

Speech Recognition and a Cymatics Based Configurable Speech Perception

S. M. Qaisar

College of Engineering, Effat University, Jeddah, KSA
sqaisar@effatuniversity.edu.sa

ABSTRACT

This paper propose an original approach of achieving a Cymatics based visual perception of isolated speech commands. The idea is to smartly combine the effective speech processing and analysis methods with the phenomena of Cymatics. In this context, an effective approach for automatic isolated speech based message recognition is proposed. The incoming speech segment is enhanced by applying the appropriate pre-emphasis filtering, noise thresholding and zero alignment operations. The Mel-Frequency Cepstral coefficients (MFCCs), Delta coefficients and Delta-Delta coefficients are extracted from the enhanced speech segment. Later on, the Dynamic Time Warping (DTW) technique is employed to compare these extracted features with the reference templates. The comparison outcomes are used to make the classification decision. The classification decision is transformed into a methodical excitation. Finally, this excitation is converted into the systematic visual perceptions via the phenomenon of Cymatics. The system functionality is tested with an experimental setup and results are presented. The approach is novel and can be employed in various applications like visual art, encryption, education, archeology, architecture, integration of impaired people, etc.

Key Words: Cymatics, Speech recognition, Mel-Frequency Cepstral Coefficients (MFCC), Dynamic time warping (DTW), Chladni plates.

1. INTRODUCTION

This work focus on realizing a systematic visual perception of the isolated speech words. Human beings naturally communicate via speech. Speech not only carries the message to convey but it also possess an important information about the speaker.

The recent technological advancements have revolutionized the speech based human machine interaction [14, 17]. It has diverse applications like smart cities, secure access, rehabilitation centres, criminals investigation analysis, banks, hands free assistance, commanding a robot, data security, virtual reality, Multimedia searches, auto-attendants, travel Information and reservation, translators, natural language understanding, etc. [15, 16, 19-24]. These systems are founded on the machine based speech comprehension and recognition. The principle is to achieve an appropriate extraction of phonetic features form the incoming speech. Later on, these extracted parameters are employed to classify the incoming speech. It allows realizing effective and robust machine based speech recognition systems [25-27].

Cymatics is the study of visualizing the effect of acoustic waves [1]. The objective of Cymatics is to make the sound waves and vibrations observable in order to take advantage of the human sense of sight, as it is the most discriminating human sense. The term “Cymatics” was first used by Hans Jenny to describe effects of sound and vibration on materials. Hans Jenny was called the father of Cymatics as he published the book “Cymatics: The study of wave phenomena and vibration” [4]. He has conducted different Cymatics experiments and invented the Tonoscope [2]. A tonoscope is a device that employs the sound waves, generated by the speaker, to vibrate its diaphragm. Sprinkling discrete substances, like powder, sand, etc., on the surface of the diaphragm helps in recognizing the nature of these vibrations as the beads try to avoid those areas that are vibrating most and collect in those areas that are vibrating least [3, 4, 5, 12].

The phenomena of Cymatics has fascinated scientists from different domains by the patterns, of artworks like appearances, acoustic vibrations can produce [2, 4]. Recently, it is employed as an investigation and design tool in various domains like Archaeology, Architecture, Music, etc. [49-51].

This work aims to achieve a Cymatics based systematic visual perception of the recognized isolated speech words. It can be realized by smartly combining the automatic speech recognition module with the phenomena of Cymatics. In this context, the proposed system smartly employs the maturity and robustness of modern speech processing and recognition techniques with the interesting features of the phenomena of Cymatics. The Section 2, describes the employed materials and methods. Section 3, describes the system prototype implementation. A case study and experimentation results are discussed in Section 4. Section 5, finally concludes the article.

2. MATERIALS AND METHODS

The principle of suggested system is depicted with the block diagram, shown on Figure 1. It shows that the system enhances the incoming speech signal and extract its features. These features are employed to classify the input speech. The classification decision is employed to generate a methodical excitation. Later on, this excitation is converted into sound and passed to the amplification stage. The augmented sound is transferred to the front-end Cymatics module for a systematic display. The different system modules are described in the following sub sections.

