inter.noise
HAMBURG 2016

On urban soundscape mapping: A computer can predict the outcome of soundscape assessments

Peter LUNDÉN[1]; Östen AXELSSON[2]; Malin HURTIG[3]

[1] SP Technical Research Institute of Sweden, Sustainable Built Environment unit, Sweden

[2] Stockholm University, Sweden

[3] Stockholm University, Sweden

ABSTRACT

The purpose of this study was to investigate whether or not a computer may predict the outcome of soundscape assessments, based on acoustic data only. It may be argued that this is impossible, because a computer lack life experience. Moreover, if the computer was able to make an accurate prediction, we also wanted to know what information it needed to make this prediction. We recruited 33 students (18 female; M_{age} = 25.4 yrs., SD_{age} = 3.6) out of which 30 assessed how pleasant and eventful 102 unique soundscape excerpts (30 s) from Stockholm were. Based on the Bag of Frames approach, a Support Vector Regression learning algorithm was used to identify relationships between various acoustic features of the acoustics signals and perceived affective quality. We found that the Mel-Frequency Cepstral Coefficients provided strong predictions for both Pleasantness (R^2 = 0.74) and Eventfulness (R^2 = 0.83). This model performed better than the average individual in the experiment in terms of internal consistency of individual assessments. Taken together, the results show that a computer can predict the outcome of soundscape assessments, which is promising for future soundscape mapping.

Keywords: Soundscape mapping, Machine learning, Urban planning I-INCE Classification of Subjects Number(s): 56.3, 52.9

1. INTRODUCTION

Currently, management of the urban acoustic environment is limited to monitoring and controlling the sound levels from transport and industry, and thus to noise abatement. Sound level is a one-dimensional measure that provides us with little information on how people perceive or experience the acoustic environment. The current approach to the management of the urban acoustic environment also miss a lot of important auditory aspects, such as the pleasant sounds of nature and the vibrant sounds of people. The soundscape approach provides an alternative that takes all of these additional aspects in to account [1]. A current limitation with the soundscape approach is that it requires socio-acoustic surveys that are resource intensive and time consuming. To overcome this limitation and to be able to map soundscape at a large scale we would need a prediction model of the relationship between physical data and human responses to the acoustic environment [2]. One method for developing such a model is machine learning. This paper presents the results of such an effort, based on similar technology as used in Music Information Retrieval (MIR) [3].

A handful of studies have investigated how soundscape can be modeled with the aid of machine learning algorithms. One approach is to use neural networks. Oldoni and co-workers [4] simulated human auditory processing with the goal of modeling how listeners, over time, would switch their attention between different sounds. This model could be used for predicting how additional sounds (e.g., chirping birds or flowing water) attract attention from undesired sound (e.g., road traffic).

Yu and Kang [5] created neural network models for predicting 'Acoustic Comfort' of soundscape (i.e., how pleasant or unpleasant they are) for 19 sites in Europe and China, based on data from large scale field surveys, including physical, social, demographic, and psychological data. However, they were unable to create one global model for all 19 sites, and created individual models for every site,

1 peter.lunden@sp.se
2 oan@psychology.su.se
3 malin.hurtig@psychology.su.se

using a different set of input features for each model. The best performing model achieved a correlation coefficient of 0.79 between the predicted and empirical values.

Another approach is the Bag of Frames (BOF) often used in MIR [3]. A digital acoustic signal is divided into a large set of short time frames, with a 50% overlap between consecutive frames. A typical length of a frame is 50 ms. Aucouturier and co-workers [6] claimed that the BOF approach is nearly optimal for classifying soundscapes. Based on Mel-Frequency Cepstral Coefficients (MFCC), they classified 16 recordings of acoustic environments from Paris into 4 classes of "avenue," "street," "market," and "park" (cf. [7]), reaching a 96% accuracy. Lagrange and co-workers [7] replicated the study of Aucouturier et al., using four sets consisting of between 16 and 100 recordings from Paris and London, including the recordings that Aucouturier et al. used. They replicate the results of Aucouturier et al., achieving an accuracy of 97%. However, for the other three sets of recordings, the accuracy was below 50%. They concluded that the BOF approach is insufficient for modeling urban soundscapes.

Torija and co-workers [8] trained a support vector machine (SVM) to classify 471 recordings from Granada, Spain, into ten soundscape categories, derived from field surveys. The input data was 14 acoustic descriptors. SVMs trained with Sequential Minimal Optimization, achieved 91% accuracy.

