SWIFTLET SOUND IDENTIFICATION USING VECTOR QUANTIZATION AND GAUSSIAN MIXTURE MODEL

SITI NURZALIKHA ZAINI BT HUSNI ZAINI

Master of Engineering (Electronic)

UNIVERSITI MALAYSIA PAHANG

Universiti Malaysia PAHANG

Engineering • Technology • Creativity

SUPERVISOR'S DECLARATION

We hereby declare that we have checked this thesis and in our opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Engineering (Electronic).

____

(Supervisor's Signature)

Full Name : DR MOHD ZAMRI BIN IBRAHIM

Position : SENIOR LECTURER

Date : 7 AUGUST 2018

____

(Co-supervisor's Signature)

Full Name : PROF MADYA DR SAIFUL NIZAM BIN TAJUDDIN

Position : SENIOR LECTURER

Date : 7 AUGUST 2018

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

____

(Student's Signature)

Full Name : SITI NURZALIKHA ZAINI BT HUSNI ZAINI

ID Number : MEL12004

Date : 7 AUGUST 2018

SWIFTLET SOUND IDENTIFICATION USING VECTOR QUANTIZATION AND GAUSSIAN MIXTURE MODEL

SITI NURZALIKHA ZAINI BT HUSNI ZAINI

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Master of Engineering (Electronic)

Faculty of Electrical & Electronics Engineering

UNIVERSITI MALAYSIA PAHANG

AUGUST 2018

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious, the Most Merciful.

All praise is due to Allah, the Creator and Sustainer of this whole universe, the Most Beneficent and the Most Merciful, for His guidance and blessing and granting me knowledge, patience and perseverance to accomplish this research successfully.

Firstly, I would like to acknowledge supportive supervisor Dr. Mohd Zamri Bin Ibrahim for his supervision and guidance to complete this thesis. Without patience and guidance from him, this research could not have been successfully completed. Special thanks to my co-supervisor Assoc. Prof. Dr. Saiful Nizam Bin Tajuddin and ex-supervisor Dr. Sunardi for their valuable motivation, moral support, encouragement and effort throughout this research.

Last but not least, most profound gratitude and respect to my beloved husband Mohd Kamarul Ridzuan bin Mohamad Himi, my son Muhammad Sayf Erfhan, my daughter Mia Sophea Emani my beloved mother Siti Hajar Binti Talib for their understanding and encouragement during my study. I would like to express my sincere appreciation to all my postgraduate's friends, who have been the ultimate source of my motivation to work hard in the completion of this research work and my entire study. Therefore, I proudly dedicate this work to both of them, may Allah SWT bless both of them.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

f	Frequency
f_s	Sampling frequency
H_{km}	Mel filter bank
H_z	Transfer function
M	Overlap window size
M_f	Number of filter bank
N	Window size
P	Magnitude Spectrum
S	Audio signal

LIST OF ABBREVIATIONS

CD	Compact Disk
D	Delta
DA	Delta-Acceleration
DCT	Discrete Cosine Transform
DTW	Dynamic Time Wrapping
EM	Expectation-Maximization
FFT	Fast Fourier Transform
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
LDA	Linear Discriminant Analysis
LPC	Linear Predictive Coding
LPCC	Linear Predictive Cepstral Coefficient
MDC	Minimum Distance Classifier
MDC	Minimum Distance Classifier
MFCC	Mel Frequency Cepstral Coefficient
SVM	Support Vector Machine
VQ	Vector Quantization
VQLBG	Vector Quantization Linde-Buzo and Gray

SWIFTLET SOUND IDENTIFICATION USING VECTOR QUANTIZATION AND GAUSSIAN MIXTURE MODEL

SITI NURZALIKHA ZAINI BT HUSNI ZAINI

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Master of Engineering (Electronic)

