The sound of emotions: Extracting the emotional content of handwriting sounds

Bartłomiej Nowak[1], David Issak-Zade[2], Ixchel Meza Solís[3], Karla Jazmín Ramírez Domínguez[3], Lennart Heepmann[4], Marcus Dancker[4], Ole Voß[4], Raphael Cassin[5], Seth Edelman[6], Vedant Janapaty[6]

Luxembourg[1], Israel[2], Mexico[3], Germany[4], UK[5], USA[6]

Dr. Vyacheslav Kalchenko

Senior Staff Scientist and Head of the Optical Imaging and Translational Bioengineering Unit at the Weizmann Institute of Science, Rehovot, Israel

Abstract

Expressive writing has demonstrated efficacy across various psychiatric applications, as evidenced by multiple studies. Crucial to its effectiveness is the complete privacy of the writing process, posing challenges for noninvasive evaluation. This study proposes an innovative approach by assessing the patient's emotional state through the acoustic properties of their writing. Prior research indicates that emotions influence the duration and frequency of writing pauses. Further exploration of the interplay between emotions and writing underscores the significance of these pauses in our analysis. This pilot study investigates the feasibility of employing several analytical and machine-learning techniques to discern writing activity. Additionally, an auxiliary AI model was developed to classify pause patterns according to distinct emotional parameters. Our findings highlight the promising potential of these methodologies for noninvasive emotional state evaluation.

Introduction

According to data from the World Mental Health Surveys, 3.9% of the world population suffers from post-traumatic stress disorder (PTSD) [6], which is characterized by symptoms such as intrusive memories, avoidance of topics or places that remind the individual of the event, and negative changes in thinking and mood. These psychological effects can also impact physical health, causing issues such as sleep problems, difficulty concentrating, and feelings of anger or shame, which can prevent individuals from living their lives normally. Therapists have identified three methods that show the best results in treating this condition:

cognitive behavioral therapy (CBT), eye movement desensitization and reprocessing (EMDR), and expressive writing.

The first study of expressive writing was carried out in 1986 by psychologists James W. Pennebaker and Sandra Klihr Beall. The test consisted of 15-20 minute sessions, where participants had to write about their deepest thoughts and feelings.[5] After 4 months of taking the tests, the participants reported significant benefits in objectively assessed and self-reported physical health, with less frequent visits to the health center and a trend towards fewer days out of role owing to illness.

Research conducted by Karen A. Baikie at the University of Cambridge has shown that practicing mindfulness can lead to improved mental and physical health, including decreased anxiety and blood pressure, improved liver function, and greater psychological well-being.[10] These benefits can also result in reduced absenteeism at work, improved memory, better self-confidence, and social life, and greater sporting performance. Additionally, mindfulness has been found to help people suffering from asthma, cancer, or postoperative pain by reducing their pain levels and the need for healthcare services.[9]

Neuroscience studies how writing affects our brain, activating parts such as the frontal lobe for reasoning and memory organization. Therapy pauses serve various purposes, from introspection to emotional expression, and neuroimaging shows brain regions associated with pauses.[12] Furthermore, neuroimaging studies have identified specific brain regions associated with pauses. In healthy individuals, pauses between clauses activate areas like the anterior superior temporal gyrus (STG) and the insula, which are linked to language processing and emotional regulation. In contrast, patients with psychiatric conditions show less activation in these areas, indicating a neural basis for their impaired pause usage and, consequently, their emotional expression and processing difficulties.[11]

In a study of 129 individuals with anxiety, depression, or stress, patients were asked to draw specific objects, and a computerized platform analyzed the data, identifying patterns in pauses. This allowed emotions to be detected by examining the duration and frequency of pauses, making the database a useful tool for assessing a patient’s mental state.[10]

Our goal was to determine the emotional state of a patient by analyzing the audio of the writing process using developed AI models. This allows for more personalized and less intrusive therapy, including analysis for doctors about the patient's emotional state and optimizing times, as well as processes involved in therapy.

Methods and results: Analytical part

To get information about the emotional state of a person writing, we used established analytical and statistical tools such as spectrograms, noise filtering techniques and autocorrelation. In the beginning, we used these tools to differentiate between writing and silence. Using more advanced tools to compare events in the audio data we have also been partially successful at detecting the difference between the sound of a writing stroke and other, equally loud noises.

We then applied these methods of detecting writing features to audio data of emotional handwriting. Thereby we wanted to find structural differences between the handwriting of positive and negative emotion that could be found with analytical tools.

Writing detection based on intensity-threshold

To collaboratively explore feature extraction techniques for handwriting analysis based on sound, we recorded a sample of *cursive* handwriting as "v.mp3". Initial waveform analysis revealed that writing periods are characterized by distinct abrupt fluctuations compared to the near-zero baseline representing non-writing periods (see fig.1.1).

