Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 06, June-2023

Karoke Audio Exarction Using Matlab

Amrutha K R
Department of Electronics and Communication Engineering
SJCE, JSS ST&U

Puneeth K M
Department of Electronics and Communication Engineering
SJCE, JSS ST&U

Abstract— To isolate sound sources in an audio scene is called audio source separation. Singing voice extraction is one use for source separation. Using the 2D Fourier Transform (2DFT), we provide a unique method for music/voice separation in this study. Our method takes use of the way periodic patterns appear in the 2D Fourier Transform and is related to studies on biological auditory systems and picture processing. We discover that our technique is relatively easy to describe, put into practice, and is competitive with existing unsupervised source separation methods that make use of related presumptions.
Keywords : Automatic karaoke, foreground/background separation, image processing, 2DFT, audio source separation, singing voice extraction.

I. INTRODUCTION

Separating sound sources within an audio scene is known as audio source separation. The bass line in a musical mix, a single voice among a boisterous crowd, and the melody of the main vocal from a song are a few examples of source separation. Numerous applications might be found for the automatic division of auditory situations into significant sources (such as singing, drums, and accompaniment). These include karaoke [3], audio remixing [2], melody transcription, and instrument identification. Extraction of singing voices is one use for source separation. For singing voice extraction, a variety of techniques have been employed, the great majority of which use the spectrogram as the input representation. Examples include melodic smoothness restrictions on a source filter model [7], deep learning-based methods [6, 5], and nonnegative matrix factorization. Repetition is one of the most straightforward and reliable methods for singing voice extraction. Repetition is used in the spectrogram by REPETSIM [9], which locates comparable frames using the similarity matrix. Strong principal component analysis is used by Huang et al. [10] to distinguish between a sparse foreground (the singing voice) and a low-rank background (the accompaniment). The work that most nearly resembles ours is REPET [3], which distinguishes between a nonperiodic foreground (vocals) and a background with periodic repetition (accompaniment). In this study, we provide a unique, straightforward technique for discriminating between periodic and nonperiodic audio that makes use of the spectrogram's two-dimensional Fourier transform (2DFT). The features of the 2DFT allow us to distinguish between periodic and non-periodic audio without having to explicitly model periodic audio or determine the time of recurrence, both of which are necessary in REPET.

II. LITERATURE REVIEW

Recent studies demonstrate the widespread usage of music and accompanying applications. These are having a need for enhancements over the song's or music signal's instrumental portion. The goal is accomplished when the present concatenation of the speech and music allows for the successful iteration to be done over the signal breaking. One concept for a virtual stage-based karaoke system that is based on virtual reality was proposed by the authors ChangWhan Sul, Kee Chang Lee, et al. [4]. They presented and demonstrated virtual karaoke in this system, and they built a virtual stage on which the designated person will perform based on the gestures that were recognized by AI and camera captures. The user or participant first chooses a song from a local library or downloads one from the song server across the network. The General Musical Instrument Digital Interface (MIDI) file format is used to encode the Songs [4]. Additionally, it contains details on the architecture and geometry of items as well as the emotions and behaviors of virtual characters [4]. By doing so, the atmosphere in which virtual characters interact with actual characters will be simulated. For this, interactive response systems can be employed (the Microsoft Kinect gadget is now popular for developing these kinds of applications as well).

III. METHODOLOGY

In a music studio, a song is often recorded in stereo using two channels: left and right. The vocal component of the audio song is almost identical in both the left and right tracks when the artist is standing in the middle of the studio stage because his voice is simultaneously and equally fed into both of these channels; however, when the music of multiple instruments is recorded, the stereo sound output would have different musical components in the left and right channels. As a result, our method of removing the speech component requires first creating two mono sound files (left and right) by dividing a recorded stereo sound track into two channels. This yields the best performance possible. The voice of the original performer is significantly diminished, and if the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 06, June-2023

recorded music is adequately balanced, the voice is totally muted. We used the OOPS approach to construct the Karaoke Machine. Out of phase stereo is referred to as OOPS. Modern stereo recordings are processed using this approach to create a "new" third channel from the two original channels. In stereo recordings, it provides us with "hidden" sounds. Additionally, I have developed a GUI (Graphical User Interface) for working purposes in order to benefit the users. using the "guide" function, a unique built-in feature. The intended GUI looks like the figure below.

