



# Audio Feature Analysis for Precise Vocalic Segments Classification in English

Szymon Zaporowski<sup>1</sup>(✉)  and Andrzej Czyżewski<sup>2</sup> 

<sup>1</sup> Audio Acoustic Laboratory, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland  
smck@multimed.org

<sup>2</sup> Multimedia Systems Department, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland  
ac@pg.edu.pl

**Abstract.** An approach to identifying the most meaningful Mel-Frequency Cepstral Coefficients representing selected allophones and vocalic segments for their classification is presented in the paper. For this purpose, experiments were carried out using algorithms such as Principal Component Analysis, Feature Importance, and Recursive Parameter Elimination. The data used were recordings made within the ALOFON corpus containing audio signal recorded employing 7 speakers who spoke English at the native or near-native speaker level with a Standard Southern British English variety accent. The recordings were analyzed by specialists from the field of phonology in order to extract vocalic segments and selected allophones. Then parameterization was made using Mel Frequency Cepstral Coefficients, Delta MFCC, and Delta Delta MFCC. In the next stage, feature vectors were passed to the input of individual algorithms utilized to reduce the size of the vector by previously mentioned algorithms. The vectors prepared in this way have been used for classifying allophones and vocalic segments employing simple Artificial Neural Network (ANN) and Support Vector Machine (SVM). The classification results using both classifiers and methods applied for reducing the number of parameters were presented. The results of the reduction are also shown explicitly, by indicating parameters proven to be significant and those rejected by particular algorithms. Factors influencing the obtained results were discussed. Difficulties associated with obtaining the data set, its labeling, and research on allophones were also analyzed.

**Keywords:** Allophone classification · Feature Importance · Random forest

## 1 Introduction

In recent years the growing interest of scientists and specialists from the industrial sector in technologies related to speech signal processing can be observed [1, 2]. According to

---

Dataset employed in this research available at website: <https://modality-corpus.org> in ALOFON corpus section.

© Springer Nature Switzerland AG 2020

A. Dziech et al. (Eds.): MCSS 2020, CCIS 1284, pp. 265–277, 2020.

[https://doi.org/10.1007/978-3-030-59000-0\\_20](https://doi.org/10.1007/978-3-030-59000-0_20)

some predictions, it is possible that more than 50% of all search queries over the internet will be utilized by voice; also, mobile device communication can be done the same way [3]. As voice assistants increase in popularity, in particular the solutions proposed by Amazon, Apple, and Google, the use of voice communication with devices will require the development of finer solutions to better understand voice commands. Finding better solutions for voice communication with computer devices will probably require not only the development of solutions related to machine learning or deep learning but also analysis of speech itself in a deeper context and meaning. Approaching the problem in this way could be more challenging due to the fact that the allophone articulation is not a simple process in terms of its analysis. First of all, an allophone is any of the various phonetic realizations of a phoneme in a language, which do not contribute to distinctions of meaning. For example, realization of “p” in word “sip” is different in word “seep”. Each of realization of “p” in these words are different allophones. Especially, the dependence on the preceding sounds, context, short duration time and individual features require a change in the approach popular in the literature. In practice, speech is quite rarely analyzed in such a detailed way as to go to the allophonic level. Works on analysis aspects related to the level of allophones are usually conducted by linguists and they are not strictly related to the application in the broadly understood machine learning.

The analysis of allophones is mainly used in relation to the quality of pronunciation in a given language. There are some relationships, such as the difference between dark /l/ and light /l/ in English, for example, as in Recasens’s paper [4]. Based on the implementation of such an allophone analysis, it is possible to determine whether a given person who speaks English comes, for example, from England or from Scotland [5]. This topic, in particular, the recognition of the speakers’ accent, has potential applications in security technologies. There are also many different fields where allophone recognition and analysis can be beneficial e.g., for biometrics with a speaker identification system, where allophone analysis can determine the identity of a person [6]. Also, allophones can help in the early detection of dyslexia. However, this issue concerns allophonic perception and its disorders [7].