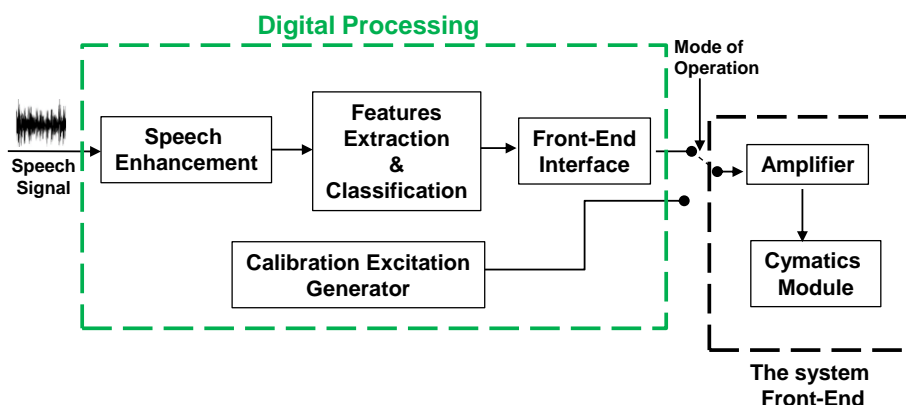


Fig. 1: The suggested system block diagram.

2.1. The Speech Processing, Analysis and Recognition Tools

The recent technological developments have revolutionized the speech processing, analysis and classification approaches. These approaches are quite mature and robust [14, 17, 25-27]. Therefore, they are extensively used for automatic speech and speaker recognition and identification. Certain applications are rehabilitation centres, biometrics, criminals' investigation analysis, banks, hands free assistance, database security, etc. [15, 16, 19-24, 28-29].

These systems work on the principle of achieving an appropriate extraction of phonetic features from the intended speech. Later on, these extracted parameters are employed to classify the intended speech. The Linear Predictive Coding (LPC) is widely used in this context [18]. However, this model does not integrate the human auditory system features. This shortcoming is treated up to a certain extent in algorithms like Linear Predictive Cepstral Coefficients (LPCC), Perceptual Linear Prediction (PLP), Mel-Frequency Cepstral Coefficients (MFCC) and Neural Predictive Coding (NPC) [23-27, 30-33].

The Mel-Frequency Cepstral (MFC) is a well-known method for extracting the speech signal features. A smart combination of the MFC and the Dynamic Time Warping (DTW) techniques can provide effective solutions especially in the case of isolated speech words recognition [23, 26, 32]. The devised system deals with the isolated speech word. Therefore, the MFC is employed for the intended speech features extraction and DTW is employed for the classification purpose. The employed speech processing, features extraction and classification chain is shown on Figure 2.

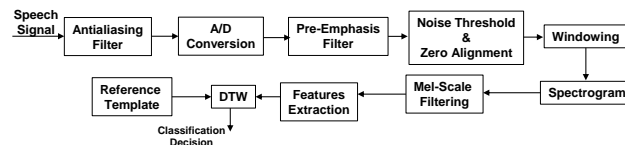


Fig. 2: The speech processing and analysis chain.

Figure 2, shows that the analog speech signal which is obtained at the output of microphone (cf. Figure 3) is passed through the antialiasing filter with a cut-off frequency of 3.5kHz. The bandlimited signal, obtained at the output of antialiasing filter is passed to the Analog to Digital (A/D) converter. It samples it at a rate of 16 kHz and each sample is represented with a 16-Bits precision.

The digitized signal is passed through the pre-emphasis filter. It provides the n^{th} output sample by subtracting the n^{th} input sample from α scaled $n-1$ th input sample. Here, α is presenting the pre-emphasis factor and its choice is application dependent. The pre-emphasis filter possess a high pass response. Therefore, it blocks the input signal DC offset which can significantly degrade the performance of post features extraction and classification modules. Moreover, it also flattens the signal spectra. It makes the signal more robust against the post processing truncation effects and improves the precision and performance of classification by minimizing the distance between the signal under test and its corresponding reference template.

A noise threshold is applied on the output of pre-emphasis filter. It removes the low amplitude noise and improves the post processing and analysis modules performance [36]. After the application of noise-thresholding, the intended speech segment is zero aligned. It improves the performance of post pattern recognition module by reducing the distance between the segment under test and its corresponding reference [36].

In practice, the hardware resources like processing speed, memory, logic resources, etc. are limited. Therefore, digital signal processing and analysis have to be performed on a finite set of data. In order to capture a finite frame of data the time windowing functions are employed [34]. It not only allows respecting the system constraints but also pro-vides a spectral leakage reduction mechanism [34]. The spectral leakage can tremendously degrades the performance of post signal processing, features extraction and classification modules.