Faculty of Electrical & Electronics Engineering

UNIVERSITI MALAYSIA PAHANG

AUGUST 2018

ABSTRAK

Pengesanan bunyi burung merupakan salah satu aplikasi dalam teknologi mengesanan audio. Pengesanan bunyi ini adalah cara terbaik untuk mengenalpasti jenis bunyi burung walet yang berlainan antara bunyi anak, dewasa dan berkumpulan. Realitinya, pengkaji burung sukar untuk membezakan ketiga-tiga jenis bunyi ini secara kasar kecuali dengan bantuan orang yang berpengalaman dalam bidang perindustrian burung walet. Pengesanan bunyi burung walet berguna dalam perindustrian burung walet dalam meningkatkan penghasilan sarang burung dan kualiti habitat disebabkan kehidupan burung wallet tertarik kepada bunyi. Tujuan kajian ini akan mengaplikasikan cara untuk mengklasifikasi tiga jenis burung walet ini menggunakan teknologi pengesanan audio. Secara ringkasnya, kajian ini akan mengekstrak ciri-ciri audio bunyi burung walet menggunakan Linear Predictive Cepstral Coefficient (LPCC), dan Mel Frequency Cepstral Coefficient (MFCC) seterusnya mengklasifikasikan menggunakan teknik Minimum Distance Classifier (MDC), Vector Quantization (VQ) and Gaussian Mixture Model (GMM). Langkah pertama, ciri-ciri audio diekstrak menggunakan LPCC dan MFCC dan disimpan di dalam data simpanan. Langkah kedua, data simpanan tadi akan digunakan ketika teknik pengklasifikasi menggunakan teknik MDC, VQ bersama saiz kod buku 8, 16, 32 dan 64 serta teknik GMM bersama 1-campuran dan 2-campuran. Langkah ketiga, keputusan yang terbaik dipilih untuk penambahan ciri seperti Delta dan Delta-Acceleration untuk tujuan penambahbaikan peratus kejituan untuk hasil keputusan yang lebih baik. Keputusan yang diperolehi dalam kajian ini, ciri-ciri pengekstrakan menggunakan MFCC bersama teknik klasifikasi GMM 2-campuran memberi keputusan 88.89% kejituan pengesanan bunyi burung walet yang tepat. Di akhir keputusan eksperimen mendapati, peningkatan 6.67% kejituan apabila menambah ciri ekstrak Delta-Acceleration sehingga mencapai 95.56% kejituan pengesanan bunyi burung yang boleh dikira peratus yang sangat baik. Kesimpulannya, cara untuk mengesanan bunyi burung walet dalam sistem ini melalui ciri pengekstrakan menggunakan MFCC dengan ciri ekstrak tambahan Delta-Acceleration dan di klasifikasi melalui teknik GMM 2-campuran.

ABSTRACT

Bird sound identification has become one of the applications in audio recognition technology. Audio recognition is a great way to classify swiftlet's sound between baby, adult, and colony. In real life, biologists are having difficulties to identify the difference between these three types of sound except for human expert hearing experience in swiftlet farming. The identification of swiftlet sound is used to increase the production nest and quality of habitat because the main characteristic of swiftlet is its attraction toward sound. The aim of this study is to implement in swiftlet sound specifically using audio recognition to identify the types of sound. In this work, swiftlet sound feature extracted using Linear Predictive Cepstral Coefficient (LPCC), and Mel Frequency Cepstral Coefficient (MFCC) then classify the sounds using Minimum Distance Classifier (MDC), Vector Quantization (VQ) and Gaussian Mixture Model (GMM). Firstly, the features extracted using LPCC and MFCC are stored in the database. Secondly, feature extraction results in the database used for classifying the swiftlets sound using MDC, VQ with codebook size is 8, 16, 32 and 64 and GMM by 1-mixture and 2-mixture for classification. Thirdly, the best performance classification selected for an additional feature in feature extraction such as Delta and Delta-Acceleration qualifier to improve accuracy for getting a better result. Based on the result of this study, the best performance was selected based on higher accuracy identification is MFCC with GMM by 2-mixture accuracy 88.89%. At the end of the experiment, the MFCC with additional features Delta-Acceleration using classification GMM by 2-mixture with improvement 6.67% compared to original and make it up to 95.56% accuracy which is considered as good percentage result. As conclusion, the best feature extraction for swiftlet sound identification is MFCC with Delta-Acceleration features by classify the sound using GMM 2-mixture.