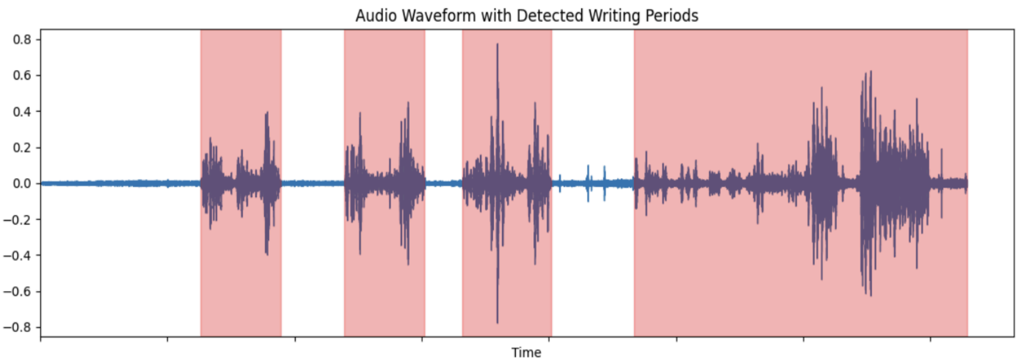


Figure 1.1. Audio Waveform of v.mp3 with highlighted detected writing periods

This observation has led us to our initial approach to writing detection based on energy threshold. To evaluate these fluctuations, we are observing the effective value of the signal by calculating the Root Mean Square of the signal using the Python library librosa [8]. This gave us a relatively stable baseline during non-writing periods, and distinct spikes during writing. Then, writing was detected by defining a silence threshold parameter that differentiates between silence and writing. (see fig 1.2)

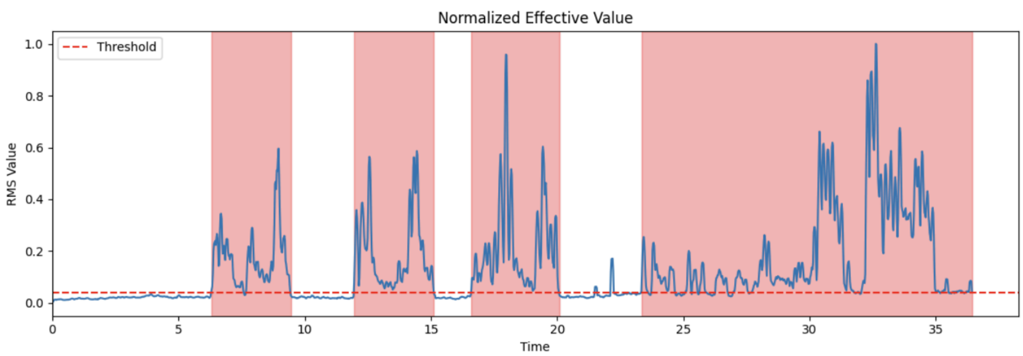


Figure 1.2. Normalized Effective Value extracted from v.mp3 recording file

Filtering and denoising methods:

Wavelet denoising

To reduce noise not only in the time domain, but also throughout different frequencies we used Wavelet denoising. Essentially, wavelet transforms decompose a signal into multiple levels, and then uses thresholding, which removes or attenuates small wavelet coefficients associated with noise. When reconstructing the signal on manipulated coefficients, it produces a signal where the observed signal components are more isolated and dominant [13].

Median filter

After applying the wavelet denoising, we implemented the median filter from the “SciPy” library to remove residual impulsive noise. For the filter, we used a kernel size of 5.

Our last measure was to normalize the data by dividing all the data points from the median filtered data by the datapoint from the array with the highest value. This was done in order to reduce the influence of physical properties of the recording setup such as the distance between microphone and pen/pencil or the sensibility of the microphone.

High-Pass Frequency Filter (HPFF)

HPFFs pass high frequency signals which cut lower frequency sounds below a certain threshold. Contrast this with low pass filters, which pass over high frequency sounds, maintaining only low-frequency noise.

Vector and Matrix scanning for stroke detection based on similarity assessment

To enhance the precision of writing stroke detection and reduce reliance on microphone sensitivity, we focused on analyzing spectral features rather than solely relying on sound intensity. Spectrograms, providing a time-frequency representation of the audio signal, effectively capture the unique envelope characteristics of pen strokes, which manifest as distinct, high-intensity bands across a wide frequency range. (see fig 2.1)

We propose a novel spectrogram-based and exemplar-driven approach for stroke detection in audio. Our method involves manually selecting an exemplar stroke and scanning the audio spectrogram for similar regions. We employed Euclidean distance, cosine similarity, and Dynamic Time Warping to quantify the similarity between the exemplar and spectrogram segments. By thresholding the similarity scores, we identified potential stroke occurrences. Our proposed method demonstrates significant improvements in stroke detection accuracy compared to traditional intensity-based approach.

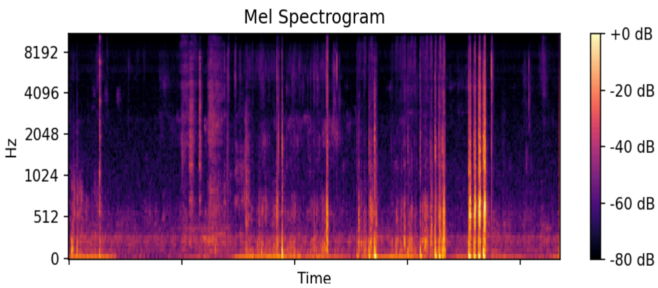


Figure 2.1: Spectrogram of a raw recording

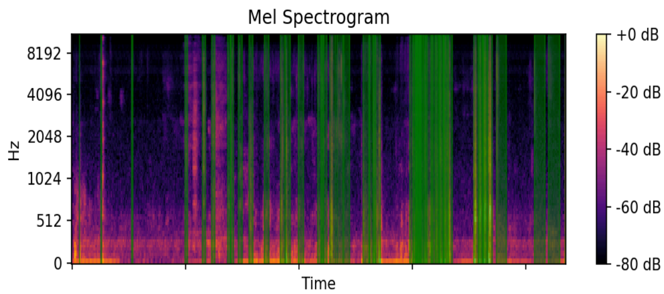
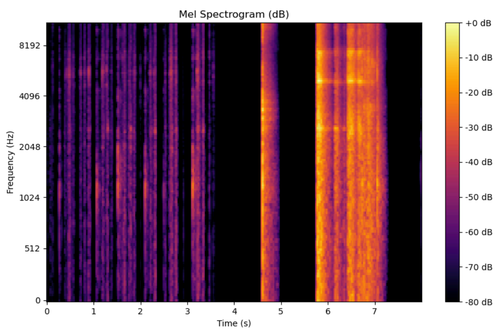