The plan of this article is as follows, in the second section, an overview of works related to the topic of speech research at the allophonic level, the parameterization process, parameter reduction, and classification of allophones and vocalic segments are presented. The third section concerns the methodology, a description of the used corpus, and its most important features. Then the parameterization method is briefly described. Next, three selected algorithms that were used to reduce and to indicate the significance of parameters in the obtained feature vector are presented. These algorithms applied are Extremely Randomized Trees, Principal Component Analysis, and Recursive Feature Elimination. The fourth section contains the description and results of the experiments. The fifth and final section includes discussion and conclusions.

## 2 Related Work

Several works have been found in the literature related to the classification of allophones. Most of this work was written as part of a larger project on the multimodal methodology

for allophone research [3, 8–10]. In most cases, simple artificial neural network architectures consisting of at least one hidden layer were used for classification. Algorithms K-Nearest Neighbors, Support Vector Machines, Self-Organizing Maps, Convolutional Neural Networks were also used [3, 8]. In the case of clustering, Gaussian Mixture Models and Hidden Markov Networks were employed [11, 12].

In the parameterization of the speech signal, various parameters are used. Based on the experiments carried out as part of the study of allophones [13, 14], a set of features consisting of acoustic descriptors, their derivatives, and statistical values, which carry information, was created. Such a feature vector includes parameters in both time and frequency domain. Time-domain parameters include the center of gravity (Temporal Centroid - TC), number of zero crossings (Zero Crossing - ZC), RMS energy (Root Mean Square Energy), and peak value to RMS value (Peak to RMS). Frequency domain-related parameters were obtained by transforming the time domain signal using a Discrete Fourier Transform. The following spectral characteristics were used: the power density spectral center (Audio Spectrum Centroid - ASC), the power density gravity center variance (Variance of Audio Spectrum Centroid - varASC), the mean square power deviation (Audio Spectrum Spread - ASSp), the mean square deviation variance power density spectra (Variance of Audio Spectrum Spread - varASSp), the skewness of the mean square power density spectrum (Audio Spectrum Skewness ASSk), the variance of the mean square power spectrum deviation (Variance of Audio Spectrum Skewness - varASSk) [15, 16]. Since Mel Frequency Cepstral Coefficients are most often used in speech signal recognition [17, 18], that is why the focus in this paper is put on this type of feature.

### 3 Methodology

#### 3.1 Data Acquisition

The recordings were made in a room prepared for voice recordings with the two acoustic dampers. They were used to reduce unwanted sounds. Two microphones, a super-directional microphone, and a lavalier microphone and an external recorder were employed to record the audio signal. The signal from microphones was recorded with a 16-bit resolution and 48 kHz/s sampling rate. Also, video and Facial Motion Capture (FMC) modalities were recorded using Six Vicon Vue cameras, Vicon reference Camera, digital camera, and sports camera. However, it must be noted that those modalities are not subject to this research. Other authors' work describes an approach that uses FMC and combined audio modality with FMC to classify allophones and vocalic segments. The arrangement of the equipment in the room used in the context of wider experimental research is presented in Fig. 1.

The recording sessions lasted two days. Seven speakers were recorded. Audio recordings acquired during these sessions were used in the experiments described in this Each speaker uttered 300 short expressions or words. The recorded subjects were native or near-native speakers speaking with varying pronunciation (with Standard Southern British English variety). The level of English pronunciation, gender, and accent of each



**Table 2.** List and number of vocalic segments recorded for all speakers

Segment	Quantity	Segment	Quantity
ae	420	I	273
e	266	A	252
ei	203	O	182
i:	182	ai	154
eu	126	a:	119
u:	91	3:	70
o:	70	U	35
au	35	aiE	7
oi	7	ir	7
eE	7	ie	7

Another problem is the difference in the duration of each segment; it could last less than 40 ms for pure vowels to almost 400 ms for diphthongs. Due to highly imbalanced classes, there was a decision to reject from classification all classes with a number of elements smaller than 35.

