INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2026, VOL. 72, NO. 1, PP. 1-7
Manuscript received January 19, 2026; revised January, 2026. doi: 10.24425/ijet.2026.157865

Acoustic analysis of selected homographs for speech recognition systems

Dominik Lentas, and Michał Łuczyński

***Abstract*—This paper presents an acoustic analysis of selected homographs in the context of automatic speech recognition (ASR) systems. The study focuses on the Polish words "Dania" (eng. Denmark) and "dania" (eng. meals), which, despite identical spelling, differ subtly in pronunciation. These differences pose challenges for ASR systems, especially when context is unavailable.**

The methodology includes spectrograms analysis MFCC (Mel-Frequency Cepstral Coefficients) extraction, and classification using a Support Vector Machine (SVM) algorithm. A custom audio database was created using recordings from ten speakers, followed by manual segmentation and normalization of samples. Spectrograms and formant trajectories were analyzed to identify phonetic distinctions, particularly the presence of the semi-vowel [j] in "Dania".

A subjective listening test involving 27 participants was conducted to assess human recognition accuracy. Results showed an average recognition rate of 58%, indicating significant ambiguity. In contrast, the machine learning model achieved up to 79% accuracy with randomly stratified data and 75% accuracy when tested on the same samples used in the subjective test.

The findings suggest that MFCC-based classification combined with SVM is a promising approach for distinguishing homographs in speech, outperforming human listeners in controlled conditions. Limitations include the small dataset and variability in speaker articulation. The study highlights the importance of phonetic exception handling in ASR systems and proposes extending the method to other homographic pairs.

***Keywords*—homographs; speech signal; machine learning; MFCC; SVM**

I. Introduction

AUTOMATIC speech recognition (ASR) systems have become indispensable in modern human-computer interaction, enabling voice-controlled interfaces, real-time transcription, and multilingual communication. As these systems evolve, their ability to accurately interpret spoken language in different contexts, e.g. in people with atypical articulation, is becoming increasingly important [1].

One of the persistent challenges in ASR is recognizing homographs – words that have identical spellings but differ in pronunciation and meaning. In languages such as Polish, where phonetic nuances are subtle and context-dependent, homographs are ~~a~~ significant obstacle to reliable interpretation of speech. Such minimal phonetic contrasts are often masked by the coarticulation and variability of the speaker, making them difficult to detect [2].

The paper examines the differences between the polish homographs "Dania" (eng. Denmark) and "dania" (eng. meals) using the International Phonetic Alphabet and the Slavic Phonetic Alphabet [3],[4]. The former is the most widely used alphabet of this type and is used to standardize phonetic notation for all languages of the world. The Slavic transcription system additionally takes into account the notation of compact-slit consonants, which increases its usefulness for writing the sounds of Slavic languages. In order to simplify the identification of the context, designations have been used, where (D) will refer to the word "Dania" understood as a country, and (d) to the word "dania" understood as a meals.

According to the International Phonetic Alphabet (IPA):
- Dania (D) as a country: [dãɲja]
- dania (d) as meals: [dãɲa]

According to the Slavic Phonetic Alphabet (AS):
- Dania (D) as a country: [dãńi̯a]
- dania (d) as meals: [dãńa]

According to these two transcriptions, "Dania" (D) and "dania" (d), are words that differ in pronunciation. In the first homograph (D) there is an additional semi-open consonant in the form [j] in the IPA notation or [i] with a glide occurring in the AS notation. The difference exists, however, due to the fact that semi-open consonants are often shorter than vowels, have less acoustic energy [5] and are usually a transient sound [2] can merge with other phonemes. Because of these factors, they are likely to be more difficult for the human ear to pick up and harder to recognize for automatic speech recognition systems.

This study investigates the acoustic characteristics of homographs and evaluates their distinguishability using both traditional signal analysis and machine learning techniques. By combining time–frequency analysis, Mel-Frequency Cepstral Coefficients (MFCC), and Support Vector Machine (SVM) classification, we aim to assess the limitations of human perception and the potential of automated systems in resolving phonetic ambiguity. The findings contribute to the development of more robust ASR systems, particularly in scenarios where contextual information is unavailable or unreliable.