Figure 2.2: Denoising and Stroke Detection

Pause analysis by summing up the intensities of spectrograms

After filtering, denoising, and normalizing the audio data, we observed that the writing strokes in the spectrograms, generated using the `librosa.feature.melspectrogram` method, appeared as homogeneous vertical bars. In contrast, intervals without deliberate noise showed near-zero intensities across all frequencies. Consequently, we summed the intensities of all frequencies at each time point, creating a one-dimensional function of time that reflects the general sound intensity at any given moment.



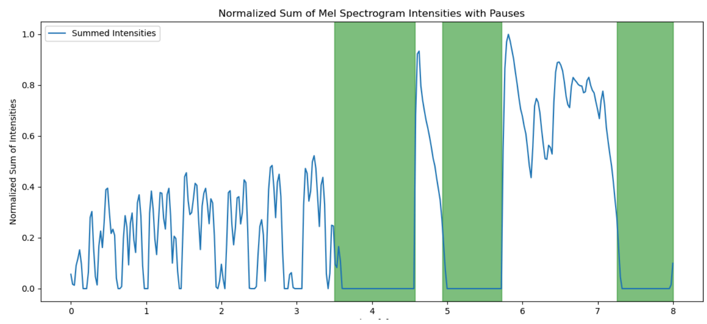


Figure 3.1: Spectrogram of example audio *3.2: Summed intensities of example audio with*
pauses

Figure 3.1 displays the mel spectrogram, and Figure 3.2 shows the summed intensities of an example recording featuring writing in the first half and loud noises in the second half. Different sounds correspond to spikes in the summed intensities plot.

Using these summed intensities, we developed a method to detect pauses by identifying intervals where values remained below a threshold for longer than a specified duration. The green bars in the right plot of Figure 3.2 indicate detected pauses with a minimum duration of 0.5 seconds.

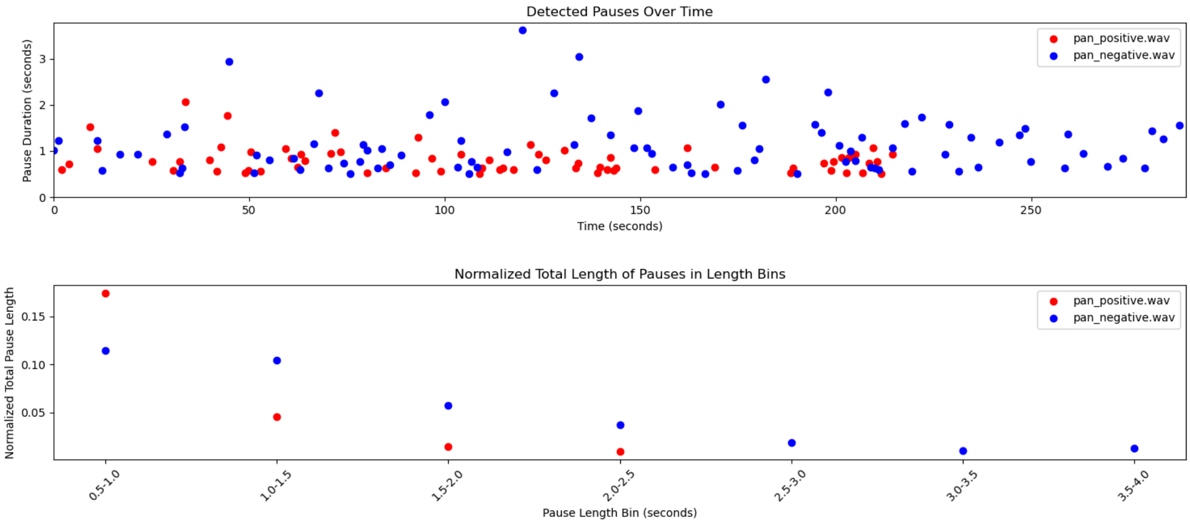


Figure 3.3: Pause analysis of negative and positive emotional handwriting

This pause detection method, with a minimum duration of 0.5 seconds, was applied to two writing files—one positive and one negative. Figure 3.3 presents the results, with red representing positive data and blue representing negative data. The upper plot shows each detected pause as a dot, with the x-position representing the start time and the y-position indicating the pause length. In the second plot, pauses were divided into length bins of 0.5-second intervals. The lengths of all pauses in each bin were summed and normalized over the total audio file length.

The plots reveal that negative writing had significantly more long pauses than positive writing. In bins with pauses longer than one second, negative writing had more pauses, with none of the positive writing pauses exceeding 2.5 seconds. The longest pause in negative writing exceeded 3.5 seconds. Only in the shortest bin (0.5 to 1 second) were there more pauses in positive writing.

We conclude that negative emotional writing has more and longer pauses than positive writing. This trend could be detected by a simple algorithm to support medical personnel.