The three studies cited in the last two paragraphs used soundscape classification. However, for soundscape mapping the critical question is, in what ways are soundscapes similar to or different from one another? According to Yu and Kang [5], the answer is Acoustic Comfort.

Fan and co-workers [9] used the BOF approach, but stepwise multiple linear regression instead of classification. To do this, there must be a model that identifies the underlying main dimensions of soundscape, which explains the similarities and differences among them. Fan et al. used the now well established model of perceived affective quality, identifying Pleasantness and Eventfulness at the two underlying main dimensions [10]. In total, 20 participants assessed 125 sound clips (6 s) from the Sound Ideas corpus [11] and the World Soundscape Project [12]. Fan et al. concluded that the results of the machine learning was well in agreement with the participants' responses.

In the present study, the BOF approach was used. A Support Vector Regression (SVR) learning algorithm was trained to identify relationships between features of the acoustic signals and perceived affective quality of 93 excerpts (30 s) from 77 audio recordings. In total, 30 students sorted the experimental sounds in a 2D space of Pleasantness and Eventfulness with the aid of a touch screen. The purpose of the study was to learn whether or not the system would be able to predict the average human responses, based on features of the acoustic signals, and if so, what feature would provide the best result.

2. METHOD

2.1 Participants

In total, 33 students (18 female, 15 male) were recruited to take part in the listening experiment. They were 20–36 years old (M_{age} = 25.4 yrs., SD_{age} = 3.6). All participants included in the analyses had a hearing threshold below 25 dB in their best ear at all tested frequencies (0.125, 0.5, 1, 2, 3, 4, and 6 kHz; Hughson Westlake's method, Interacoustics Diagnostic Audiometer AD226). Two participants did not meet the criteria of normal hearing and were excluded from the analyses. One participant was excluded after data screening, due to low internal consistency in the assessments. All participants received course credits or a gift card of 200 SEK as compensation for volunteering.

2.2 Experimental Sounds

A large number of acoustic environments in the greater Stockholm region were documented by the aid of audio recordings using a higher order Ambisonics format. This format preserves the 3D spatial acoustic information, and provides for very realistic reproduction of the recorded acoustic environments [13, 14]. In total, 77 Ambisonics recordings (15 min) were conducted in the summer period of 2014. The recording locations were selected with the aim to identify all combinations of Pleasantness and Eventfulness [10], and thus to fill the 2D space of perceived affective quality.

From the 77 recordings, 102 excerpts (30 s) were selected to be used in a listening experiment. Previous research results indicate that Pleasantness is positively associated with the sound of nature (e.g., chirping birds) and negatively associated with the sound of technology (e.g., road traffic), and that Eventfulness is positively associated with the sound of people [10, 15]. Therefore, the 102 audio excerpts were selected to include all practically possible combinations of a low, medium and high presence of the sound of people, nature and technology.

2.3 Equipment

2.3.1 Recording

The audio recordings were conducted with the use of two microphone units: one free-field measurement microphone (Brüel & Kjær 4190/2669; Brüel & Kjær 5935 Conditioner Amplifier), and one spherical microphone array (Eigenmike EM32) with 32 pressure sensitive electret microphones. The measurement microphone was connected to a RME Fireface UC audio interface, whereas the microphone array used its own interface. Both interfaces were then connected to a Mac Book Pro computer, and synchronized at sample level using word clock. The sampling rate was 48 kHz (32-bit floating point). The signal from the microphone array was processed by a higher order near field compensated (HOA-NFC) 4th order Ambisonics encoder [16], developed in the project. All audio signals were then recorded with the use of the Reaper digital audio workstation software, running on the Mac Book Pro computer. In order to estimate the sound levels at the recording locations, so that the audio recordings could be reproduced as authentically as possible, the measurement microphone was used to record a 1 min long calibration tone (1 kHz, 94 dB) at the beginning of each recording session (Brüel & Kjær Type 4231 sound calibrator).

2.3.2 Playback

In the listening experiment, the 102 audio excerpts were presented to the participants using a loudspeaker array consisting of 29 active digital loudspeakers (Genelec 8130A), and two active digital sub-basses (Genelec 7270). The 29 loudspeakers were mounted in a spherical configuration with four horizontal circles of loudspeakers, which contained 6, 8, 8, and 6 loudspeakers each, plus 1 loudspeaker at the top of the sphere. A HOA-NFC 4th order Ambisonics decoder for 30 loudspeakers was designed, using software developed in the project. The Sphere Partition Toolbox [17] was used to obtain an optimal solution for the loudspeakers' positions on the sphere. In this solution, all 30 loudspeakers covered an equal part of the sphere. The 30th, bottom loudspeaker, was removed, because there was no space available for it. This has a negligible influence on the quality of the reproduction. The acoustic axes of the loudspeaker were all directed towards the center of the sphere, with a radius of 1.5 m. The sounds where played back with the use the Reaper digital audio workstation software, running on a Macintosh Pro computer with a RME MADIface XT. The loudspeakers were fed with digital audio signals, distributed to the loudspeakers through a RME ADI-6432R MADI to AES/EBU converter.