The paper is arranged as follows: in Chapter II, the state of the art in the field of homograph analysis, extraction of speech signal features, application of machine learning models to speech classification problems is presented. Chapter III presents preparation the audio sample database, the pre-processing of audio signals, the tools and software used, the extraction of features, and the description of the machine learning model for classification purposes. Chapter IV describes the experiments

Dominik Lentas and Michał Łuczyński are with Wroclaw University of Science and Technology, Poland (e-mail: 263707@student.pwr.edu.pl, michal.luczynski@pwr.edu.pl).

and their results. Finally, a discussion of the results and conclusions were presented.

II. State of the Art

A. Homographs and Phonetic Ambiguity

Homographs [5] are a well-known source of ambiguity in both written and spoken language. In speech, their correct interpretation relies heavily on contextual information, as subtle phonetic cues alone are often insufficient. Research has shown that even trained listeners may struggle to resolve homographs without immediate contextual support, with post-homograph context playing a critical role in disambiguation [6], and lexical ambiguity generally imposing processing challenges until broader sentence context is integrated [7].

Recent research has focused on improving homograph disambiguation using deep learning and contextual embeddings. For example, Nicolis and Klimkov proposed a lightweight classifier using contextual word embeddings, achieving 99.1% accuracy on English homographs without rule-based systems [8]. Similarly, Rezáčková et al. fine-tuned a T5 transformer model for homograph disambiguation in Text-to-Speech (TTS) systems, outperforming previous approaches [9].

B. MFCC Feature Extraction

MFCC are among the most widely used features in speech recognition [10]. They model the spectral envelope of speech signals in a perceptually relevant way, emphasizing frequency bands that are most informative to human hearing [11],[12]. MFCCs have proven effective in capturing phonetic distinctions, especially in scenarios with limited speech data or restricted computational resources, and in noisy conditions.

Recent studies have extended MFCC applications to emotion recognition [13], stuttering detection [14], and spoofing detection [15], demonstrating their versatility and robustness across domains.

C. SVM classifiers

Support Vector Machines (SVM) are supervised machine learning models that construct optimal decision boundaries between classes. In speech applications, SVMs have demonstrated high accuracy in tasks such as emotion recognition [16],[17], or accent recognition [18].

Hybrid approaches combining SVM with rule-based systems have also shown promise in homograph disambiguation, particularly in resource-constrained environments [19]. These methods balance interpretability and performance, making them suitable for deployment in mobile and embedded ASR systems.

III. Materials and Methods

A. Speech Database

To investigate the acoustic differentiation of selected homographs in Polish, a dedicated speech corpus was constructed. The dataset comprises recordings from ten native Polish speakers (four female, six male), who were instructed to read a text containing multiple instances of the target words: “Dania” (referring to the country) and “dania” (plural of “meal”). Due to coarticulation effects, manual segmentation was required to isolate target words. The editing process involved normalization to -0.1 dBFS and conversion to mono. The segmentation was performed using Ableton Live and Audacity, with Praat used for phonetic annotation and formant analysis. The text was composed to simulate natural speech and avoid artificial emphasis on the homographs.

Recordings were conducted in two acoustically treated environments suitable for voice recording:

- a professional studio at Akademickie Radio LUZ,
- a quiet room with sound-absorbing treatment using a Zoom H1n portable recorder.

Each speaker read the same text, which included five occurrences of “Dania” and seven of “dania”, embedded in semantically neutral sentences to minimize contextual bias.

B. Audio Preprocessing

From the raw recordings, 119 audio samples were manually extracted:

- 47 samples corresponding to Dania,
- 72 samples corresponding to dania.

The higher count results from retaining correctly pronounced words that occurred in repeated takes during recording sessions. Several recordings were excluded due to pronunciation errors, disturbance, or segmentation errors that could negatively impact feature extraction.

All samples were saved in lossless WAV format, with a bit depth of 24 bits and a sampling rate of 44.1 kHz. The preprocessing pipeline included:

- segmentation of individual word utterances,
- normalization of amplitude to -0.1 dBFS,
- conversion to single-chanel audio.