Markov Chains

In essence, Markov chains are stochastic models where the present event is determined by a previous event and the next event is determined by the present event. In order to determine if a significant difference between writing and no writing periods could be determined by Markov chains, writing and no writing audio samples were recorded. The following table shows a sample of two Markov Chain probability matrices.

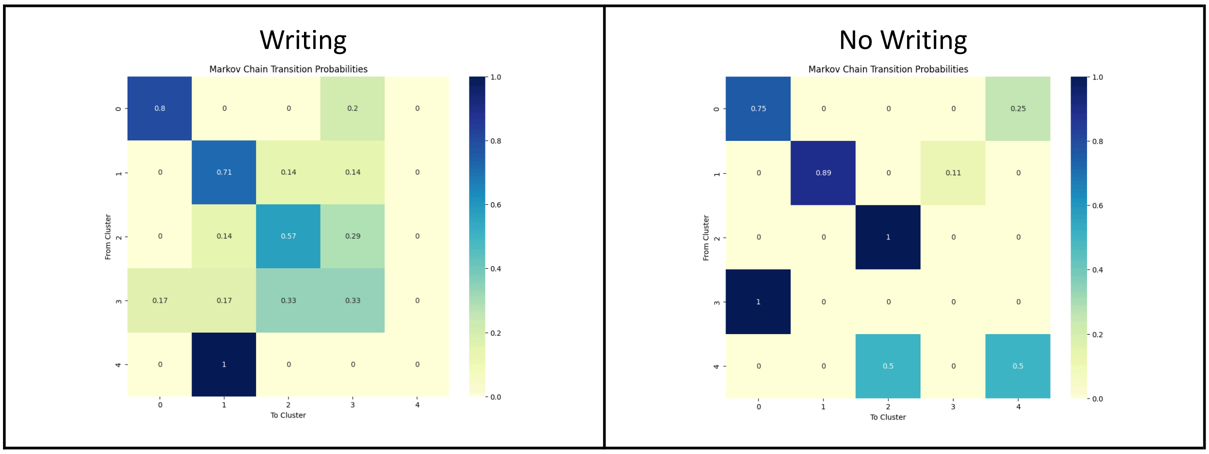


Figure 4.1 & 4.2: Markov Chain Transition Matrices for Writing and No Writing Audio Samples

Using the Markov Chain transition matrices, generated sequences for the audio files were created. The data is summarized in the following table.

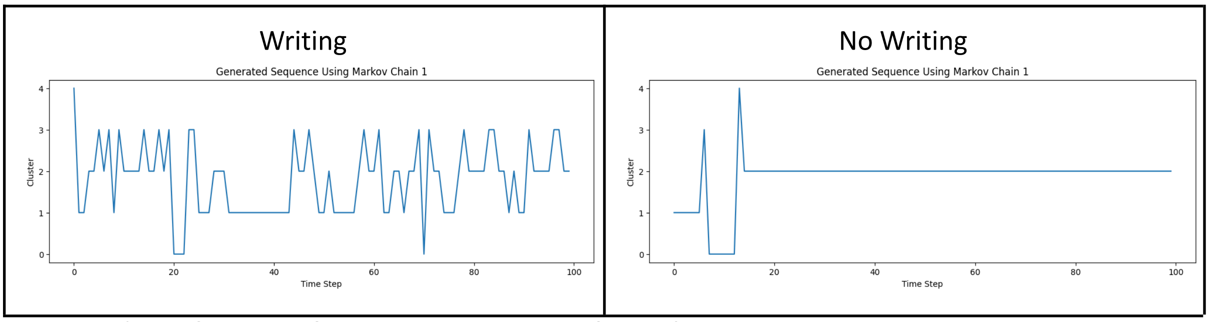


Figure 4.3 & 4.4: Generated Sequences Using Markov Chains for Writing and No Writing

Using the results of figure 4.3 and 4.4, it can be deduced that the non-writing generated sequences have flatter regions while the writing-generated sequence involves more variability and frequent changing of states.

Based on this data, it was predicted that with a negative emotion, the generated sequence graph would display more variability (since the frequency of strokes would be greater, causing greater shifts in the generated sequence of Markov Chains) than that of a neutral or positive emotion. The following table shows samples of data collected from negative, neutral, and positive emotion writing samples.

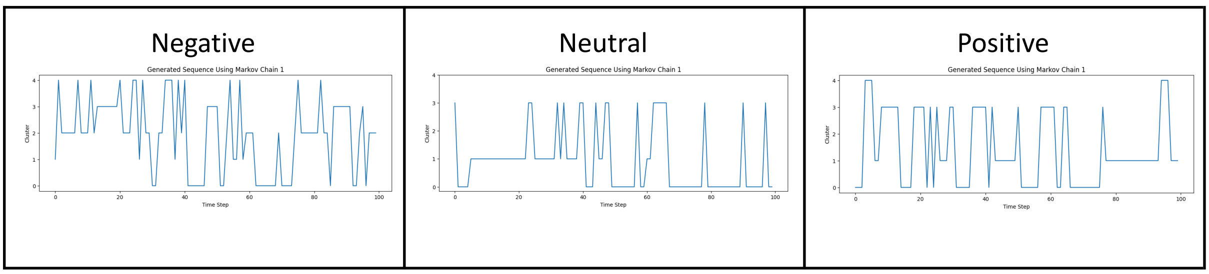


Figure 4.5: Generated Sequences Using Markov Chains for Negative, Neutral, and Positive Emotion Audio Samples

Methods and results: Machine learning part

In order to improve the analysis of writing sounds, automate complex processes, and distinguish more accurately between writing, pauses and background noises, we decided to implement artificial intelligence models to help us with this mission.
The main idea was to first detect the audio signals and their forms, using different methods to detect writing and non-writing processes, then we calculated the writing times and pauses in various samples so that once this was done and with the help of research in In the cognitive part, a conclusion could be obtained regarding the emotions found in the samples collected.