Dataset presented in this research is available to download from the MODALITY corpus webpage available in the ALOFON corpus tab [19]. Detailed information about the above corpora is mentioned in the authors' previous works [8, 20].

### 3.2 Parameterization

This research presented in this paper uses the first 40 MFCCs – 20 first MFCC's average values, 20 first MFCC's variance values, and 20 delta MFCC and 20 delta-delta MFCC Coefficients. Using the formula (1) it is possible to calculate the first and second-order derivatives of MFCC:

$$d_t = \frac{\sum_{n=1}^2 n(c_t + n - c_t - n)}{2 \sum_{n=1}^2 n^2} \quad (1)$$

where  $c$  is the  $n$ th cepstral coefficient and first-order dynamic coefficient, and  $t$  stands for time.

The MFCC parameters were calculated for 512 samples of speech frames (with an overlap of 50%) and then the average and the variance values were calculated, which results in an 80-dimensional vector of parameters.

### 3.3 Feature Selection and Reduction Algorithms

Principal Component Analysis is one of the most popular and widely used algorithms for dimension reduction [21, 22]. PCA algorithm is used when there is a need for extracting crucial information from the data, compressing the size of the data vector and clarify

the dataset description. One of the advantages of this algorithm is the fact that is one of the easiest ways to analyze the structure of the observations and the variables. PCA was used to reduce the dimension of gathered data with minimal loss of accuracy in classification.

Extremely Randomized Trees (ERT) algorithm is deriving concept from the random forest algorithm [23]. It provides a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and is also characterized by the same distribution for all trees in the forest. The error connected with generalization for forests converges to a limit as the number of trees in the forest grows. The generalization error of a forest of tree classifiers depends on the correlation between trees in the forest and on the strength of the individual trees in the whole set [24, 25]. The algorithm was implemented using scikit python library [26]. ERT algorithm was used with 10 estimators, entropy criterion, minimum sample split equal two, minimum samples per leaf equal one, a maximum number of features equal twenty and with balanced class weight to balance an uneven number of examples for classes of classified vocalic segments. Bootstrap and warm start settings were not used. The biggest advantage of the ERT is obtaining detailed information about the most important parameters forming the data vector.

The goal of the Recursive Feature Elimination (RFE) algorithm is to sort out features using recursive reduction of the dataset content. It uses an external estimator, which allocate weight for each feature. The algorithm works as follows. In the first step, the initial set of features is used to train the estimator. The importance of feature is received for the given attributes. In the next step, the less important feature is pruned out of the current set of features. This procedure is repeated recursively until the desired number of features is reached [27]. RFE model was implemented using scikit library [26].

### 3.4 Classification

A simple feedforward architecture was used as an artificial neural network classifier. It was a neural network used with three hidden layers consisting of 80 neurons, each with a Relu activation function and softmax functions on the last layer, which size is determined by the size of the classes being the object of classification. The network was implemented with the use of Keras library in the Python programming language [28]. For the training, the Stochastic Gradient Descent (SGD) algorithm with the learning rate 0.05 was used, and also a categorical cross-entropy loss function has been applied with 10 times cross-validation of data.

The classifier using the Support Vector Machine algorithm used in this work was implemented using the scikit-learn library in the Python programming language [26]. The data processing module used was identical to the one used for neural networks. Two types of the kernel were used: the polynomial kernel and the RBF (Radial Basis Function) kernel. In addition, it was decided to use weights for the classes used due to their heterogeneous number in order to balance them. The automatically selected gamma coefficient was also employed, while the shape of the decision function was adopted as one vs the rest. Data were split into 60/40 for training and testing sets.

For both classifications presented in this section, there were used 3 algorithms presented in Sect. 3.3. The results of the classification are presented in the next section.

## 4 Results

In Table 3, 4 and 5 results for ANN classification are presented employing ERT, RFE and PCA algorithm. Training took 1500 epochs for each validation. The number of features selected for each algorithm was 39 for ERT and 20 for both PCA and RFE. This number of coefficients was chosen according to obtained results. Presented results are average from 10 classification algorithm executions.