Particular care was taken during segmentation to account for coarticulation effects, which often obscure phoneme boundaries, especially in spontaneous speech. Normalization to -0.1 dBFS was chosen to prevent clipping and ensure consistent loudness across all samples, which is a standard practice in phonetic and ASR-related studies to maintain comparability of acoustic features.

C. Tools and Software

The acoustic analysis and machine learning experiments were conducted using the following tools:

- Praat: for formant tracking, pitch analysis,
- Audacity: for waveform visualization and manual editing,
- Python (v3.10) with the following libraries:
  - librosa: for MFCC computation and audio feature extraction,
  - scikit-learn: for SVM implementation and evaluation,
  - numpy, pandas, matplotlib: for data handling and visualization.

D. Feature Extraction

Each audio sample was parameterized using MFCC. The MFCC extraction process included:

- application of a mel-scaled filter bank,
- logarithmic transformation of the power spectrum,
- discrete cosine transform to obtain cepstral coefficients.

Thirteen MFCCs were computed per frame, a commonly adopted setting in speech processing that balances spectral

detail and computational efficiency. Samples of varying duration were zero-padded to ensure uniform input dimensions for classification. This step was necessary due to the variability in utterance duration and the requirements of *NumPy*-based data structures used in *scikit-learn*. Padding ensured compatibility with the SVM classifier and allowed consistent feature vector lengths. Additionally, trajectories of the first three formants (F1–F3) were analyzed, as these carry the most relevant phonetic information for distinguishing the target homographs, particularly the presence of the semi-vowel [j] in Dania. Higher formants (F4 and above) were not included because they contribute minimally to vowel and glide differentiation in Polish and add computational complexity without significant benefit for this task

E. Classification Model

A Support Vector Machine classifier with a radial basis function (RBF) kernel was trained to distinguish between the two homograph classes. The classification pipeline involved:

- flattening MFCC matrices into feature vectors,
- stratified splitting of the dataset into training and test sets using three configurations (25 measurements were taken for each aspect ratio, calculating the average accuracy):
  - 80/20,
  - 70/30,
  - 60/40.

To address variability in sample length, all feature vectors were equalized using zero-padding based on the longest sample. The model was evaluated using standard metrics [11,12]:

- accuracy, i.e. a measure proving the accuracy of a given model. It is the ratio of correct predictions to all predictions.
- precision, i.e. a measure of correct class A recognition in relation to all correct class A and B recognitions
- recall, i.e. a measure of class A recognition in relation to all correct recognitions and incorrect A recognitions
- f1, talks about the balance between the precission and recall measures.

In a separate experiment, the classifier was tested on a manually selected subset of 40 samples used in a subjective listening test. The test set was manually matched to the samples used in the subjective test, allowing direct comparison between human and machine classification. Ten classification trials were conducted to assess model stability and identify consistently misclassified samples. In addition to standard stratified splits (80/20, 70/30, 60/40), a manually selected test set was prepared to reflect the conditions of a subjective listening experiment. This configuration was used to evaluate the model's performance under controlled phonetic ambiguity, with results presented in Section IV.

IV. EXPERIMENTS AND RESULTS

A. Subjective test

A auditory evaluation was conducted to assess the human ability to distinguish between the homographs "Dania" (country) and "dania" (meals) in Polish. A total of 27 participants completed an online listening test comprising 40 audio samples (20 per class). Each participant was asked to classify each sample using one of three labels:

- "D" – if the participant believed the word was "Dania" (country),
- "d" – if the participant believed the word was "dania" (meals),
- "Don't know" – if the participant was unsure.

It was considered that the "I don't know" option could also turn out to be important information, because the samples in the test often sounded ambiguous. In addition, this reduced the chances that the respondents would fill out the test randomly. The test was designed to be completed in under 10 minutes, with instructions recommending the use of headphones in a quiet environment. The interface allowed repeated playback of each sample to minimize random guessing.

An analysis of the responses of 27 subjective test participants showed that recognizing homographs without context is a difficult task. The average success rate was 58% and the median correct response was 24 out of 40 samples (60%). Of the 40 samples, 16 were correctly recognized by 50% or less of the participants, while 24 samples were recognized by more than 50% of the participants. The most difficult sample turned out to be one containing the word "dania" (ang. meals) pronounced with an atypical glide, which visually resembled "Dania" (ang. Denmark) on the spectrogram; it was correctly recognized by only 4 participants.