The frequency responses of all loudspeakers were equalized at the center of the sphere. Also the sound levels of the 102 audio excerpts used in the listening experiment were calibrated at the center of the sphere. The sound levels of the acoustic signals played back in the listening room were recorded with an identical measurement microphone unit as used in situ. The playback levels were then calibrated against the RMS values obtained in situ.

The empty listening room had a background noise lower than NC15 and a reverberation time of less than 10 ms, except for the lowest frequencies that had a reverberation time of less than 30 ms.

2.4 Data Collection Tool

In the listening experiment, the participants controlled the playback of the experimental sounds and assessed them with the aid of a 10.5 inch Android tablet. In the graphical user interface (developed in the project), numbered icons in the shape of small rectangles, represented the sounds. Initially, all icons were located at the top of the screen, and the participants assessed the sounds by moving the icons into a rectangular area representing the 2D space of perceived affective quality. The horizontal axis of the area represented Pleasantness and the vertical represented Eventfulness. To guide the participants in their assessments the right pole of the Pleasantness dimension was marked "Pleasant", and the left pole marked "Annoying". The top pole of the Eventfulness dimension was marked "Eventful" and the bottom pole was marked "Uneventful". The bottom right corner of the rectangular area was marked "Calm", the top right corner "Exciting", the top left corner "Chaotic", and the bottom left corner "Monotonous". This corresponds to the measuring model of perceived affective quality that Axelsson and co-works have developed [10, 18] (cf. [9, 19]). The playback started when clicking on an icon, and paused when clicking on the empty background of the rectangular area. The sounds were looped until paused or another sound was played.

2.5 Procedure and Design

The experiment took place in a soundproof listening room. All participants were tested

individually. First, a hearing test was performed to ensure that the participant had a normal hearing ability. During the experiment the participant sat on a chair at the center of the loudspeaker array, surrounded by the 29 loudspeakers. The task was to assess the experimental sounds on how pleasant and eventful the participant perceived them to be, by the aid of the data collection tool.

In total, the participant assessed 120 sounds, which were divided in four predefined sets of 30 sounds in each (i.e., always the same 30 sounds in each set for all participants). In order to evaluate the internal consistency of the assessments, the 120 sounds included 18 duplicates. They were distributed in such a way that within every set, 2 different sounds occurred twice ($2 \times 4 = 8$ duplicates), and 5 sounds occurred once in Set 1 and 3, and another 5 sounds occurred once in Set 2 and 4 ($5 \times 2 = 10$ duplicates). For each of the four sets, the duplicates were selected to represent opposite extremes of the Pleasantness-Eventfulness space. Also the 18 duplicates were predefined, and thus always the same for all participants. In every session, the 30 sounds were presented in a unique random order. The order in which the four sets were presented to the participants was also randomized.

All 30 sound icons remained on the screen in a session, and the participant was free to alter the assessments throughout the session. However, once the participant moved on to the next set of 30 sounds, it was not possible to return to a previous set.

2.6 Machine Learning

To handle the extraction of acoustic features from the recordings, the Bag of Frames (BOF) approach was used [20]. First, a set of acoustic features were extracted from the 30 s time segments of the mono recordings (obtained by the free-field measurement microphone) that corresponded to each of the 102 unique excerpts that the participants assessed in the listening experiment. However, in 9 cases the time segments of the mono signal accidentally included part of the calibration tone, and could not be used for the machine learning. Thus, 93 mono signals were used. Each of these 93 mono signals were divided into 1406 time frames, of 2048 samples each, corresponding to 42.67 ms. There was a 50% overlap between consecutive frames (i.e., 1024 samples, or 21.33 ms).