The data are presented by showing values for specific speakers instead of the average value for the whole set, because otherwise it would not be possible to see differences between individual speakers.

Table 6, 7 and 8 consists of results for SVM combined with ERT, RFE and PCA algorithms respectively. The number of features selected for each algorithm was the same as in the ANN approach. Results are also the average of 10 classification attempts for each algorithm. Figure 2 presents the results of the Feature Importance algorithm for the top 20 coefficients. Figure 3 shows the results of Feature Importance algorithm for each speaker separately.

**Table 3.** Results of data classification using ERT algorithm with ANN

Speaker ID	Accuracy	Precision	Recall
1	0.943	0.969	0.969
2	0.971	0.984	0.984
3	0.943	0.969	0.969
4	0.857	0.769	0.755
5	0.886	0.952	0.938
6	0.829	0.906	0.906
7	0.829	0.769	0.755

**Table 4.** Results of data classification using RFE algorithm and ANN

Speaker ID	Accuracy	Precision	Recall
1	0.943	0.936	0.969
2	0.914	0.964	0.953
3	0.971	0.964	0.984
4	0.914	0.805	0.802
5	0.943	0.956	0.969
6	0.914	0.955	0.953
7	0.857	0.923	0.922

**Table 5.** Results of data classification using PCA algorithm and ANN

Speaker ID	Accuracy	Precision	Recall
1	1	1	1
2	0.914	0.817	0.787
3	0.914	0.968	0.787
4	0.943	0.984	0.969
5	0.886	0.952	0.938
6	0.371	–	–
7	0.8	0.931	0.724

**Table 6.** Results of data classification using ERT algorithm and SVM

Speaker ID	Accuracy	MSE
1	0.857	0.517
2	0.857	1.025
3	0.828	0.907
4	0.771	0.841
5	0.857	1.204
6	0.457	0.041
7	0.914	0.954

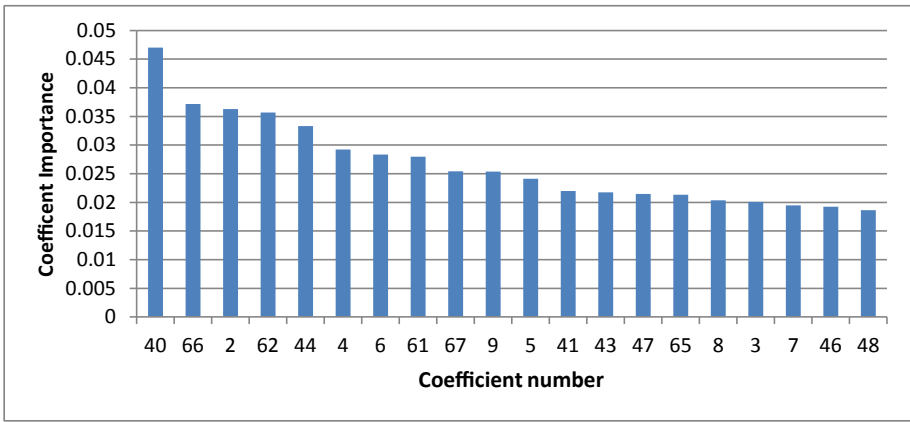
**Table 7.** Results of data classification using RFE algorithm and SVM

Speaker ID	Accuracy	MSE
1	0.857	0.471
2	0.886	0.897
3	0.886	0.4661
4	0.743	0.798
5	0.886	0.90
6	0.8	1.19
7	0.857	0.74