The models used are based on machine learning, so to train the neural networks used, all the members of the research collected samples of writing audio on their own using different devices. The small database created was approximately 200 elements in which audios with all kinds of characteristics were included, such as writing with background noises, audios without writing, audios of writing altering the emotional state of the person who wrote it, etc.

The models that provided the best results in terms of machine learning to detect between writing and non-writing were the following.

1. Energetic Approach:

Returning to the model used in the analytical part on the use of the intensity threshold for writing and non-writing detection, to improve the writing detection model using AI, we incorporated machine learning techniques to learn and generalize from labeled data. We used a convolutional neural network (CNN) for this purpose, as CNNs are well-suited for analyzing spectrograms of audio data. By using a CNN, we could automatically learn to detect these periods based on features extracted from the audio signal, which can improve accuracy. To implement this technique we needed the library TensorFlow/Keras to define, train, and evaluate deep learning models.
This method was not very effective, since the energy of writing through audio is not very exact, devices are needed that faithfully record the audio and that there are no background noises.

2. Hybrid approach:

Spectral analysis of strokes (see fig. 5.1) shows a distinctive spectral makeup: Along the frequency dimension, strokes are characterized by a broad uniform sound. Along the time dimension, the sound is quite stable and unchanging. These two features (broad spectrum and time stability) should be easy for a neural net to detect.

Hypothesis: A simple neural net should be able to recognize the characteristic spectrogram of a stroke.

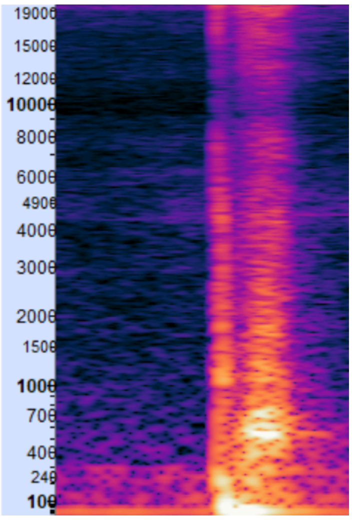


Figure 5.1: spectrogram of a pencil stroke

We implemented the spectrogram with a standard implementation of Mel-Frequency-Cepstral-Coefficients (MFCC) by Librosa [8]. The `librosa.feature.mfcc()` function uses a Fast Fourier Transform (FFT) to calculate the frequency makeup of a very short time window of the analyzed file. The frequencies are aggregated into overlapping bins. Lower frequencies were sorted into smaller bins so that smaller differences can be detected. For high frequencies the bins were wider, resulting in a lower resolution. This method is highly effective for voice recognition, where the number of bins is typically set to 13.

To quantify the stability of the signal over time, we also extracted the first and second-order derivatives of the MFCCs using Librosa's feature.delta() function.

To pinpoint the onset of a pause as precisely as possible, the MFCCs and deltas of just a single window were considered. Thus, the input for the neural network for n MFCC-bins (initially n = 13) consisted of 3*n values. We designed the hidden layers with varying sizes, where the first hidden layer is the largest and subsequent ones are half the size. The output was a binary classification into writing or not writing. We opted for Tensorflow-Keras to implement the neural network due to its intuitive structure, enabling fast model implementations.

To improve performance, multiple parameters were varied in training, resulting in different models. These parameters were the size of the first hidden layer, learning rate, number of training epochs, batch size, number of MFCC-bins and size of the FFT window. Performance was measured against a validation dataset. Care was taken that the validation dataset was completely independent from the training data to prevent any overfitting.

Model results: By systematically varying these parameters, we were able to achieve an accuracy of up to 97%.

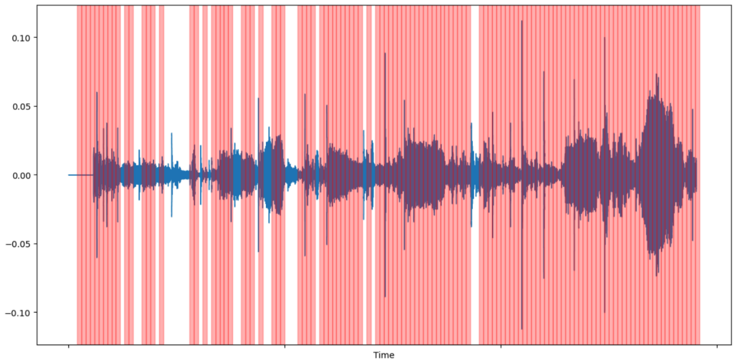


Figure 5.2: The hybrid model's analysis of an example recording. Periods the model detected as writing are colored red.