With the aim to find a set of features that is sufficiently small and that results in a model which is good enough for the purpose of the study, the following features were used (for a review and descriptions of available features, see [21, 22, 23]):

a) Auto-correlation (AC)
b) Linear Prediction Coefficients (LPC)
c) Mel-Frequency Cepstral Coefficients (MFCC)
d) Mel Spectrum (MEL)
e) Octave Band Signal Intensity (OBSI)
f) Octave Band Signal Intensity Ratio (OBSIR)
g) Relative Loudness (RL)
h) SpectralCrestFactorPerBand (SCFPB)

Second, a Gaussian Mixture Model (GMM) with 13 components [24, 25] was used to cluster the data extracted from each of the 93 mono signals. The GMM was fitted by an expectation maximization (EM) method. This procedure collapses the time domain, and reduces the amount of the data to a manageable size. Collapsing the time domain prevents the order of events in the acoustic signal to influence the result. Some of the features used have adjustable parameters. In these cases, the parameter space was searched to find the best settings. Before clustering, the feature data was standardized (i.e., $M = 0$, $SD = 1$).

Third, the dissimilarities between the GMMs were calculated, using the likelihood of the feature data for a given GMM. The resulting dissimilarity matrix was used to train two separate Support Vector Regression (SVR) learning algorithms [26]; one for each dimension of the soundscape model (i.e., one for Pleasantness and one for Eventfulness). The results of the listening experiment was used as the target values in the training of the SVRs. A Gaussian radial basis function kernel [27] was used for non-linear mapping of the extracted features.

The 10-fold cross-validation method was used [25, 28]. The 93 mono signals were divided into 10 sets of equal size. One set was put aside, and the remaining nine sets were used to train the system. The performance of the system was then evaluated by how well it could predict the value of the stimuli in the set that was put aside (i.e., the validation set). This process was repeated until all ten sets were used as validation set. The average of the ten validations was used as the result.

To find the optimal combination of features and parameter settings, a genetic search algorithm was developed (cf. [29]). The SVRs systems were trained with 50 random points in the search space.

The best individuals (the points) are then selected to create a new generation using cross breading and mutation to simulate an evolutionary process. The process was stopped after 50 generations. The process was repeated 5 times and the best was used as the result.

The machine learning system was programmed in Python using the SciPy library for scientific computation [30, 31] and the Scikit-learn package for machine learning [32, 33]. Feature extraction where conducted with the Yaafe package [22, 34].

3 RESULTS

First, the data from every participant was screened for the internal consistency of the individual assessments. Pearson's coefficient of correlations (r) were calculated for the first and second time the participant assessed the 18 duplicated sounds on Pleasantness and Eventfulness. As described in Section 2.1, one participant was excluded from the analyses after the screening, because the correlation coefficient for Pleasantness was 0.06, and 0.46 for Eventfulness. Table 1 presents the internal consistency of the individual assessments for all the remaining 30 participants in the form of the minimum, maximum and average correlation coefficients. For comparison with the results of the machine learning, Table 1 also includes the squared correlation coefficients (R^2).

Table 1 – Internal consistency of individual assessments

	Pleasantness		Eventfulness	
	r	R^2	r	R^2
Min	0.67	0.45	0.61	0.37
Max	0.97	0.94	0.97	0.95
Mean	0.81	0.66	0.84	0.71

Figure 1 presents the arithmetic mean values of the Pleasantness and Eventfulness scores from the listening experiment. These results were used as target values in the training of the SVRs.

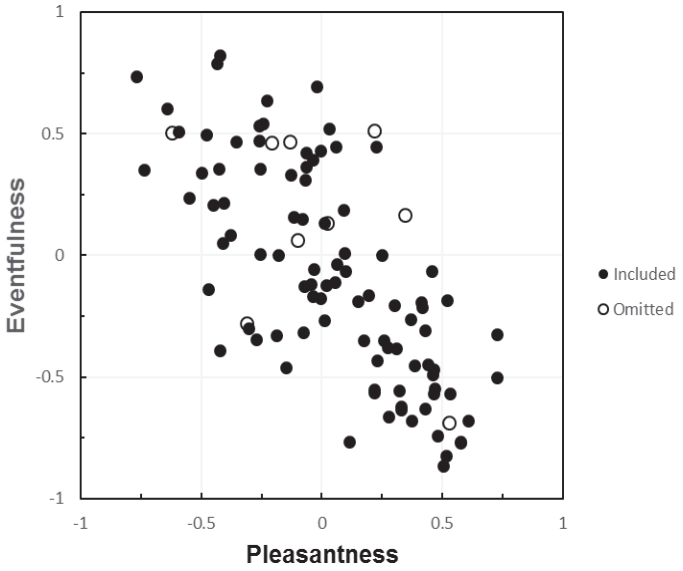

Figure 1 – Arithmetic mean values of the Pleasantness and Eventfulness scores from the listening experiment. Filled symbols represent the 93 sounds that were used for machine learning, and the open symbols the 9 that were omitted.