Factors influencing the difficulty of recognition included:
- poor articulation (e.g. lack of a clear phoneme [j]),
- coarticulation, i.e. smooth transitions between words,
- lack of context, which in natural communication facilitates interpretation.

B. Spectrogram analysis

Spectrograms were generated for selected samples to visualize time–frequency characteristics and formant trajectories.

The most obvious difference is the additional phoneme. The changes should occur at the time of transition of phonemes [ɲja] and [ɲa] (notation consistent with International Phonetic Alphabet). However, due to the fact that the ~~additional~~ semi-open consonant [j] occurring only in the word "Dania" (D) is a transitional sound, it can be difficult to notice. The semi-open consonant [j] forms with the vowel [a] glide, which should be visible in the smooth transition of the second and third formants between phonemes. During the analysis, it was noted that in the word "Dania" (D) the second and third formants are often closer to each other (at the moment of transition from the phoneme [j] to the phoneme [a]) than in the case of the word dania (d). In addition, "Dania" (D) tends to exhibit greater pitch variation, whereas "dania" (d) is usually pronounced with a relatively flat intonation pattern. Unfortunately, these methods of assessing differences are not strict rules but rather useful guidelines. There are numerous factors that complicate the identification of differences between the two words.

Particular attention was paid to formants F1 and F2, intonation patterns, and the presence of the palatal approximant [j], classified as a consonant in Polish phonetics.

For illustration, spectrograms comparing "Dania" (country) and "dania" (meals) are shown for four speakers: A (Figures 1-2), B (Figures 3–4), G (Figures 5–6), and J (Figures 7–8).

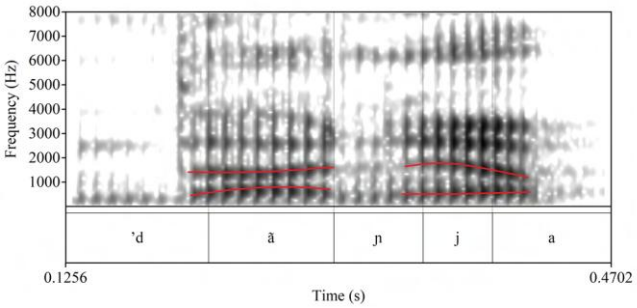

Fig. 1. The word "Dania" (D) on a spectrogram with the course of the formants marked in red (speaker A)

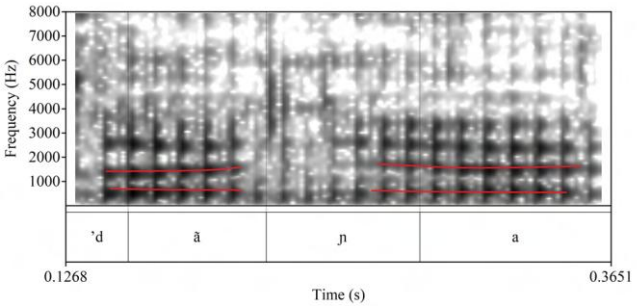

Fig. 2. The word "dania" (d) on a spectrogram with the course of the formants marked in red (speaker A)

The formants of the word "dania" (d) shown in Figure 2 are flat and look like straight lines, and the formants of the word "Dania" (D) visible in Figure 1 form curved lines going upwards, reaching a maximum at the time of transition to the phoneme [j], and then descending in the vicinity of the phoneme [a].

