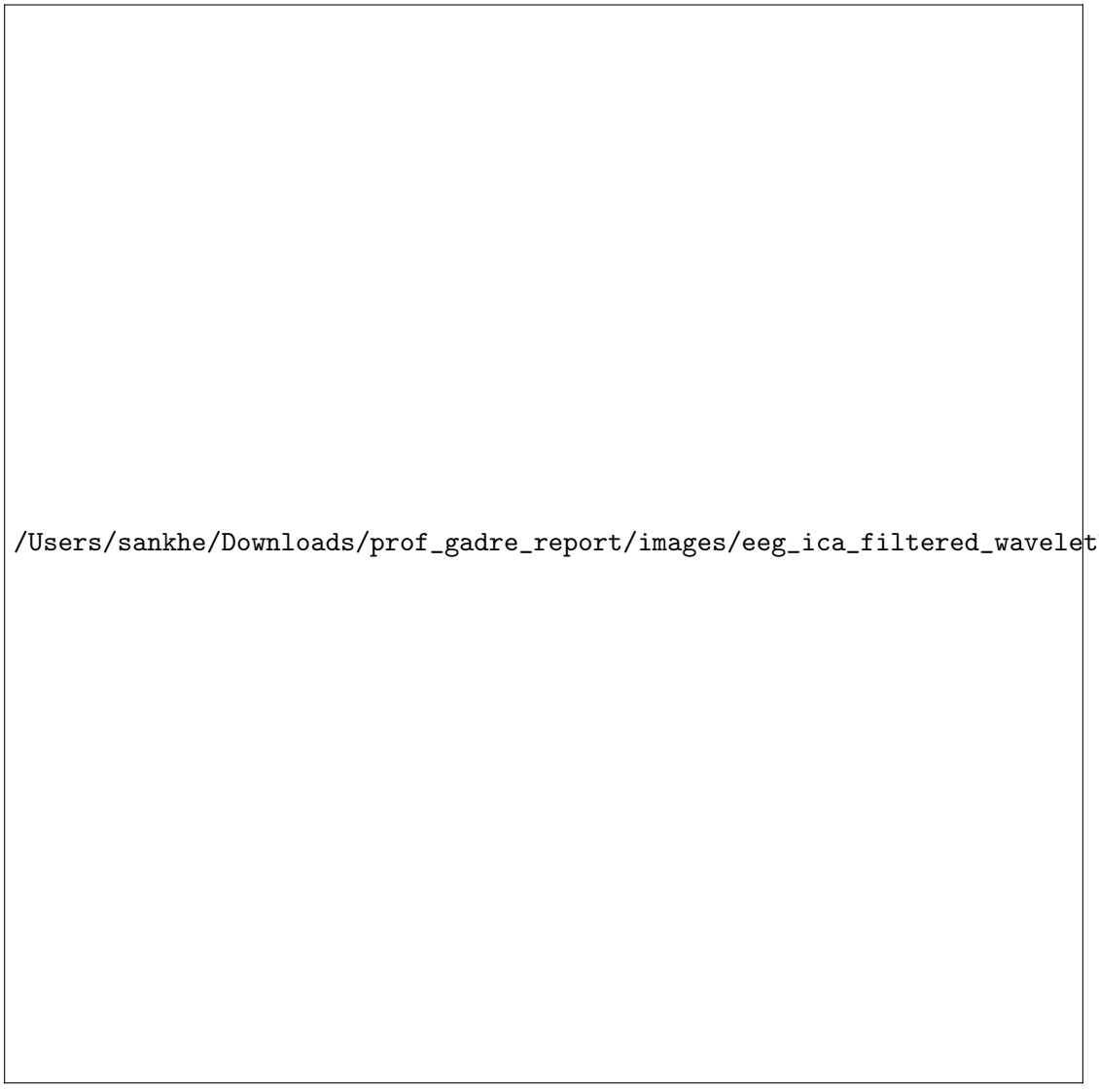

FIGURE 2.1: EEG data

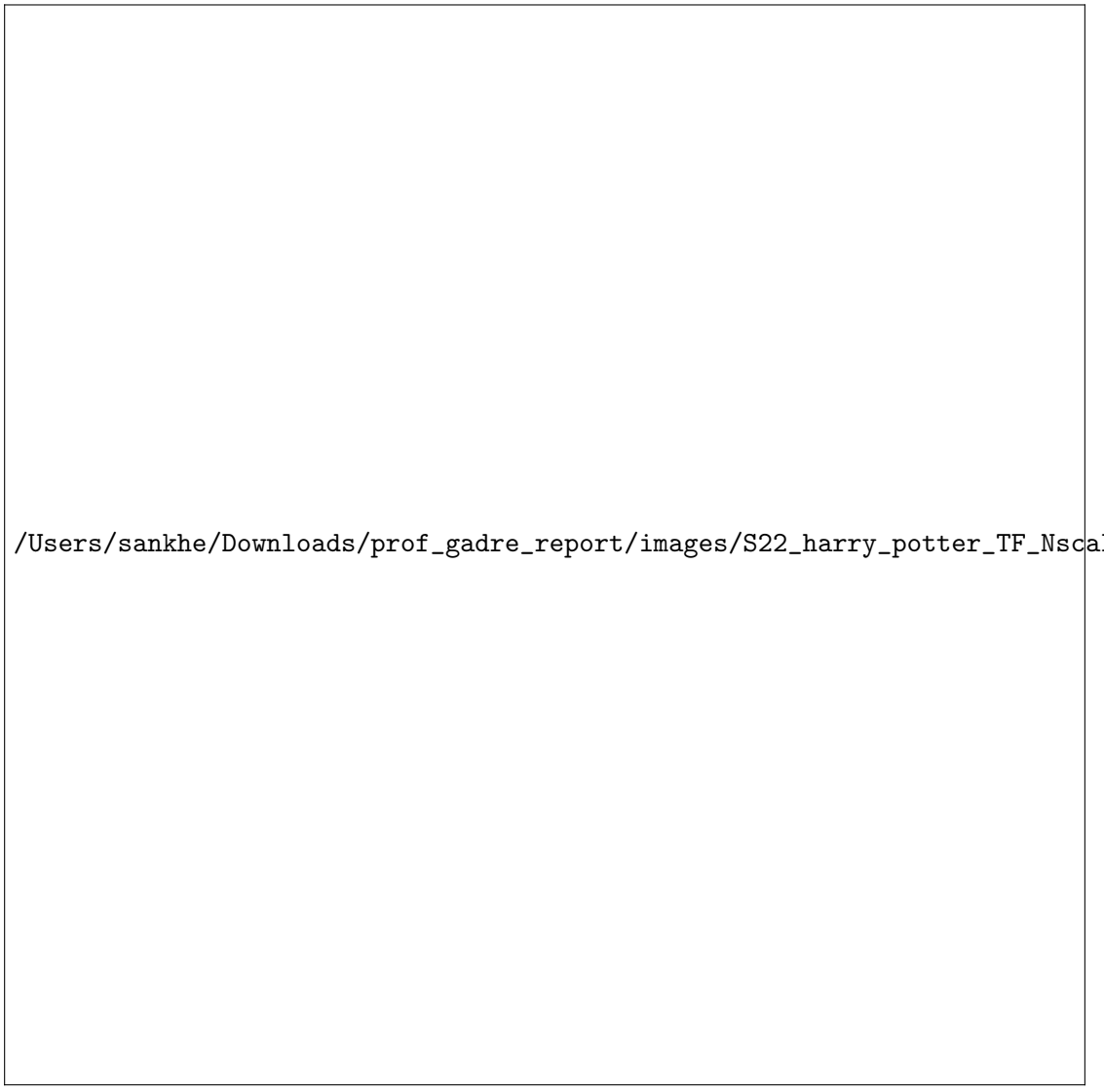

FIGURE 2.2: Harry potter theme

Bibliography

[1] *Introduction to Wavelets.* Department of Electrical Engineering, IIT Bombay

[2] Robi Polikar, *The Wavelet Tutorial, http://users.rowan.edu/ polikar/WAVELET-S/WTpart1.html*

[3] Christian F. Beckmann and Stephen M. Smith *Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging* Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital, Headley Way, Headington, Oxford, UK

[4] Sebastian Stober, Avital Sternin, Adrian M. Owen and Jessica A. Grahn *Towards Music Imagery Information Retrieval: Introducing the OpenMIIR Dataset of EEG Recordings from Music Perception and Imagination.* In: Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR15), pages 763-769, 2015.