Table 2 presents the Mean Squared Error (MSE) values (the sum of both dimensions), as well as the squared rank-order correlation coefficients (R^2) for Pleasantness and Eventfulness, obtained by comparing the empirical solution presented in Figure 1 with the solution that the SVRs predicted for each type of feature. Table 1 is organized by the MSE values, from the best to the worse model fit. A criterion for selecting a machine learning model is that it performs at least as well as the internal consistency of a person. For this reason, R^2 values equal to or larger than the mean values of internal

consistency (Table 1) are presented in boldface font.

Table 2 – Model fit

		R^2	
Feature	MSE	Pleasantness	Eventfulness
MFCC	0.07	**0.74**	**0.83**
OBSI+ OBSIR	0.09	**0.71**	**0.74**
OBSI	0.11	**0.66**	0.66
LPC	0.11	0.63	0.66
MEL	0.12	**0.67**	0.62
RL	0.12	0.52	0.68
OBSIR	0.13	0.47	0.69
SCFPB	0.13	0.52	0.64
AC	0.20	0.42	0.33

Note: R^2 values equal to or larger than the mean values of internal consistency (Table 1) are presented in boldface font.

Table 2 shows that two of the features performed better than the average participant. The best result was achieved with MFCC (34 coefficients).

4 DISCUSSION

The results show that it is possible to predict human responses to acoustic environments based on data from acoustic signals. In this study, the Mel-Frequency Cepstral Coefficients provided the best machine learning model. It performed better than any other feature, as well as better than the average participant in the listening experiment, in terms of the internal consistency of the assessments.

One may argue that a machine should not be able to predict soundscape assessments, because people base their judgments on experience that a machine cannot have, and that there are some higher order cognitive functions involved in the assessments. The present results indicate the opposite. When disregarding individual variation in the data, predicting the average responses, all necessary information is available in the acoustic signal, which is very promising for soundscape mapping.

It is possible that it is the BOF approach that makes machine learning work. It collapses the time domain and creates a representation of the long-term statistics of the spectral information. There are indications that the auditory system uses long-term statistics in processing sounds that can be classified as 'sound textures' [35]. Such sounds, like distant road traffic, is typical for the urban environment.

There is not yet any consensus on how to conduct machine learning studies on soundscape. A review of the available literature provides a rather messy picture. No two studies use comparable approaches. Consequently, it is impossible to compare the present with previous results. Perhaps this also reflects a lack of consensus among soundscape researchers on the purpose and objectives of soundscape research, in general. Nevertheless, we disagree with Lagrange et al. [7] that the BOF approach is insufficient for modeling urban soundscapes. It is also interesting to note that the Mel-Frequency Cepstral Coefficients reoccur in the literature and frequently provide good results. This indicates that it is important to include features based on the frequency domain of sounds.

A potential limitation in the present study is that we did not manage to identify all combinations of perceived affective quality of soundscape. The lower left and the upper right sectors of Figure 1 are empty. This means that the machine learning system that we created in this study may be incomplete. For this reason we continue our effort and aim to learn more about the causes of perceived affective quality of soundscape, in order to fill the space in future studies, and to achieve a complete machine learning system.

ACKNOWLEDGEMENTS

This research project was funded by Grant MMW2012.0033 from the Marianne and Marcus Wallenberg Foundation. We also acknowledge The Royal Society, and Sweden's innovation agency VINNOVA.