REFERENCES

Anderson, S. E., Dave, A. S., & Margoliash, D. (1996). Template-based automatic recognition of birdsong syllables from continuous recordings. *The Journal of the Acoustical Society of America*, *100*(2 Pt 1), 1209–1219. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8759970

Anusuya, M., & Katti, S. (2009). Speech recognition by machine: A review. *International Journal of Computer Science and Information Security*, *6*(3), 181–205. https://doi.org/10.1109/PROC.1976.10158

Astuti, W., & Aibinu, A. (2011). Animal sound activity detection using multi-class support vector machines. In *In Mechatronics (ICOM), 2011 4th International Conference On (pp. 1-5). IEEE.* (pp. 1–5). IEEE. Retrieved from https://ieeexplore.ieee.org/abstract/document/5937122/

Barbu, T. (2007). A Supervised Text-Independent Speaker Recognition Approach. *World Academy of Science, Engineering and Technology*, *33*(9), 206–210.

Bimbot, F., Bonastre, J.-F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S. Reynolds, D. A. (2004). A Tutorial on Text-Independent Speaker Verification. *Journal on Advances in Signal Processing (EURASIP)*, *2004*(4), 430–451. https://doi.org/10.1155/S1110865704310024

Borde, P., Varpe, A., Manza, R., & Yannawar, P. (2015). Recognition of isolated words using Zernike and MFCC features for audio visual speech recognition. *International Journal of Speech Technology*, *18*(2), 167–175. https://doi.org/10.1007/s10772-014-9257-1

Briggs, F., Raich, R., & Fern, X. Z. (2009). Audio Classification of Bird Species: A Statistical Manifold Approach. In *2009 Ninth IEEE International Conference on Data Mining* (pp. 51–60). IEEE. https://doi.org/10.1109/ICDM.2009.65

Cai, J., Ee, D., Pham, B., Roe, P., & Zhang, J. (2007). Sensor Network for the Monitoring of Ecosystem: Bird Species Recognition. In *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information* (pp. 293–298). IEEE. https://doi.org/10.1109/ISSNIP.2007.4496859

Chou, C.-H., Lee, C.-H., & Ni, H.-W. (2007). Bird Species Recognition by Comparing the HMMs of the Syllables. In *Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)* (pp. 143–143). IEEE. https://doi.org/10.1109/ICICIC.2007.199

Chou, C.-H., & Liu, P.-H. (2009). Bird Species Recognition by Wavelet Transformation of a Section of Birdsong. In *Ubiquitous, Autonomic and Trusted Computing 2009 (UIC-ATC'09).* (pp. 189–193). IEEE. https://doi.org/10.1109/UIC-ATC.2009.85

Chou, C.-H., Liu, P.-H., & Cai, B. (2008). On the Studies of Syllable Segmentation and Improving MFCCs for Automatic Birdsong Recognition. In *2008 IEEE Asia-Pacific Services Computing Conference* (pp. 745–750). IEEE. https://doi.org/10.1109/APSCC.2008.6

Clemins, P. J., & Johnson, M. T. (2003). Application of speech recognition to African elephant (Loxodonta africana) vocalizations. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing 2003 (ICASSP '03).* (Vol. 1, p. I-484-I-487). IEEE. https://doi.org/10.1109/ICASSP.2003.1198823

Coles, R. B., Konishi, M., & Pettigrew, J. D. (1987). Hearing and Echolocation in the Australian Grey Swiftlet , Collocalia Spodiopygia. *Journal of Experimental Biology*, *129*, 365–371.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines : and other kernel-based learning methods*. Cambridge University Press.

Dave, N. (2013). Feature Extraction Methods LPC , PLP and MFCC In Speech Recognition. *International Journal for Advance Research in Engineering and Technology*, *1*(Vi), 1–5.