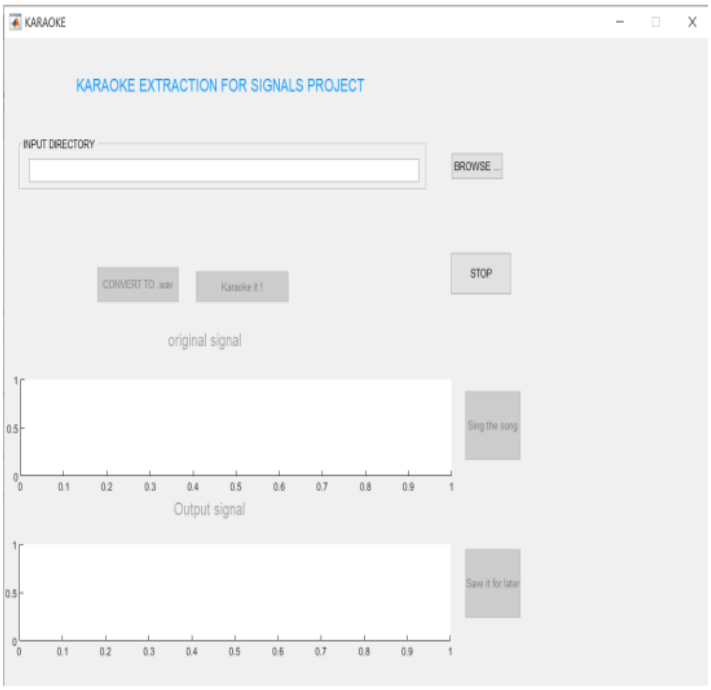

Fig. 1 GUI design of the model

IV. RESULT

The browse and stop buttons are initially active in the GUI window that appears when the program is executed; the other buttons are all code-disabled.

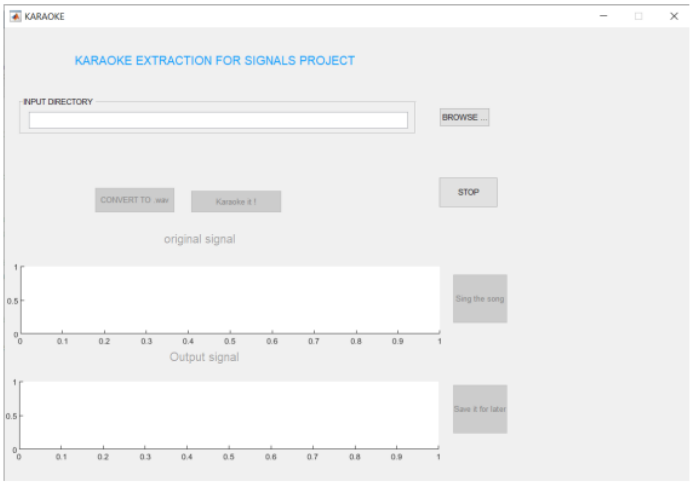

Fig. 2. The GUI is ready to function with browse button enabled

By placing the mouse in the explore bar and selecting the browse button, we may access the system's directories to retrieve the input signal. Two different forms of audio signal formats,.wav and.mp3, are available here. A different method to acquiring karaoke is taken if the supplied file is a.wav file. If the inserted file is in.mp3 format, a notification will appear on the screen instructing you to convert it to a.wav file, and the convert to.wav button will turn on. It is necessary to click that button in order to convert the MP3 file to a WAV format.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 06, June-2023

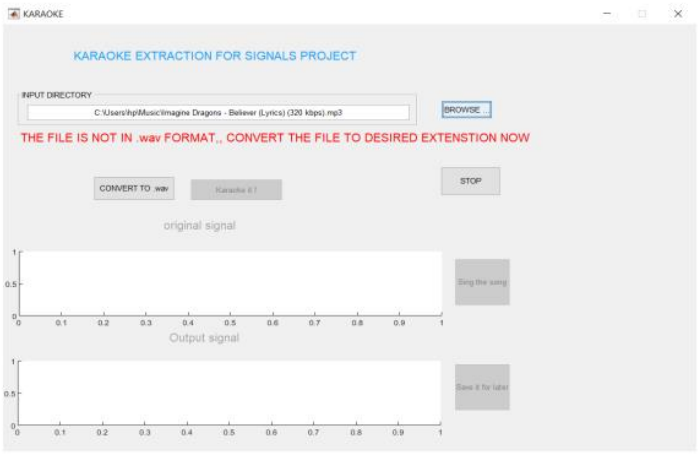

Fig.3. The file format doesn't match the input format

The.wav file will be saved in a system location after creation. So once more, have to browse to find the.wav file. The "KARAOKE it" button should be clicked once the.wav file has been inserted. The original signal's FFT will then be shown on the first axis board, and the karaoke output will be shown on the second axis board.