REFERENCES

1. ISO 12913-1:2014. Acoustics—Soundscape—Part 1: Definition and Conceptual Framework. Geneva, Switzerland: International Organization for Standardization (ISO); 2014.
2. Aletta F, Kang J, Axelsson Ö. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landscape and Urban Planning 2016; 149: 65–74.
3. Casey M, Veltkamp R, Goto M, Leman M, Rhodes C, Slaney M. Content-Based Music Information Retrieval: Current Directions and Future Challenges. Proceedings of the IEEE 2008; 96: 668–696.
4. Oldoni D, De Coensel B, Boes M, Rademaker M, De Baets B, Van Renterghem T, Botteldooren D. A computational model of auditory attention for use in soundscape research. The Journal of the Acoustical Society of America 2013; 134: 852–861.
5. Yu L, Kang J. Modeling subjective evaluation of soundscape quality in urban open spaces: An artificial neural network approach. The Journal of the Acoustical Society of America 2009; 126: 1163–1174
6. Aucouturier J-J, Defreville B, Pachet F. The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music. Journal of the Acoustical Society of America 2007; 122: 881–891.
7. Lagrange M, Lafay G, Défréville B, Aucouturier J-J. The bag-of-frames approach: A not so sufficient model for urban soundscapes. Journal of the Acoustical Society of America 2015; 138: EL487–EL492
8. Torija AJ, Ruiz DP, Ramos-Ridao AF. A tool for urban soundscape evaluation applying Support Vector Machines for developing a soundscape classification model. Science of the Total Environment 2014; 482-483: 440–451.
9. Fan J, Thorogood M, Riecke BE, Pasquier P. Automatic recognition of Eventfulness and Pleasantness of soundscape. In: Proceedings of the Audio Mostly 2015 on Interaction with Sound. New York, NY, USA: ACM; 2015. p. 12:1–12:6
10. Axelsson Ö, Nilsson M.E, Berglund B. A principal components model of soundscape perception. Journal of the Acoustical Society of America 2010; 128(5): 2836–2846.
11. http://www.soundideas.com.
12. http://www.sfu.ca/truax/wsp.html.
13. Davies WJ, Bruce NS, Murphy JE. Soundscape reproduction and synthesis. Acta Acustica United with Acustica 2014; 100(2): 285–292.
14. Guastavino C, Katz BF, Polack J, Levitin DJ, Dubois D. Ecological validity of soundscape reproduction. Acta Acustica United with Acustica 2005; 91: 333–341.
15. Axelsson Ö. How to measure soundscape quality. In: Proceedings of Euronoise 2015. Maastricht, The Netherlands: Nederlands Akoestisch Genootschap and ABAV - Belgian Acoustical Society; 2015. Paper 67.
16. Moreau S, Daniel J, Bertet S. 3D sound field recording with higher order Ambisonics: Objective measurements and validation of spherical microphone. Audio Engineering Society Convention 120, May 2006; 2006. Paper 6857.
17. Leopardi P. A partition of the unit sphere into regions of equal area and small diameter. Electronic Transactions on Numerical Analysis 2006; 25: 309–327.
18. Axelsson Ö, Nilsson ME, Berglund B. A Swedish instrument for measuring soundscape quality. In: Kang J, editor. Proceedings of Euronoise 2009: Action on Noise in Europe. Edinburgh, Scotland: Institute of Acoustics; 2009. Paper EN09_0179.
19. Russell JA, Weiss A, Mendelsohn GA. Affect grid: A single-item scale of pleasure and arousal. Journal of Personality and Social Psychology 1989; 57(3): 493–502.
20. Wang, W. editor. Machine Audition: Principles, Algorithms and Systems. Hershey, PA: Information Science Reference; 2011.
21. Peeters G. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. CUIDADO I.S.T. Project Report; 2004.
22. Mathieu B, Essid S, Fillon T, Prado J, Richard G. YAAFE, an easy to use and efficient audio feature extraction software, In: Proceedings of the 11th ISMIR conference, Utrecht, Netherlands; 2010. p. 441–446.
23. Davis SB, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech and Signal Processing 1980; 28(4): 357–366.
24. McLachlan G. editor. Mixture Models. New York, NY: Marcel Dekker; 1988.
25. Murphy KP. Machine Learning: A Probabilistic Perspective. Cambridge, MA: MIT Press; 2012.
26. Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing 2004; 14: 199–222.

27. Schölkopf B, Smola AJ. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Cambridge, MA: MIT press; 2002.
28. Duan K, Keerthi SS, Poo AN. Evaluation of simple performance measures for tuning SVM hyperparameters. Neurocomputing 2003; 51: 41–59.
29. Fröhlich H, Chapelle O, Schölkopf B. Feature selection for support vector machines by means of genetic algorithm. In: Tools with Artificial Intelligence, 2003. Proceedings. 15th IEEE International Conference on. IEEE; 2003. p. 142–148.
30. https://www.scipy.org/
31. Jones E, Oliphant E, Peterson P, et al. SciPy: Open Source Scientific Tools for Python, 2001-, http://www.scipy.org/ [Online; accessed 2016-05-16].
32. http://scikit-learn.org/
33. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 2011; 12: 2825–2830.
34. http://yaafe.sourceforge.net/
35. McDermott JH, Schemitsch M, Simoncelli EP. Summary statistics in auditory perception. Nature Neuroscience 2013; 16: 493–498.