Deepika, M., & Rajan, N. (2014). Automatic Identification of Bird Species from the Recorded Bird Song Using ART Approach. *International Journal of Innovative Research in Science, Engineering and Technology*, *3*(3), 668–675. Retrieved from https://www.rroij.com/open-access/automatic-identification-of-bird-species-fromthe-recorded-bird-song-using-art-approach.pdf

Deller, J. R., Hansen, J. H. L., & Proakis, J. G. (1993). *Discrete-time processing of speech signals*. Wiley-IEEE Press.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1976). Maximum Likelihood from Incomplete Data via the EM Algorithm. Retrieved from https://dash.harvard.edu/handle/1/3426318

Diment, A., Heittola, T., & Tuomas, V. (2013). *Semi-Supervised Learning for Musical Instrument Recognition*. Finland: Tampere University of Technology Department of Signal Processing.

Doupe, A. J., & Kuhl, P. K. (1999). BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms. *Annual Review of Neuroscience*, *22*(1), 567–631. https://doi.org/10.1146/annurev.neuro.22.1.567

Fagerlund, S. (2004). *Automatic Recognition of Bird Species by Their Sounds*. *Laboratory of Acoustics and Audio Signal Processing*. Helsinki University of Technology.

Fagerlund, S. (2007). Bird Species Recognition Using Support Vector Machines. *EURASIP Journal on Advances in Signal Processing*, *2007*(1), 038637. https://doi.org/10.1155/2007/38637

Fullard, J. H., Barclay, R. M. R., & Thomas, D. W. (1993). Echolocation in Free-Flying Atiu Swiftlets (Aerodramus sawtelli). *Biotropica*, *25*(3), 334. https://doi.org/10.2307/2388791

Gersho, A., & Gray, R. M. (1992). Vector Quantization I:Structure and Performance. In *Vector Quantization and Signal Compression* (pp. 309–343). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-3626-0_10

Gill, M. K., Kaur, R., & Kaur, J. (2010). Vector Quantization based Speaker Identification. *International Journal of Computer Applications*, *4*(2), 1–4. https://doi.org/10.5120/806-1146

Gish, H., & Schmidt, M. (1994). Text-independent speaker identification. *IEEE Signal Processing Magazine*, *11*(4), 18–32. https://doi.org/10.1109/79.317924

Harma, A. (2003). Automatic Identification of Bird Species based on Sinusoidal Modeling of Syllables. In *Signal Processing*. Finland: Laboratory of Acoustics and Audio Signal.

Hobbs, J. J. (2004). Problems in the harvest of edible birds' nests in Sarawak and Sabah, Malaysian Borneo. *Biodiversity and Conservation*, *13*(12), 2209–2226. https://doi.org/10.1023/B:BIOC.0000047905.79709.7f

Huang, C.-J., Yang, Y.-J., Yang, D.-X., & Chen, Y.-J. (2009). Frog classification using machine learning techniques. *Expert Systems with Applications*, *36*(2), 3737–3743. https://doi.org/10.1016/J.ESWA.2008.02.059

Ibrahim, Z. (2014). *A novel lip geometry approach for audio-visual speech recognition*. © Mohd Zamri bin Ibrahim. Retrieved from https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/16526

Idris, A., Abdullah, A.-A., & Abd-Rehman, M. (2014). An Overview of the Study of the Right Habitat and Suitable Environmental Factors that Influence the Success of Edible Bird Nest Production in Malaysia. *Asian Journal of Agricultural Research*, *8*(1), 1–16. https://doi.org/10.3923/ajar.2014.1.16

Jordan Price, J., P. Johnson, K., & H. Clayton, D. (2004). The evolution of echolocation in swiftlets. *Journal of Avian Biology*, *35*(2), 135–143. https://doi.org/10.1111/j.0908-8857.2004.03182.x

Kogan, J. A., & Margoliash, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *The Journal of the Acoustical Society of America*, *103*(4), 2185. https://doi.org/10.1121/1.421364

Kwan, C., Mei, G., Zhao, X., Ren, Z., Xu, R., Stanford, V., … Ho, K. C. (2004). Bird classification algorithms: theory and experimental results. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 5, p. V-289-92). IEEE. https://doi.org/10.1109/ICASSP.2004.1327104

Lee, C.-H., Han, C.-C., & Chuang, C.-C. (2008). Automatic Classification of Bird Species From Their Sounds Using Two-Dimensional Cepstral Coefficients. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(8), 1541–1550. https://doi.org/10.1109/TASL.2008.2005345