3. Transfer learning with YAMNet

Creating model dataset: Raw data for the dataset originates from audio samples generated at the ISSI, which are firstly classified as "writing" or "no writing". The model input samples are 0.96-second frames, taken from a random index from a random audio file. Rudimentary pre-processing was also applied to boost model performance [1]. To reduce noise a rolling window with a hop of 1 was used, and if the signal value was lower than the parameter "noise threshold" *(see Appendix A for parameter value)*, then the signal value was reduced to 0. As the final step min-max scaling was applied to the whole dataset. Described by this equation:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Where:

x: feature values in dataset

One important step in creating the dataset is the confirmation of the sample classification. Usually, samples coming from audio files labeled as "writing" would be automatically classified as such. However, it quickly became clear that this simple approach risked misclassifying pauses, which are inherent in any writing. These noisy labels would radically reduce model test accuracy by worsening the model's generalizability [2]. To limit the likelihood of such mistakes, any sample with a higher fraction of 0 (=noise) than the parameter “label threshold” *(see Appendix A for parameter value)* was classified as “no writing” even if it came from a “writing” audio file.

Pause and writing detection using YAMNet: The presence of pauses is a crucial signal of emotional state *(see chapter Methods: Cognitive science behind the expressive writing)*, therefore, creating a robust model for writing and pause detection was essential to our research.

Transfer learning, in other words, using a pre-trained model as a starting point for a second model, was the best approach for this task, due to the quite limited data in our possession and the availability of state-of-the-art audio detection models such as YAMNet [3].

YAMNet serves as a powerful feature extractor, on account of its audio embeddings, which are one of its 3 outputs [4]. The embedding vector is the average-pooled output that feeds to the final YAMNet classifier [4]. However, they can also be used as input to another classifier, as was done in this research.

We formed the embeddings from the training data *(see section Creating model dataset)*, which we then inputed to a dense neural network, acting as a binary classifier. Owing to the powerful base model, the classifier could remain small to reduce training time *(see Appendix B for model architecture and Appendix C for training parameters)*.

Model results: The model performance on clean training samples was superb, with an accuracy of 98.1%.

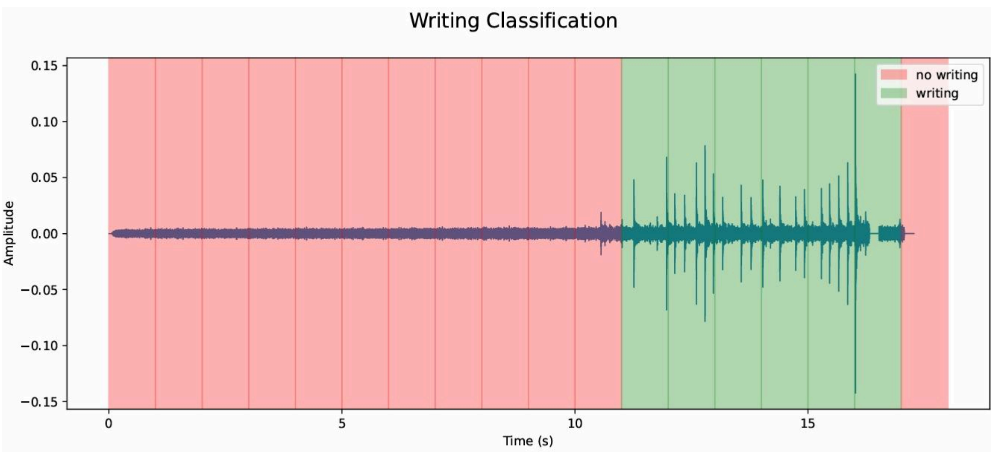


Figure 6.1: Classifier writing prediction on clean test data

Effect of speech on model results: To assess a potential limitation arising from the presence of speech, which could occur during a medical treatment, we tested our model on audio data with sections, where both speech and writing were present.

Speech was detected with YAMNet (no other classifier included), as it is one of the 512 classes included in the model [4].

The results of this test can be seen in Fig. 6.2 and Fig. 6.3. Particular attention has to be paid to the segment starting at 1:06 and ending at 1:14, where both speech and writing were present, but the model detected writing in only 50% of the sections (length 0.96 seconds). In sections where writing occurs on its own (e.g. 0:12 to 0:19, 0:35 to 0:45), the model performs substantially better.