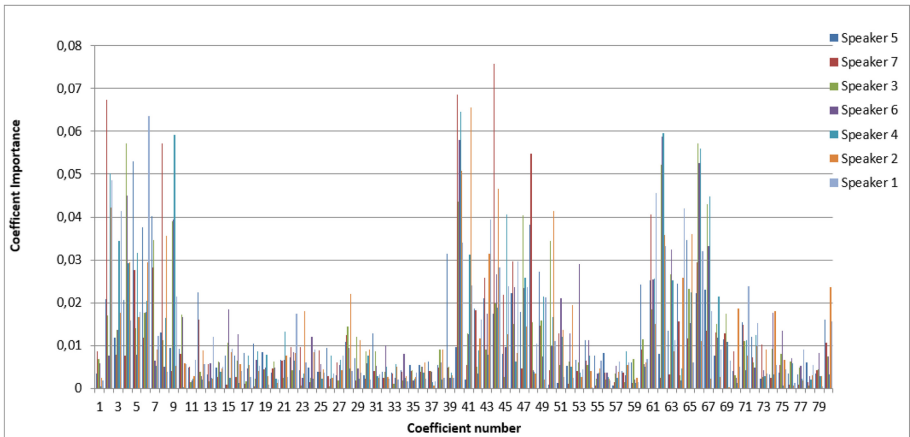


**Table 8.** Results of data classification using PCA algorithm and SVM

Speaker ID	Accuracy	MSE
1	0.857	0.591
2	0.886	1.075
3	0.914	0.491
4	0.829	0.751
5	0.743	1.549
6	0.457	0
7	0.886	1.025



**Fig. 2.** Results of Feature Selection algorithm for all speakers – mean value for top 20 coefficients



**Fig. 3.** Results of Feature Selection algorithm for all speakers – importance for all coefficients according to each speaker

## 5 Discussion

In the vast majority of scientific papers on subjects related to allophones, too little importance is attached to the significance of parameters found in the vector, which is directed to the classifier input. In the overwhelming majority of cases, the Principal Component Analysis (PCA) method is used, while not paying attention to which of the parameters considered by the algorithm are significant, thus they should be transferred to the classifier input, and which ones were treated as noise to be removed.

In the case of classification using neural networks, the RFE algorithm performed best, slightly behind the solution using Extremely Randomized Trees. The PCA algorithm worked the worst. It is puzzling that the results are unstable for the speaker ID 6. In the case of RFE and ERT algorithms, the results do not differ significantly from the results obtained for other speakers. However, the case of the PCA algorithm could not be exceeded by up to 50% accuracy. Furthermore, the recall and precision metrics could not be calculated due to the significant number of errors in classification. However, even in the absence of counting speaker ID 6 to the average value of the accuracy metric for individual algorithms, PCA performs the worst. The decrease in the value of the recall metric for ERT and PCA algorithms in relation to the RFE algorithm is also interesting. This indicates that the selection of features made for ERT and PCA is not necessarily the most optimal.

In the case of classification using Support Vector Machine, the RFE algorithm was again the best. The PCA algorithm performed slightly worse, but visibly weaker, while the ERF algorithm performed worse than it did. It is important that the same hyperparameter settings were used for each speaker in order to enable results comparison. Attention should be paid to the problem with speaker ID 6 again, the accuracy results of which in two out of three classifiers did not exceed 50%. Moreover, the MSE value for this speaker was divergent. In the case of the ERF algorithm, where the accuracy score was 80%, MSE indicates the possibility of under-training the algorithm. For PCA and RFE, MSE values are overfitting. For speakers ID 2 and ID 5, MSE values also indicate the possibility of matching hyperparameters.

Based on the conducted experiments, it was observed that the classifier using a neural network is more effective and more universal than the Support Vector Machine. In most cases, recall and precision metrics indicate the actual classifier has learned to recognize individual allophones but does not try to always guess the same etiquette, which is common for unbalanced classes.

According to the results of Feature Importance algorithm applied for all speakers, it is noteworthy that for each speaker, the most significant coefficients indicated by the algorithm differed. Nevertheless, it can be seen which coefficients most often appeared as the most important ones. Based on these results, a ranking of the top 20 most significant coefficients was created by calculating the average value for each coefficient.