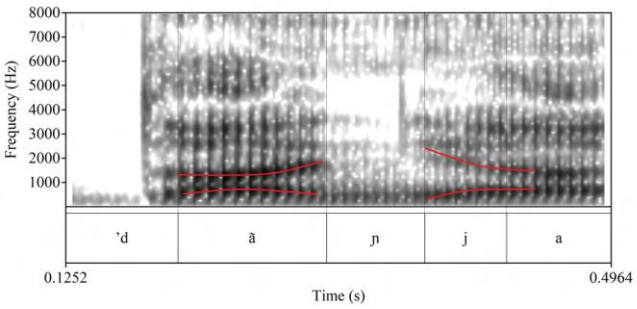

Fig. 3. The word "Dania" (D) on a spectrogram with the course of the formants marked in red (speaker B)

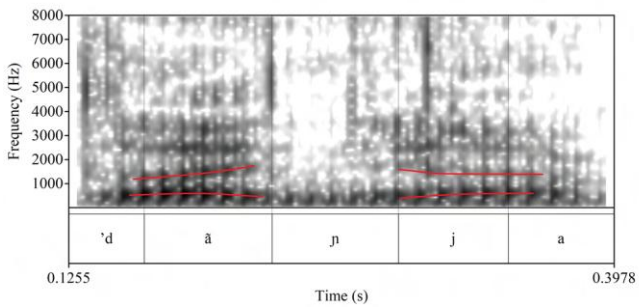

Fig. 4. The word "dania" (d) on a spectrogram with the course of the formants marked in red (speaker B)

Formants 3 and 4 are much less curved in Figure 4 than in Figure 3, but their course is very similar. In the case of the utterance of the word "dania" (d) (according to the context), it was possible to distinguish 5 phonemes by ear and on the spectrogram, which corresponds to the word "Dania" (D). The phoneme [j] in this case is not as clearly visible as in Figure 3, but it does occur, which suggests that the spoken word depicted in Figure 4 could be perceived in two ways.

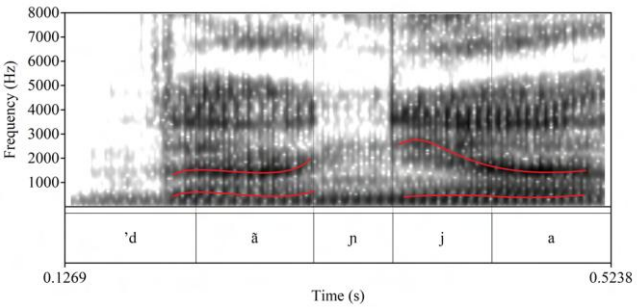

Fig. 5. The word "Dania" (D) on a spectrogram with the course of the formants marked in red (speaker G)

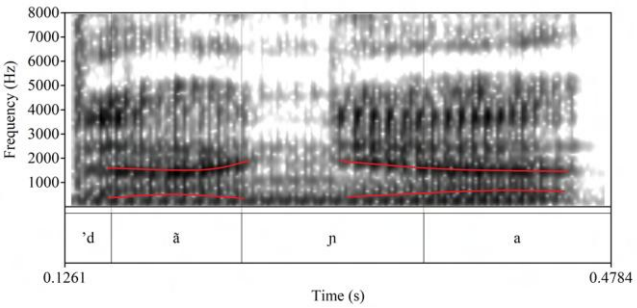

Fig. 6. The word "dania" (d) on a spectrogram with the course of the formants marked in red (speaker G)

The main difference between the words spoken by speaker G is that when articulating the word "dania" (d), the formants have a more stable course. This is especially evident in the transitions between the phonemes [ñja] and [ña]. The formants in Fig.6 are arranged flat, when in Fig.5 the outline of the phoneme [j] can be clearly seen.

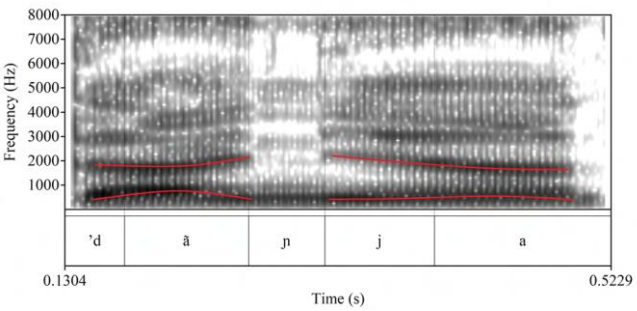

Fig. 7. The word "Dania" (D) on a spectrogram with the course of the formants marked in red (speaker J)