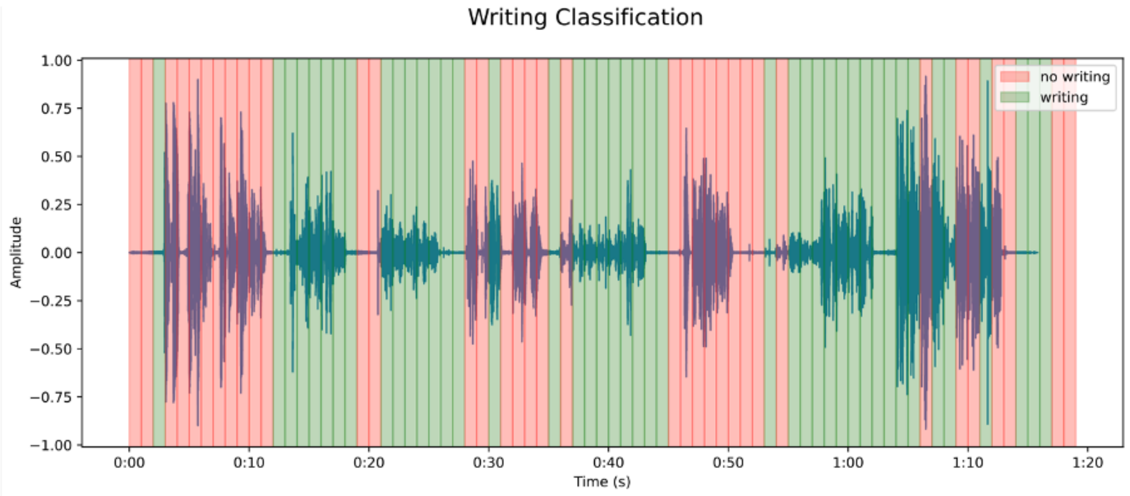


Figure 6.2: Classifier prediction for writing and no writing on mixed test data

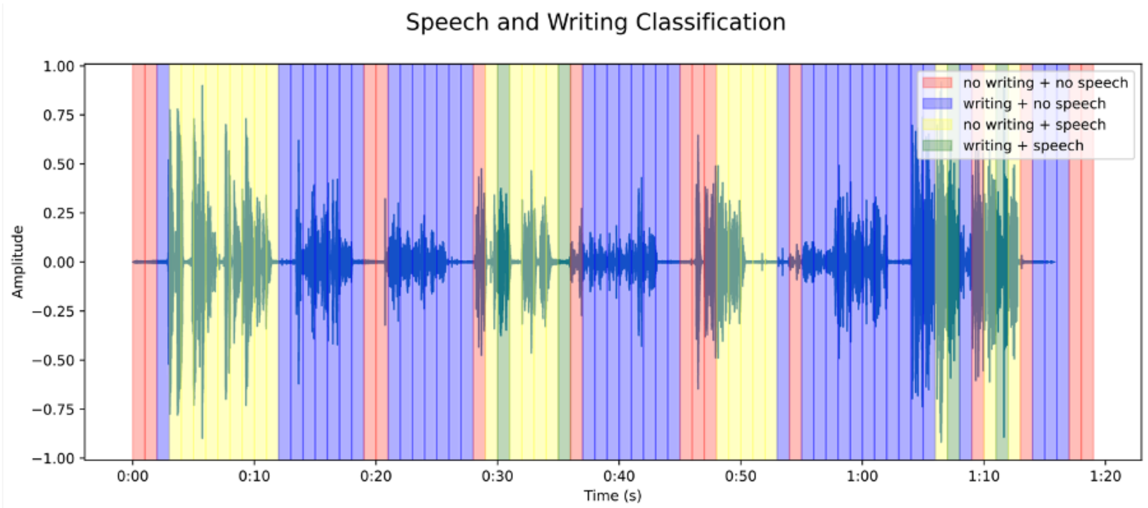


Figure 6.3: Classifier prediction combined with YAMNet speech prediction on mixed test data

Detecting Emotions

We assume a relationship between the length of the pauses in writing, the length of the time between the pauses with the current mood of the writer**.** A neural network seems very suitable to make this connection visible. Therefore, the goal was to develop and train a neural network that is able to detect the mood of the writer from a list of the length of pauses and a second list of the length of times between the pauses.

We have already developed methods to extract this data from the audio files and they have proven to be quite reliable.

Training data: The biggest obstacle to fulfilling that task was the training data for the neuronal network. Because we were not able to record and label the large amount of data needed for the training of our neural network, we sought out accessible training data. The EMOTHAW Database [7] was very promising for this. In that study, the handwriting of participants was recorded at 100 data points per second. In addition to the position and tilt of the pen, it included a parameter indicating whether the pen was touching the tablet or not. This enabled us to use an algorithm to bring the EMOTHAW data into the data structure of lists of pauses and intervals intended for our neural network. The study labeled their data using the DASS survey testing for depression, anxiety and stress. It found significant correlations between the length of pauses and said emotional states. This makes it possible to train a neural network to learn the association between writing pauses and these emotional states. The neural network

Our neural network received the two lists combined in a tensor as input. 512 pauses and intervals are intended in each case. A larger number would probably increase the complexity of the neural network too much, a much smaller number would probably not be enough for a significant decision. If the input file does not fulfill the specified number of pauses, it can be cropped or padded with zeros.

The neural network had a total of 7 dense layers where each layer has half the number of neurons of the previous layer. The output layer had 4 neurons for the intended mood assignments: depression, Anxiety, stress and none of them.

Model Results: We will be able to train the model as soon as we get access to the whole EMOTHAW database.

Discussion and Conclusion

This study presents a novel approach to optimizing expressive writing therapy for PTSD by utilizing audio analysis of the writing process. Preliminary findings show that sound recordings from standard devices might be able to detect handwriting characteristics and emotional states, as evidenced by the different analytical methods used and the AI models developed.

Our exploration of various methods to analyze the audio characteristics of handwriting has yielded promising results, though it also highlighted specific areas requiring further investigation. The initial approach of writing detection based on intensity-threshold provided valuable insights into the nature of handwriting sounds. However, its heavy reliance on microphone sensitivity and susceptibility to background noise significantly impacted its robustness and overall performance.

Vector and matrix scanning for stroke detection based on similarity assessment showed potential improvements in accuracy. However, addressing the impact of noise remains crucial. Further investigation into alternative similarity metrics and their possible combinations is recommended to achieve more precise stroke detection and better differentiation between writing and non-writing periods.

Pause analysis, achieved by summing the intensities of the spectrogram, showed promising trends. However, since the presented data consists of only two files with a total length of 510 seconds, further research on a larger dataset is needed to verify these trends and methods. Future research should also aim to determine a precise minimum pause duration or better intensity threshold for pause detection to improve the accuracy of emotional state assessment.

The hybrid approach and the approach using YAMNet demonstrated that, for samples with distinct and clear writing audio, the models proposed were satisfactory. It can be used, along with other methods, to help measure engagement and emotions when the data is free of substantial noise. This highlights the importance of clean audio samples for accurate emotional state detection.