In this context, the zero aligned signal is windowed. A variety of smoothening window functions is available like Hamming, Hanning, Blackman, Kaiser, etc. The choice of an appropriate window function and its length is application dependent [34]. Normally the windowing process is applied in a sliding mode fashion.

The windowed speech segments are transformed from time to frequency domain. It is achieved by employing an appropriate binary weighted length Fast Fourier Transform (FFT) [34]. It allows exploiting the signal spectra for extracting valuable information [35].

The spectrogram of the intended speech signal is computed by using the windowed segments spectrum. The Mel-Scale filtering is applied on each windowed section linear scale spectra. The Mel-Scale filtering is based on the human auditory system characteristics. This module is composed of a filters bank. Like human auditory system, it focus more on the frequencies up to 1kHz and jumps logarithmically above 1kHz. It is realized by employing narrow spaced triangular filters below 1kHz and logarithmically spaced triangular filters above 1kHz [36, 37]. It provides a precise information of energy, exists at lower frequencies, and the precession reduces for higher frequencies. The focus is on estimating the energy at concerned frequency regions. Let a set of P Mel

filters is employed. Then in order to calculate the filter bank energies each filter is multiplied with the signal spectrum. Afterwards the obtained filtered spectral components are added up. In this way P values are obtained which provides an indication of each filter energy. The spectrum, obtained after multiplying the signal spectrum with Mel filters bank is known as the Mel spectrum.

The Discrete Cosine Transform (DCT) is applied to convert the Mel spectrum into MFCCs. It reaches the aim of regulating the speech signal into minimal dimensions coefficients [36, 37]. In speech recognition, higher order MFCCs are more susceptible to the influence of noise than the lower order MFCCs. Therefore, for feature matching, only the lower half of the P coefficients is kept.

The MFCCs only present the static characteristics of the intended speech word. However, the human auditory system sensitivity is higher for the dynamic characteristics of sound [24, 26]. The intended word dynamic features can be obtained by derivation of MFCCs. Following this approach, the Delta and Delta-Delta coefficients are computed. The Delta coefficients are obtained by computing the first order derivatives of MFCCs. The Delta-Delta coefficients are computed as the second order derivatives of MFCCs. It increases the dimension of the intended speech segment feature vector and thereby increases the recognition accuracy.

The considered MFCCs, Delta and Delta-Delta coefficients are employed for the features matching. The DTW algorithm is employed for the intended speech recognition. It compares computed coefficients of the speech segment under test with the reference template. . In this way, the word under test is classified by finding the least distance between features of word under test and the reference templates [36].

2.2. *The Front-End Interface and Calibration Generator*

The suggested system functions in two different modes. These are the operation mode and the calibration mode. The choice of mode is made by the user and it is achieved via a software-based switch.

During operation mode, the front-end interface provides a liaison between the speech processing, analysis and classification unit and the system front-end. It converts the recognized word into a methodical sound via a predefined lookup table. For each considered speech word a specific sound is generated. This lookup table is designed as a function of employed display medium and actuator.

In calibration mode, the calibration excitation generator is used to generate a sequence of monotone excitations. Amplitudes and frequencies of these excitations are systematically varied within a predefined range. The choice of this range is a function of targeted application and employed system front-end module characteristics. These excitations are converted into sounds and are applied to the front-end module. In this way, excitations which result in distinguishable and repeatable Chladni patterns are registered in an excitation lookup table.

2.3. *The Front-End Module*

This module composes of an amplifier and a Cymatics module. The Cymatics module consists of an actuator and a display medium.

The amplifier adapts the excitation sound magnitude, generated by the front-end interface. It is done in order to adapt it to the employed actuator acceptable level.

The actuator transforms the augmented excitation acoustic waves into vibrations. These vibrations are employed to excite the display media. It transforms the excitation, generated by actuators, in a visual perception. A unique and repeatable visual pattern is generated corresponding to each specific excitation. In the studied case, the metallic Chladni plates are sand particles are employed as the display media.

According to Chladni's law, for fixed centre flat surfaces, a relationship between the frequency of mode of vibration and complexity of the resulted pattern can be formally expressed by Equation 1 [6]. Where, f_b is the frequency of mode of vibration. C and b are coefficients which depend on the type of surface material. m and n are respectively representing the number of linear and radial nodes. Equation 1, shows that there is a proportional relationship between the frequency and number of both radial and linear nodes. Therefore, when the frequency is increased the number of both nodes increase which results in more complex generated patterns.

$$f_b = C(m + 2n)^b \quad (1)$$

It concludes that for a systematic excitation the generated Cymatics pattern is a function of the employed medium shape, its properties and the excitation characteristics. Therefore, the pattern Q over the given display medium can be represented by Equation 2.

$$Q = F(A, f, S, O_x) \quad (2)$$

In Equation 9, A and f are respectively presenting the acoustic wave excitation amplitude and frequency. S is presenting the display media shape. O_x is presenting the display media other properties like volume, bulk modulus, density, etc.