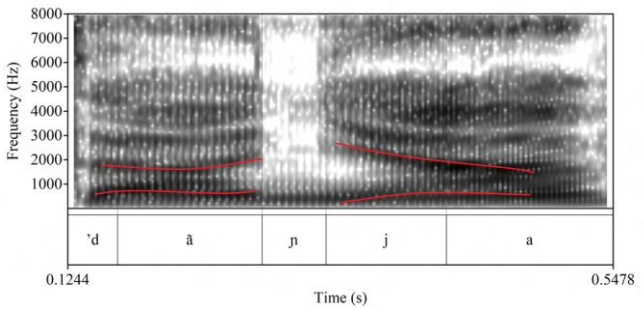

Fig. 8. The word "dania" (d) on a spectrogram with the course of the formants marked in red (speaker J)

In the case of speaker J, the authors were unable to distinguish the correct amount of phonemes for the word "Dania" (D) (Figure 7). Figure 8 shows an example of the wrong way of pronouncing the words "dania" (d). In the whole sentence from which it was cut would be perfectly understandable and read as dania (d). However, if we do not have context and only have the sample itself and its spectrogram image, the sample seems to be a typical example of the word Dania (D).

Analysis of the spectrograms revealed significant acoustic differences between the words “Dania” (D) and “dania (d). Key observations:

Differences in formants for the word "Dania" (country):

- the curvature of the controls, especially F2, was visible at the moment of passing through the phoneme [j].

- formants were arranged in ascending and descending lines, which indicates the presence of a glide.

Differences in controls in the case of the word "dania" (meals):

- formants were flatter, resembling straight lines.

- the absence of a clear phonemic transition [j], suggesting a simpler phonetic structure.

Intonation and acoustic energy:

- “Dania” (country) was often more strongly intoned, with higher energy in the 500–3000 Hz band.
- “dania” (meals) had a flat intonation and shorter duration of phonemes.

In addition, it was checked in how many cases it was possible to distinguish the number of phonemes by auditory and of spectrographic analysis. All samples were checked and the number of phonemes that could stand out was entered next to each one. Then, for each sample, it was checked whether the number of phonemes highlighted was consistent with the number of phonemes in the IPA phonetic notation of that word.

In 19 cases out of 119, it was not possible to distinguish the appropriate number of phonemes. For example, according to the IPA phonetic notation, there should be 5 distinguishable phonemes in a sound sample signed as Dania (D), but the authors distinguished only 4 of them on the spectrogram. In addition, 14 out of 19 samples undistinguished refer to the wrong amount of phonemes in the case of the occurrence of "Dania" (D), and only 5 others in the case of the occurrence of "dania" (d), which suggests that much more often it is "Dania" (D) as a country that can be misinterpreted as "dania" (d) as meals.

Of the nineteen samples previously identified as mispronounced (due to incorrect articulation or phoneme count), eight were included in the subjective test. Of these eight samples, six involved the word “Dania” (D), while the remaining two involved “dania” (d). As expected, these samples were more difficult for participants to recognize.

C. Automatic classification

An SVM classifier with an RBF kernel was implemented using MFCC features extracted from the audio samples. To explore the impact of train/test proportions, the full dataset (119 samples) was evaluated under three configurations:

– 80/20 — average accuracy: 76%,
– 70/30 — average accuracy: 79%,
– 60/40 — average accuracy: 75%.

Each configuration was tested in 25 runs, and mean accuracy was calculated to identify the most stable split. Based on these results, the 70/30 ratio was considered optimal.

In the final experiment, the test set was not randomly sampled but fixed to 40 samples (20 Dania, 20 dania) to match the subjective listening test. This design ensured direct comparability between human and machine performance. Although this proportion (~66/34) differs slightly from 70/30, it remains close to the configuration identified as optimal in the preliminary analysis, preserving methodological consistency.

The classifier achieved an overall accuracy of 75% on this test set. Detailed metrics were as follows:

– Dania: precision = 0.92, recall = 0.55, F1-score = 0.69 (support = 20),

– dania: precision = 0.68, recall = 0.95, F1-score = 0.79 (support = 20).

The confusion matrix is shown in Figure 9, illustrating that most errors involved misclassifying Dania as dania (9 cases), while the reverse occurred only once.