Information on Feature Importance for all speakers indicates that the most important parameter was the variance of the MFCC. Interestingly, the list of 20 most important coefficients for the entire dataset did not include other coefficients regarding variance (positions 21–40). The list includes 5 delta MFCC coefficients (66 and 62 in second and fourth place), 8 coefficients constituting average MFCC values (2 in third place, 4 in sixth place, 6 in seventh place) and 5 delta MFCC coefficients (44 in fourth place). Based on

the results obtained from the Feature Importance algorithm, it cannot be clearly stated that one of the categories of coefficients is redundant. However, it can be observed that the MFCC coefficients with the highest indices and their derivatives are not as important as the coefficients with the lower indices. It indicates that it is not worth making parameterization using more than 15 factors. The only exception is the aforementioned variance parameter, which came from a factor of 20 and was the only one to represent the variance of MFCC coefficients in the prepared statement.

One of the things that are worth mentioning is the fact that the speaker's accent did not significantly affect the results of the classification. Also, the pronunciation level cannot be clearly linked to the quality of the classification.

## 6 Conclusions

In this paper, an approach to the classification of allophones and the significance of vocalic segments parameters using the Artificial Neural Network and Support Vector Machine is presented. Using the Extremely Randomized Trees, Principal Component Analysis and Recursive Feature Elimination algorithms, efforts were made to obtain the resulting feature vector to obtain the best classification result for both types of classifiers and at the same time check which coefficients in the feature vector are the most important using the Feature Importance algorithm. The obtained results indicate that the Recursive Feature Elimination algorithm works best, ahead of the Extremely Randomized Trees algorithm. It was found using the Feature Importance algorithm that it may not be clearly indicated which group of features is the most important. However, there are indications that features above MFCC and its derivatives do not affect classification. The same happens with the variance of MFCC coefficients.

It should be noted that the tests were carried out on a relatively small group of speakers. In order to conduct research on a larger scale, it would be necessary to solve the biggest problem associated with the classification of allophones, i.e., to develop automatic labeling and detection of allophones in recordings.

Further work related to the classification of allophones can be used, for example, in voice biometry, to verify users, because as is seen in the significance ranking of features, each speaker has a different set of the most important features, which can be used as a unique feature.

**Acknowledgment.** Research sponsored by the Polish National Science Centre, Dec. No. 2015/17/B/ST6/01874.

## References

1. Karpagavalli, S., Chandra, E.: A review on automatic speech recognition architecture and approaches. *Int. J. Sig. Process. Image Process. Pattern Recogn.* **9**(4), 393–404 (2016)
2. Xu, D., Qian, H., Xu, Y.: The state of the art in human–robot interaction for household services (chap. 6.1). In: Xu, D., Qian, H., Xu, Y. (eds.) *Household Service Robotics*, pp. 457–465. Academic Press, Oxford (2015)