3. PROTOTYPE IMPLEMENTATION

A system prototype is realized in order to exhibit the appealing features of suggested system. The commercially available appropriate components are employed to achieve an expeditious prototype development. The system experimentation setup is shown on Figure 3.

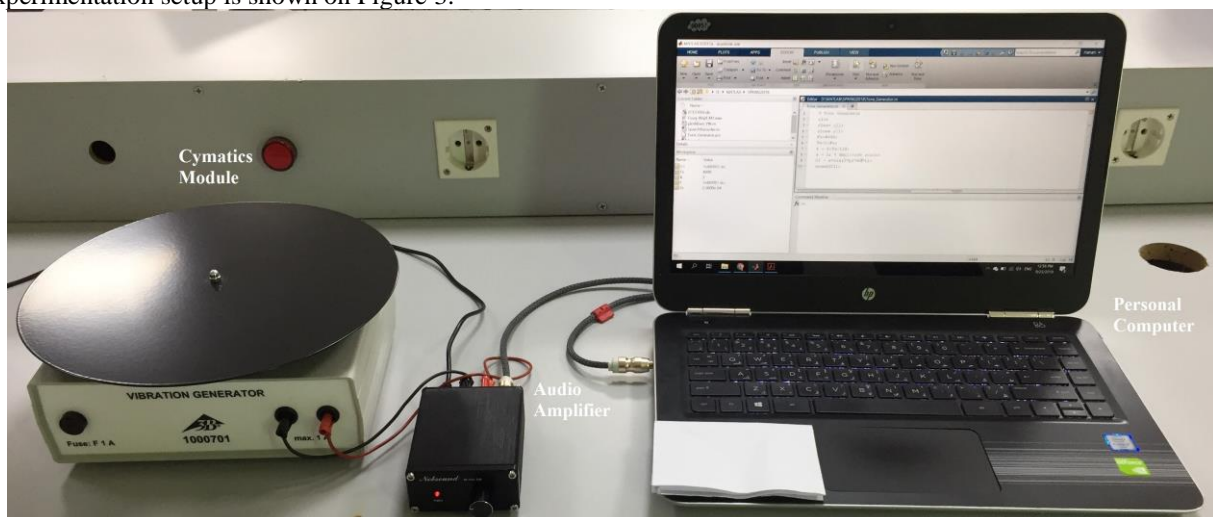


Fig. 5: The system experimentation setup.

The bandwidth of the speech signal is limited with a second order Chebyshev low pass filter. It acts as an antialiasing filter. Its cut-off frequency is chosen equal to 3.5kHz. The filtered output is digitized with the personal computer (PC) sound card. The sampling rate is chosen equal to 16 kHz. Each sample is quantized with a resolution of 16-Bits. The antialiasing filter acts as a bridge between microphone and the PC. It is implemented on a PCB with audio jack input and output.

The digitized signal is pre-emphasized. $\alpha=0.98$ is chosen. It allows the pre-emphasis filter to attenuate the input signal low frequency components more than the high frequency ones. It achieves the goal of spectrum flattening and improves the robustness of signal against the post processing truncation effects. The noise threshold and zero alignment is performed on the output of pre-emphasis filter. Later on the signal finite time framing is performed by employing the Hamming window.

The windowed signal spectrum is computed by applying the 512 points FFT. The spectrum energies are

computed and the redundant information is removed by applying the Mel filters bank. The filters bank is designed for a frequency band between 300Hz to 3500Hz. It composes of 24 filters. After the application of DCT and the choice of lower half of the obtained coefficients, it results into 16 MFCCs. These MFCCs are employed to compute Delta coefficients and later on, these Delta coefficients are used to compute Delta-Delta coefficients. The MFCCs, Delta coefficients and Delta-Delta are onwards employed for the intended word recognition.

The classification decision is employed to generate a methodical excitation. A specific excitation is generated for each identified speech word. It is achieved by employing a lookup table, designed during the calibration process (cf. Section 2.2).