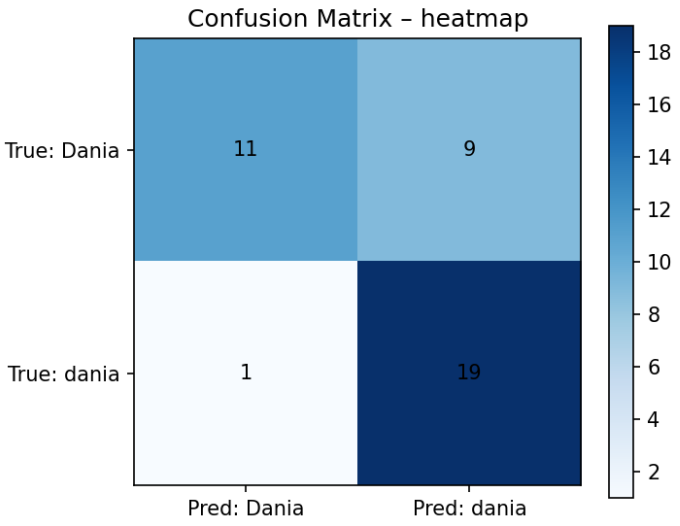

Fig. 9. Confusion matrix of the SVM classifier on the 40-sample test set

V. Discussion

Experiments confirm that recognizing homographs without context is a difficult task for humans. The average success rate of subjective test participants was 58%, indicating significant uncertainty in the samples. In the automatic classification experiment, the SVM model based on MFCC features achieved 75% accuracy on the 40-sample test set, which, although higher than human performance, demonstrates that phonetic ambiguity remains challenging even for machine learning models.

One of the key factors influencing the difficulty of classification is coarticulation, which is the smooth transitions between phonemes that can mask the presence of essential features, such as the palatal approximant [j] in the word "Dania" (country). Additionally, individual differences in articulation between speakers – resulting from fatigue, speech habits or accent – further complicate recognition.

Both humans and the model struggled with the same problematic samples, suggesting that some utterances lack sufficient acoustic cues for unambiguous classification. For example, certain recordings (e.g., samples with unclear [j] transitions) were misclassified by the model and also caused errors in the subjective test. In such cases, even spectrogram analysis did not allow for a clear determination of phoneme boundaries consistent with IPA notation.

It is worth noting that each of the approaches – subjective test, spectrogram analysis and automatic classification – can complement each other. Their combination increases the chance of correct identification, which is especially important in ASR systems operating without contextual information. Although the study focuses on isolated words, the proposed MFCC+SVM approach could serve as a supplementary module in ASR pipelines, particularly for short commands or keyword spotting where contextual information is unavailable.

Application potential:

- ASR systems for languages with high phonetic complexity (e.g. Polish),
- Speech-to-Text applications, where the correctness of recognition affects the quality of transcription,
- Voice interfaces in mobile and telecommunications devices, where context may be limited.

Conclusions

For the purposes of the paper, the focus was solely on the analysis of the differences between the homographs "Dania" (country) and "dania" (meals). The aim of the paper is primarily to draw attention to the problem of identifying words with similar sounds in the context of speech recognition systems and to propose solutions. These two homographs are just an example of a pair of words that can be a problem for automated systems and humans to recognize. Although the study focused on a single pair of homographs, the methodology can be extended to other phonetic ambiguities in Polish and other languages. This is particularly relevant in ASR systems where contextual information is unavailable. The problem can be extended to other words with similar sounds. In telecommunications, this problem could influence the decision to implement contextual speech recognition or to use exception handling mechanisms based on automatic analysis of acoustic signal parameters.

The analysis confirmed that the use of MFCC coefficients in combination with the SVM classifier is an effective method of recognizing difficult phonetic cases, such as homographs in Polish. The model proved to be more precise than human perception, even in cases where phonetic differences were minimal.

This study shows that the MFCC in combination with SVM can effectively resolve phonetic ambiguities in Polish homographs, surpassing human perception.

The automatic system showed the greatest agreement with the phonetic notation of the words "Dania" (country) and "dania" (meals), which makes it a promising tool in the context of speech recognition without access to contextual information. Even with a limited number of samples, the model achieved high performance, suggesting that its performance could increase with a larger dataset.