3. Piotrowska, M., Korvel, G., Kurowski, A., Kostek, B., Czyżewski, A.: Machine learning applied to aspirated and non-aspirated allophone classification—An approach based on audio ‘Fingerprinting’. In: 145 Audio Engineering Society Convention (2018)
4. Recasens, D.: A cross-language acoustic study of initial and final allophones of //l. *Speech Commun.* **54**(3), 368–383 (2012)
5. German, J.S., Carlson, K., Pierrehumbert, J.B.: Reassignment of consonant allophones in rapid dialect acquisition. *J. Phon.* **41**(3–4), 228–248 (2013)
6. Dražanský, M., Orság, F.: Fingerprints and speech recognition as parts of the biometry. In: *Proceedings of 36th International Conference MOSIS*, pp. 177–183 (2002)
7. Noordenbos, M.W., Segers, E., Serniclaes, W., Mitterer, H., Verhoeven, L.: Allophonic mode of speech perception in Dutch children at risk for dyslexia: a longitudinal study. *Res. Dev. Disabil.* **33**(5), 1469–1483 (2012)
8. Zaporowski, S., Cygert, S., Szwoch, G., Korvel, G., Czyżewski, A.: Rejestracja, parametryzacja i klasyfikacja alofonów z wykorzystaniem bimodalności (2018)
9. Piotrowska, M., Korvel, G., Kostek, B., Ciszewski, T., Czyżewski, A.: Machine learning-based analysis of English lateral allophones. *Int. J. Appl. Math. Comput. Sci.* **29**(2), 393–405 (2019)
10. Zaporowski, S., Czyżewski, A.: Selection of features for multimodal vocalic segments classification. In: Choróś, K., Kopel, M., Kukla, E., Siemiński, A. (eds.) *MISSI 2018. AISC*, vol. 833, pp. 490–500. Springer, Cham (2019). [https://doi.org/10.1007/978-3-319-98678-4\\_49](https://doi.org/10.1007/978-3-319-98678-4_49)
11. Jeevan, M., Dhingra, A., Hanmandlu, M., Panigrahi, B.K.: Robust speaker verification using GFCC based *i*-vectors. In: Lobiyal, D., Mohapatra, D.P., Nagar, A., Sahoo, M. (eds.) *Proceedings of the International Conference on Signal, Networks, Computing, and Systems. LNEE*, vol. 395, pp. 85–91. Springer, New Delhi (2017). [https://doi.org/10.1007/978-81-322-3592-7\\_9](https://doi.org/10.1007/978-81-322-3592-7_9)
12. Xu, J., Si, Y., Pan, J., Yan, Y.: Automatic allophone deriving for Korean speech recognition. In: *Proceedings - 9th International Conference on Computational Intelligence and Security, CIS 2013*, pp. 776–779 (2013)
13. Korvel, G., Kostek, B.: Examining feature vector for phoneme recognition. In: *2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 394–398 (2017)
14. Kostek, B., Piotrowska, M., Ciszewski, T., Czyżewski, A.: No comparative study of self-organizing maps vs subjective evaluation of quality of allophone pronunciation for non-native English speakers. In: *Audio Engineering Society Convention 143* (2017)
15. Kostek, B., et al.: Report of the ISMIS 2011 contest: music information retrieval. In: Kryszkiewicz, M., Rybinski, H., Skowron, A., Raś, Z.W. (eds.) *ISMIS 2011. LNCS (LNAI)*, vol. 6804, pp. 715–724. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-21916-0\\_75](https://doi.org/10.1007/978-3-642-21916-0_75)
16. Sikora, T., Kim, H.G., Moreau N.: *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. Wiley, Hoboken (2005)
17. Eringis, D., Tamulevičius, G.: Modified filterbank analysis features for speech recognition. *Baltic J. Mod. Comput.* **3**(1), 29–42 (2015)
18. Zheng, F., Zhang, G., Song, Z.: Comparison of different implementations of MFCC. *J. Comput. Sci. Technol.* **16**, 582–589 (2001). <https://doi.org/10.1007/BF02943243>
19. Multimedia Systems Department: *Modality Corpus* (2018). <http://modality-corpus.org/>. Accessed 13 Feb 2020
20. Cygert, S., Szwoch, G., Zaporowski, S., Czyżewski, A.: Vocalic segments classification assisted by mouth motion capture. In: *2018 11th International Conference on Human System Interaction (HSI)*, pp. 318–324 (2018)
21. Bro, R., Smilde, A.K.: Principal component analysis. *Anal. Methods* **6**, 2812–2831 (2014)

22. Abdi, H., Williams, L.J.: Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2**(4), 433–459 (2010)
23. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *Mach. Learn.* **63**(1), 3–42 (2006). <https://doi.org/10.1007/s10994-006-6226-1>
24. Louppe, G., Wehenkel, L., Sutura, A., Geurts, P.: Understanding variable importances in forests of randomized trees. *Adv. Neural. Inf. Process. Syst.* **26**, 431–439 (2013)
25. Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P.: Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **43**(6), 1947–1958 (2003)
26. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2012)
27. Mao, Y., Pi, D., Liu, Y., Sun, Y.: Accelerated recursive feature elimination based on support vector machine for key variable identification. *Chin. J. Chem. Eng.* **14**, 65–72 (2006)
28. Chollet, F.: Keras (2015). <https://keras.io>. Accessed 10 Sept 2018