The speech processing, analysis and classification chain along with the front-end interface and the calibration excitation generator are realized with the MATLAB based specifically developed applications [43].

The excitation, generated by the front-end interface is converted into sound by using the PC built in speaker. This sound signal is further augmented with the amplifier module. It is done in order to make the excitation compatible with the actuator module. A 100-watt audio amplifier is employed in this context [39]. It possess a Signal to Noise Ratio (SNR) of 103dB and frequency range between 20Hz to 20kHz. The interface between PC and amplifier is realized with an audio cable with male audio jacks on both sides [42].

The actuator is realized with a SF-9324 wave driver [40]. The amplified excitation is applied to SF-9324. The interface between amplifier and wave driver is realized with an audio cable with audio jack on one end and the crocodile connectors on the other end. The SF-9324 frequency range is between 0.1Hz to 5kHz. It can accept a maximum input of 6V. The maximum input current is limited to 1 ampere. The wave driver transforms the input sound waves into vibrations. Amplitudes and frequencies of these vibrations are directly proportional to the amplitude and frequency of input sound waves. The wave driver actuates the Chladni Plate that is employed as a display medium [41]. A 24 cm diameter circular metallic Chladni plate is employed in the studied case.

4. RESULTS AND DISCUSSION

In order to confirm and demonstrate the proposed system functionality, a case study is conducted. In this context an isolated speech words database is used. The employed database is composed of 3 auditory commands: Forward, Reverse and Center. These are pronounced by 10 different speakers including 5 males and 5 females. The recording is carried out in a noise controlled environment. Each speaker uttered every command 30 times. Therefore, in total 900 recordings are collected.


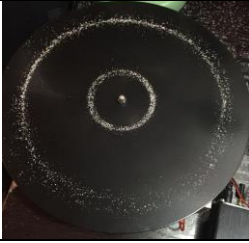

Each digitized word is processed and analyzed in order to extract its features. Later on, these features are employed for the classification purpose.

In the studied case, 900 sound waves are recorded for 3 considered commands. The 600 utterances from this database are employed for training and the 300 remaining utterances are used for testing.

The obtained average speaker dependent percentage recognition accuracy for all intended audio commands is 98%. The obtained average speaker independent percentage recognition accuracy for all commands is 95%. It shows that the average speaker dependent recognition accuracy is 3% higher than the average speaker independent recognition accuracy.

The classification decision made for the incoming speech segment by using the proposed speech processing and analysis chain, is employed to generate a methodical excitation. A specific monotone excitation of 0.5V amplitude is generated for each identified speech word. It is done by employing a lookup table, designed during the calibration process. The process is clear from Table 1.

Table 1: Summary of speech commands, their respective acoustic excitation features and Chladni patterns.

Speech Command	Acoustic Excitation Frequency (Hz)	Chladni Pattern
Forward	116	
Reverse	455	
Center	597.8	

The excitation, generated by the front end interface, is converted into sound. This sound signal is ten times augmented with the amplifier. The augmented excitation is applied to the wave driver. The wave driver actuates the Chladni plate via vibrations. The plate vibrates and in this way a repeatable and specific pattern is displayed on its surface by using the sand particles.

It allows achieving a systematic perception of each intended isolated speech word. The system is configurable as it generates a specific pattern for each intended word on the same display medium. It is realized by smartly combining an effective isolated speech word recognition process along with the physical phenomenon of Cymatics.

The results, discussed, show that the proposed speech enhancement and recognition module outperforms in terms of the speaker independent recognition accuracy, as compared to the existing state of the art solutions [26, 37, 47, 52]. Moreover, the speaker dependent recognition accuracy, achieved by the proposed solution is also comparable with the state of the art solutions. The main limitations in the performance of suggested method are the employed speech acquisition system limited precession. It introduce artefacts in the acquired speech signal and reduces its quality. A better performance can be achieved by employing a higher precision speech acquisition system.

Table 1, shows that the devised system achieves a methodical and specific visual perception of each intended isolated word. The system is configurable as it generates a specific pattern for each considered isolated auditory command on the same display medium. It is realized by smartly combining an effective isolated speech recognition method with a methodical acoustic excitation conversion and the physical phenomenon of Cymatics. The obtained results have demonstrated the Chladni's law [6]. They present a directly proportional relationship between the excitation frequency and the number of both radial and linear nodes, generated on the square metallic Chladni plate. From Table 1, it is clear that the complexity of generated patterns is increased with the

increase in acoustic excitation frequency and vice versa.