Recommendations for further research

- Expand the dataset with more samples and speaker diversity to improve model generalization.
- Integration of deep learning methods (e.g., neural networks, contextual models) to take into account both acoustic and semantic features.
- Analysis of other homograph pairs in Polish and other languages to assess the scalability of the proposed approach.
- Research on the impact of noise and recording conditions on the effectiveness of classification.

Acknowledgements

The authors of the work are very grateful to the participants of the research, both people who agreed to record their voices for the purposes of making a database of recordings, and subjective tests participants.

References

[1] Martin, A., MacDonald, R. L., Jiang, P.-P., et al. (2025). Project Euphonia: Advancing inclusive speech recognition. Frontiers in Language Sciences, 4, Article 1569448.
[2] Martínez-Celdrán, E. (2004). Problems in the classification of approximants. Journal of the International Phonetic Association, 34, 201–210.
[3] Wiktionary. Dania— Wiktionary,Free dictionary. [Online; accessed 2-grudzień-2024].2023. url: https://pl.wiktionary.org/w/index.php?title=Dania&oldid=8283315
[4] Wiktionary. danie— Wiktionary,Free dictionary. [Online; accessed 2-grudzień-2024]. 2024. url: https://pl.wiktionary.org/w/index.php?title=danie&oldid=8375551
[5] Crystal, D. (2008). A Dictionary of Linguistics and Phonetics. Wiley.
[6] Bar-On A, Dattner E, Braun-Peretz O. Resolving homography: The role of post-homograph context in reading aloud ambiguous sentences in Hebrew. Applied Psycholinguistics. 2019;40(6):1405-1420. https://doi.org/10.1017/S0142716419000316
[7] Rodd, Jennifer M., 'Lexical Ambiguity', in Shirley-Ann Rueschemeyer, and M. Gareth Gaskell (eds), The Oxford Handbook of Psycholinguistics, 2nd edn, Oxford Library of Psychology (2018; online edn, Oxford Academic, 10 Sept. 2018), https://doi.org/10.1093/oxfordhb/9780198786825.013.5 ,
[8] Nicolis, D., & Klimkov, M. (2021). Lightweight contextual classifier for homograph disambiguation.
[9] Rezáčková, M., et al. (2024). Homograph disambiguation in TTS using fine-tuned T5 transformer.

[10] Ittichaichareon, C., Suksri, S., & Yingthawornsuk, T. (2012). Speech recognition using MFCC.
[11] Dhingra, S. D., Nijhawan, G., & Pandit, P. (2013). Isolated Speech Recognition Using MFCC and DTW. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Energy, 2, 4085–4092.
[12] Hanifa, R., Isa, K., & Mohamad, S. (2021). A review on speaker recognition: Technology and challenges. Computers & Electrical Engineering, 90, 107005.
[13] Lathiyaa, R., et al. (2025). Speech Emotion Recognition Using MFCC and SVM Classification. Journal of Neonatal Surgery, 14(24s), 943–950.
[14] Banerjee, S., et al. (2025). Stuttering detection using MFCC and deep learning.
[15] Sonawane, A., & Inamdar, V. (2023). Spoofing detection in ASR using MFCC and CNN.
[16] Staroniewicz, Piotr. "Recognition of emotional state in Polish speech-comparison between human and automatic efficiency." European Workshop on Biometrics and Identity Management. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. https://doi.org/10.1007/978-3-642-04391-8_5
[17] Al Dujaili, Mohammed Jawad, Abbas Ebrahimi-Moghadam, and Ahmed Fatlawi. "Speech emotion recognition based on SVM and KNN classifications fusion." International Journal of Electrical and Computer Engineering 11.2 (2021): 1259. https://doi.org/10.11591/ijece.v11i2.pp1259-1264
[18] Singh, M. K., Kishore, D., & Kumar, R. A. (2024). Accent Recognition of Speech Signal Using MFCC-SVM and k-NN Technique. EVERGREEN Journal, 11(2), 305–312.
[19] Gorman, K., et al. (2023). Hybrid rule-based and SVM approach to homograph disambiguation.