This accessible and non-intrusive technique offers significant potential advantages, including real-time feedback for doctors about the patient's emotional involvement, optimizing processes in therapy. Additionally, this approach allows for the evaluation of the emotional meaning of the patient's writings and the formality of their interactions with the therapist, contributing to more personalized and effective treatment plans. The ability to carry out this method remotely would enable doctors to monitor and evaluate a patient's commitment and emotional involvement continuously and objectively, making expressive writing therapy more accessible for those who may not be able to attend face-to-face sessions regularly.

Our future plans are to improve the models we currently have and create a larger database of writing audio samples to better train our neural networks, ensuring that the results are mostly reliable and accurate. Moreover, addressing the issue of poor performance with files containing speech is crucial, as patients cannot be guaranteed to remain silent during therapy. This could be potentially achieved by including training data with speech or filtering it out during pre-processing. However, the strength of its signal could make the latter difficult. Our long-term goal is to develop an easy-to-use system that psychologists and doctors can implement to treat PTSD safely and reliably, thus obtaining better results in terms of the evaluation and treatment of this condition. Ultimately, we aim to generate a positive impact on society by providing an effective, accessible method for emotional state evaluation and therapeutic intervention.

In conclusion, while our study has demonstrated the potential of audio-based handwriting analysis for emotional state evaluation, further research is essential to address the identified limitations. Expanding the dataset, improving noise reduction techniques, and refining detection algorithms will be critical steps in advancing this technology. Future studies should focus on these areas to enhance the robustness and applicability of this approach in therapeutic settings, ultimately improving the effectiveness of expressive writing therapy and broader mental health interventions.

References

1. Nawi, N. M., Atomi, W. H., & Rehman, M. Z. (2014, January 29). *The effect of data pre-processing on optimized training of Artificial Neural Networks*. Procedia Technology. https://www.sciencedirect.com/science/article/pii/S2212017313003137

2. Song, H., Kim, M., Park, D., Shin, Y., & Lee, J.-G. (2022, March 10). *Learning from noisy labels with Deep Neural Networks: A survey*. arXiv.org.https://arxiv.org/abs/2007.08199

3. Drossos, K., Mimilakis, S. I., Gharib, S., Li, Y., & Virtanen, T. (2020, February 2). *Sound event detection with depthwise separable and dilated convolutions*. arXiv.org. https://arxiv.org/abs/2002.00476

4. *Google: Yamnet*. Kaggle. (n.d.). https://www.kaggle.com/models/google/yamnet/tensorFlow2/yamnet/1?tfhub-redirect=true

5. Baikie KA, Wilhelm K. Emotional and physical health benefits of expressive writing. *Advances in Psychiatric Treatment*. 2005;11(5):338-346. doi:10.1192/apt.11.5.338

6. Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., et al (2018, October 1). *Posttraumatic stress disorder in the World Mental Health Surveys*. Psychological medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6034513/

7. Likforman-Sulem, L., Esposito, A., Faundez-Zanuy, M., Clemencon, S., & Cordasco, G. (2017). EMOTHAW: a novel database for emotional state recognition from handwriting and drawing. *IEEE Transactions on Human-machine Systems*, *47*(2), 273–284. https://doi.org/10.1109/thms.2016.2635441

8. McFee, Brian et al. (2023). librosa/librosa: 0.10.1 https://doi.org/10.5281/zenodo.8252662

9. Katherine M. Krpan, Ethan Kross, ,et al. An everyday activity as a treatment for depression: The benefits of expressive writing for people diagnosed with major depressive disorder,Journal of Affective Disorders, Volume 150, Issue 3, 2013, Pages 1148-1151, ISSN 0165-0327,https://doi.org/10.1016/j.jad.2013.05.065.

10. Likforman-Sulem, L., Esposito, A., Faundez-Zanuy, M., Clemencon, S., & Cordasco, G. (2017). Emothaw: A novel database for emotional state recognition from handwriting and drawing. IEEE Transactions on Human-Machine Systems, 47(2), 273–284. https://doi.org/10.1109/thms.2016.2635441

11. Matsumoto, K., Kircher, Ti., Stokes, P., Brammer, M. J., LIddle, P., & McGuire, P. K. (2013, September 26). Frequency and neural correlates of pauses in patients with formal thought disorder. Frontiers. https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2013.00127/full

12. The State of Queensland; ou=Queensland Health; ou=; ou=; (2022, July 12). Brain Map Frontal Lobes. Queensland Health https://www.health.qld.gov.au/abios/asp/bfrontal#:~:text=The%20frontal%20lobes%20are%20important,order%20to%20achieve%20a%20goal.

13. Dautov, C. P., & Ozerdem, M. S. (2018, May). *(PDF) wavelet transform and signal denoising using wavelet method*. ResearchGate. https://www.researchgate.net/publication/326276573_Wavelet_transform_and_signal_denoising_using_Wavelet_method

Appendix

A: YAMNet dataset pre-processing parameters

Parameter name	Parameter value
Noise threshold	0.00075
Label threshold	0.75

B: Classifier Architecture

Parameter name	Parameter value
Input size	1024
Output size	2
Hidden layer count	4
Activation function (all layers)	ReLu
Hidden layer 1 neuron count	1024
Hidden layer 2 neuron count	512
Hidden layer 3 neuron count	256
Hidden layer 4 neuron count	128

C: Classifier training parameters

Parameter name	Parameter value
Learning rate	0.001
Epoch count	30
Batch size	64