Preliminary, a case of only five auditory commands is considered. However, studies are extendable in order to find out specific symbols for an extensive vocabulary of isolated words.

The extension of proposed system study towards different material and shape mediums is a future work. It can result into appealing 2-D and 3-D visual perceptions and encrypted codes of the intended isolated words.

In current study, the acoustic excitation amplitude is kept constant (cf. Table 1). A study on the relationship between the excitation amplitude variations, while keeping the frequency constant, and the generated patterns is a future work. Moreover, a study on the relationship between the acoustic excitation harmonics, for various amplitudes, and its impact on the generated Chladni's pattern is another point to study. During this study the system might deviate the theory of wave mechanics because of limitations of different system modules like amplifier, wave driver etc.

A certain number of excitation generators, actuators and display mediums can be employed in parallel and in a variety of topologies as a function of the intended application. Moreover, tangibility can also be embedded in these mediums. In addition, the embedded implementation and miniaturisation of these systems can also be realized by employing the microcontrollers, Field Programmable Gate Arrays (FPGAs), low power wireless interfaces and Microelectromechanical systems (MEMS). This approach possesses a strong potential of opening new prospects in the development of smart fixed, portables and wearables that can be employed in various applications like visual arts, encryption, education, integration of impaired people, architecture, archaeology, etc. The embedded implementation, miniaturisation and investigation of proposed system principle usage for potential applications are rich axis to explore.

5. CONCLUSION

An inventive approach of achieving a Cymatics based visual perception of isolated speech commands has been proposed. In this context, speech enhancement and recognition processes are smartly combined with a methodical acoustic excitation generation. These excitations are converted into systematic sound waves. The sound waves are augmented and are later on employed for generating specific patterns on the configurable Cymatics based display media. A circular metallic Chladni plate with sand particles is employed as the display media. The media is called configurable because it adapts its visual pattern configuration as a function of the applied acoustic excitation. It has been demonstrated with the help of results, summarized in Table 1, that a specific visual pattern is successfully achieved for each considered auditory command.

A system prototype is successfully realised and tested. It has been shown that the proposed speech recognition module achieves an average speaker dependent isolated word recognition accuracy of 98%. The average speaker independent isolated word recognition accuracy is 95%. It expresses that the average speaker dependent recognition accuracy is 3% higher than the average speaker independent recognition accuracy. It concludes that the proposed system is a better candidate for mono user applications as compared to the poly user ones. Moreover, the proposed speech recognition module outperforms in terms of the speaker independent recognition accuracy, as compared to the existing state of the art solutions [26, 37, 47, 52]. In addition, the speaker dependent recognition accuracy, achieved by the proposed solution is also comparable with the state of the art solutions.

This proposed approach is novel and its principle can be employed in various applications like visual arts, encryption, education, integration of impaired people, architecture, archaeology, etc. The embedded implementation, miniaturisation and investigation of proposed system principle usage for potential applications are future works. The employment of signal driven speech acquisition, enhancement and recognition will improve the system efficiency in terms of resources utilization and power consumption [44-46]. The study and integration of these features in the system is another axis to explore.

6. ACKNOWLEDGEMENT

Author is thankful to Dr. M. Akbar and Dr. A. Subasi for the fruitful discussions. Author is also thankful to Eng. H. Alshareef for her help during the experimentation.