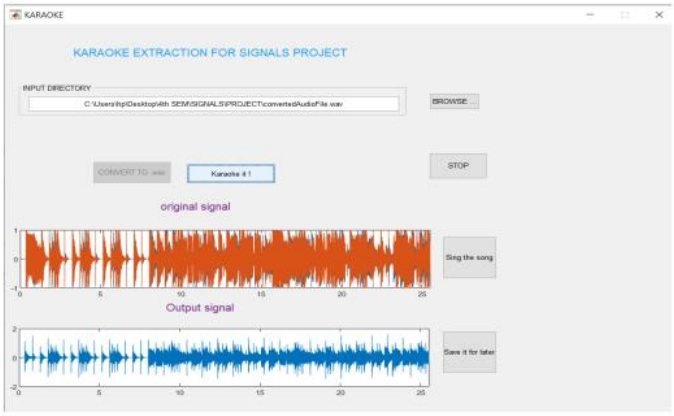

Fig.3. The corrected format is taken as input and output was produced

Additionally, there are two choices: one is to store the output for later use, and the other is to play the result. The only musical accompaniment will be the performed song. We utilize the "Stop" button to end the entire procedure and halt program execution.

V. CONCLUSION

The tendencies in the current situation indicate that retrieving musical knowledge is important. These days, there is a large need for musical karaoke processing tools and programs as well as digital music companions. In this essay, we attempted to visualize and identify an alternative aspect that is highly optimized and simple to utilize. We described a straightforward and original method for music/voice separation. Our method takes use of the way periodic patterns appear in the scalerate domain and is related to work in both image processing and biological hearing systems. We discover that our system can successfully separate unsupervised sources from supervised sources using comparable underlying assumptions.

REFERNECES

[1] "Automatic music transcription and audio source separation," Cybernetics &Systems, vol. 33, no. 6, pp. 603–627, 2002.
[2] J. F. Woodruff, B. Pardo, and R. B. Dannenberg, "Remixing stereo music with scoreinformed source separation.," in ISMIR, pp. 314–319, 2006.
[3] Z. Rafii and B. Pardo, "Repeating pattern extraction technique (repet): A simple method for music/voice separation," Audio, Speech, and Language Processing, IEEE Transactions on, vol. 21, no. 1, pp. 73–84, 2013.
[4] S. Uhlich, F. Giron, and Y. Mitsufuji, "Deep neural network-based instrument extraction from music," in Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on, pp. 2135– 2139, IEEE, 2015.
[5] J.-L. Durrieu, B. David, and G. Richard, "A musically motivated mid-level representation for pitch estimation and musical audio source separation," IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 6, pp. 1180–1191, 2011.
[6] A. Liutkus, D. Fitzgerald, Z. Rafii, B. Pardo, and L. Daudet, "Kernel additive models for source separation," Signal Processing, IEEE Transactions on, vol. 62, no. 16, pp. 4298–4310, 2014
[7] N. Monji and K. Ushida, "Making Karaoke Parties Lively by Reordering Songs Based on Pop Music Concert Program Data," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 6779-6780, doi: 10.1109/BigData55660.2022.10020370.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 06, June-2023

[8] K. Chen, S. Yu, C. -i. Wang, W. Li, T. Berg-Kirkpatrick and S. Dubnov, "Tonet: Tone-Octave Network for Singing Melody Extraction from Polyphonic Music," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 621-625, doi: 10.1109/ICASSP43922.2022.9747304.

[9] X. Gao, C. Gupta and H. Li, "PoLyScriber: Integrated Fine-Tuning of Extractor and Lyrics Transcriber for Polyphonic Music," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 1968-1981, 2023, doi: 10.1109/TASLP.2023.3275036.

[10] Qian, J., Liu, X., Yu, Y. *et al.* Stripe-Transformer: deep stripe feature learning for music source separation. *J AUDIO SPEECH MUSIC PROC.* **2023**, 2 (2023).

[11] Y. Wang, S. Tanaka, K. Yokoyama, H. -T. Wu and Y. Fang, "Karaoke Key Recommendation Via Personalized Competence-Based Rating Prediction," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 286-290, doi: 10.1109/ICASSP39728.2021.9414524.

[12] P. Gao, C. -Y. You and T. -S. Chi, "A Multi-Dilation and Multi-Resolution Fully Convolutional Network for Singing Melody Extraction," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 551-555, doi: 10.1109/ICASSP40776.2020.9053059.

[13] N. Monji and K. Ushida, "Making Karaoke Parties Lively by Reordering Songs Based on Pop Music Concert Program Data," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 6779-6780, doi: 10.1109/BigData55660.2022.10020370.

[14] Bittner, R.M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., Bello, J.P.: Medleydb: A multitrack dataset for annotation-intensive mir research. In: ISMIR. Volume 14. (2014) 155–160