REFERENCES

- [1]. Oh, You Jin, and Sojin Kim. "Experimental Study of Cymatics." *International Journal of Engineering and Technology* 4.4 (2012): 434.
- [2]. Lewis, S. (2010). *Seeing Sound: Hans Jenny and the Cymatic Atlas* (Doctoral dissertation, University of Pittsburgh).
- [3]. Sykes, Lewis. "The augmented tonoscope." *Manchester Institute for Research and Innovation in Art and Design* (2011).
- [4]. Jenny, H. (2001). *Cymatics: a study of wave phenomena and vibration*. MACROmedia.
- [5]. Rhodenborgh, A. (2016). *Discovering an uncanny world: Cymatics software and the journey to the creation of knowledge within the field of contemporary Cymatics* (Master's thesis).
- [6]. Rossing, Thomas D. "Chladni's law for vibrating plates." *American Journal of Physics* 50.3 (1982): 271-274.
- [7]. Tuan, P. H., Wen, C. P., Chiang, P. Y., Yu, Y. T., Liang, H. C., Huang, K. F., & Chen, Y. F. (2015). Exploring the resonant vibration of thin plates: reconstruction of Chladni patterns and determination of resonant wave numbers. *The Journal of the Acoustical Society of America*, 137(4), 2113-2123.
- [8]. Krasil'nikov, V. A., & Krylov, V. V. (1984). *Introduction to physical acoustics*.
- [9]. Pettersson, Peter. "Cymatics: the science of the future." (2009).
- [10]. Echols, R. *Physics of Music: Making Waves in a Science Classroom*.
- [11]. Misseroni, D., Colquitt, D. J., Movchan, A. B., Movchan, N. V., & Jones, I. S. (2016). Cymatics for the cloaking of flexural vibrations in a structured plate. *Scientific reports*, 6, 23929.
- [12]. Wang, X. Q., & So, R. M. C. (2015). Timoshenko beam theory: A perspective based on the wave-mechanics approach. *Wave Motion*, 57, 64-87.
- [13]. Cooper, Thomas W. "FROM CONSCIOUSNESS TO TECHNOLOGY: CYMATICS, WAVE PERIODICITY, AND COMMUNICATION." *INTEGRATIVE EXPLORATIONS* 2.1 (1994): 52.
- [14] Bennett, Ian M., et al. "Distributed real time speech recognition system." U.S. Patent No. 9,076,448. 7 Jul. 2015.
- [15] E. C. Anderson, "Automatically storing and presenting digital images using a speech-based command language," Dec. 2002.
- [16] B. Burger, I. Ferrané, F. Lerasle, and G. Infantes, "Two-handed gesture recognition and fusion with speech to command a robot," *Autonomous Robots*, vol. 32, no. 2, pp. 129–147, 2012.
- [17] M. E. Deisher, R. P. Banginwar, and R. C. Knauerhase, "Speech recognition command via an intermediate mobile device," Feb. 2007.
- [18] Markel, J. D., & Gray, A. J. (2013). *Linear prediction of speech* (Vol. 12). Springer Science & Business Media.
- [19] Wang, K., An, N., Li, B. N., Zhang, Y., & Li, L. (2015). Speech emotion recognition using Fourier parameters. *IEEE Transactions on Affective Computing*, 6(1), 69-75.
- [20] Payton, T., & Claypoole, T. (2014). *Privacy in the age of Big data: Recognizing threats, defending your rights, and protecting your family*. Rowman & Littlefield.
- [21] Alvarez, C. (2017). *Lean customer development: Building products your customers will buy*. " O'Reilly Media, Inc."
- [22] Kristensen, M., Penner, J., Nguyen, A., Moy, J., & Lam, S. (2017). *COMPANY SYNOPSIS FOR: AMAZON.COM, INC.*
- [23] Ramlrez, M., Sotaquirá, M., De La Cruz, A., Maria, E., Avellaneda, G., & Ochoa, A. (2016, August). An automatic speech recognition system for helping visually impaired children to learn Braille. In *Signal Processing, Images and Artificial Vision (STSIVA), 2016 XXI Symposium on* (pp. 1-4). IEEE.
- [24] Darabkh, K. A., Haddad, L., Sweidan, S. Z., Hawa, M., Saifan, R., & Alnabelsi, S. H. (2016). An efficient speech recognition system for arm-disabled students based on isolated words. *Computer Applications in Engineering Education*.
- [25] Price, M., Glass, J., & Chandrakasan, A. P. (2015). A 6 mW, 5,000-word real-time speech recognizer using WFST models. *IEEE Journal of Solid-State Circuits*, 50(1), 102-112.

- [26] Li, L., Xu, W., Wu, J., He, S., & Li, X. (2014, December). The Hokkien isolated word recognition system based on FPGA. In *Anti-counterfeiting, Security, and Identification (ASID), 2014 International Conference on* (pp. 1-5). IEEE.
- [27] Liu, L. (2015). *Acoustic models for speech recognition using Deep Neural Networks based on approximate math* (Doctoral dissertation, Massachusetts Institute of Technology).
- [28] Sarria-Paja, M., Senoussaoui, M., & Falk, T. H. (2015, May). The effects of whispered speech on state-of-the-art voice based biometrics systems. In *Electrical and Computer Engineering (CCECE), 2015 IEEE 28th Canadian Conference on* (pp. 1254-1259). IEEE.
- [29] Blamey, P. J., Maat, B., Baskent, D., Mawman, D., Burke, E., Dillier, N., ... & Huber, A. M. (2015). A retrospective multicenter study comparing speech perception outcomes for bilateral implantation and bimodal rehabilitation. *Ear and hearing*, 36(4), 408-416.
- [30] Wang, C. Y., Shih, M., Tai, T. C., Lin, P. C., Huang, S. T., Zhao, J. H., & Wang, J. C. (2015, June). VLSI design for SC-based speaker recognition. In *Industrial Electronics and Applications (ICIEA), 2015 IEEE 10th Conference on* (pp. 335-338). IEEE.
- [31] Mantena, G., Achanta, S., & Prahallad, K. (2014). Query-by-example spoken term detection using frequency domain linear prediction and non-segmental dynamic time warping. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(5), 946-955.
- [32] Mansour, A. H., Salh, G. Z. A., & Mohammed, K. A. (2015). Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms. *International Journal of Computer Applications*, 116(2).
- [33] Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., & Paavilainen, P. (2016). Predictive coding of phonological rules in auditory cortex: A mismatch negativity study. *Brain and language*, 162, 72-80.
- [34] E. C. Ifeachor and B. W. Jervis, *Digital signal processing: a practical approach*. Pearson Education, 2002.
- [35] N. Almayouf, S. M. Qaisar et al. (2017). A speech to machine interface based on the frequency domain command recognition. *Signal and Image Processing (ICSIP), 2017 IEEE 2nd International Conference on*.
- [36] Mohan, B. J. (2014, January). Speech recognition using MFCC and DTW. In *Advances in Electrical Engineering (ICAEE), 2014 International Conference on* (pp. 1-4). IEEE.
- [37] Eljhani, M. M. (2015). *Front-end of Wake-Up-Word Speech Recognition System Design on FPGA*. Florida Institute of Technology.
- [38]. Dávila, Juan, Manuel Del Pino, and Juncheng Wei. "Concentrating standing waves for the fractional nonlinear Schrödinger equation." *Journal of Differential Equations* 256.2 (2014): 858-892.
- [39]. "100 Watt Class T Digital Audio Amplifier." Dayton Audio, 2014.
- [40]. "SF-9324 Wave Driver" Pasco scientific, 2015.
- [41]. "Chladni Plates Kit" Pasco scientific, 2015.
- [42]. "3.5mm Stereo Plug/Plug Male/Male Cable - 6ft" Secomp, 2014.
- [43]. Ingle, V. K., & Proakis, J. G. (2016). *Digital Signal Processing Using MATLAB: A Problem Solving Companion*. Cengage Learning.
- [44]. Qaisar, S. M., Fesquet, L., & Renaudin, M. (2014). Adaptive rate filtering a computationally efficient signal processing approach. *Signal Processing*, 94, 620-630.
- [45]. Qaisar, S. M., Akbar, M., Beyrouthy, T., Al-Habib, W., & Asmatulah, M. (2016, June). An error measurement for resampled level crossing signal. In *Event-based Control, Communication, and Signal Processing (EBCCSP), 2016 Second International Conference on* (pp. 1-4). IEEE.
- [46]. Qaisar, S. M. (2015, June). An event driven technique for filtering computational complexity reduction. In *Event-based Control, Communication, and Signal Processing (EBCCSP), 2015 International Conference on* (pp. 1-4). IEEE.
- [47]. Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A review on speech recognition technique. *International Journal of Computer Applications*, 10(3), 16-24.
- [48]. Sarma, M., & Sarma, K. K. (2015). *Speech Recognition in Indian Languages—A Survey*. In *Recent Trends in Intelligent and Emerging Systems* (pp. 173-187). Springer, New Delhi.
- [49]. Silva, F. (2010). *Secrets in the Fields: The Science and Mysticism of Crop Circles*. Charlottesville, VA: Hampton Roads.
- [50]. Cox, T. J. (n.d.). *Acoustic Diffusers: the Good, the Bad, and the Ugly*. Proceedings of the Institute of Acoustics. Retrieved from:
http://www.rpginc.com/docs%5CTechnology%5CWhite%20Papers%5CAcoustic%20Diffusers_The%20Good,%20The%20Bad%20and%20The%20Ugly.pdf
- [51]. Cox, T. J. & D'Antonio, P. (2009). *Acoustic Absorbers and Diffusers: Theory, Design and Application*. Florence KY: Taylor & Francis.
- [52]. Cox Rashmi, C. R. (2014). Review of algorithms and applications in speech recognition system. *International Journal of Computer Science and Information Technologies*, 5(4), 5258-5262.