Lee, C.-H., Lee, Y.-K., & Huang, R.-Z. (2006). Automatic Recognition of Bird Songs Using Cepstral Coefficients. *Journal of Information Technology and Applications*, *1*(1), 17–23. https://doi.org/10.1590/S0001-37652004000200037

Lim, C. K. (2011). International Conference and Training on Swiftlet Ranching (ICOTOS 2011). In *Opportunity and sustainability of swiftlet farming in Malaysia* (pp. 17–19). Kuala Terengganu, Malaysia: Universiti Sultan Zainal Abidin.

Linde, Y., Buzo, A., & Gray, R. (1980). An Algorithm for Vector Quantizer Design. *IEEE Transactions on Communications*, *28*(1), 84–95. https://doi.org/10.1109/TCOM.1980.1094577

Lopes, M. T., Gioppo, L. L., Higushi, T. T., Kaestner, C. A. A., Silla Jr., C. N., & Koerich, A. L. (2011). Automatic Bird Species Identification for Large Number of Species. *IEEE International Symposium on Multimedia*, 117–122. https://doi.org/10.1109/ISM.2011.27

Lourie, S. A., & Tompkins, D. M. (2008). The diets of Malaysian swiftlets. *Ibis*, *142*(4), 596–602. https://doi.org/10.1111/j.1474-919X.2000.tb04459.x

Marcone, M. F. (2005). Characterization of the edible bird's nest the "Caviar of the East." *Food Research International*, *38*(10), 1125–1134. https://doi.org/10.1016/J.FOODRES.2005.02.008

McIlraith, A. L., & Card, H. C. (1995). Birdsong recognition with DSP and neural networks. In *IEEE WESCANEX 95. Communications, Power, and Computing. Conference Proceedings* (Vol. 2, pp. 409–414). IEEE. https://doi.org/10.1109/WESCAN.1995.494065

Mitrovic, D., Zeppelzauer, M., & Breiteneder, C. (2006). Discrimination and Retrieval of Animal Sounds. *12th International Multi-Media Modelling Conference*, (January), 339–343. https://doi.org/10.1109/MMMC.2006.1651344

Mulia, A. H. (2007). *Strategi Jitu Memikat Walet - A. Hendri Mulia, SE., CMA - Google Books* (1st editio). Jakarta: AgroMedia Pustaka. Retrieved from https://books.google.com.my/books?id=W5LJyW_E_7oC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

O'Shaughnessy, D. (2000). *Speech communications : human and machine*. Institute of Electrical and Electronics Engineers.

Qasem, M., Du, X., & Ahalt, S. C. (1999). Efficient codebook search for vector quantization: exploiting inherent codebook structure. In S. K. Park & R. D. Juday (Eds.), *International Society for Optics and Photonics (AeroSense '99)* (pp. 149–154). https://doi.org/10.1117/12.354702

Rabiner, L. R., & Juang, B. H. (Biing-H. (1993). *Fundamentals of speech recognition*. PTR Prentice Hall. Retrieved from https://dl.acm.org/citation.cfm?id=153687

Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker Verification Using Adapted Gaussian Mixture Models. *Digital Signal Processing*, *10*(1–3), 19–41. https://doi.org/10.1006/DSPR.1999.0361

Reynolds, D. A., & Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, *3*(1), 72–83. https://doi.org/10.1109/89.365379

Shaw, S. (2011). *An Evaluation of Birdsong Recognition Techniques*. *Department of Computer Science*. University of Sheffield.

Shrawankar, U., & Thakare, V. M. (2013). Techniques for Feature Extraction In Speech Recognition System : A Comparative Study. Retrieved from http://arxiv.org/abs/1305.1145

Thomassen, H. A., & Povel, G. D. E. (2006). Comparative and phylogenetic analysis of the echo clicks and social vocalizations of swiftlets (Aves: Apodidae). *Biological Journal of the Linnean Society*, *88*(4), 631–643. https://doi.org/10.1111/j.1095-8312.2006.00648.x

Vilches, E., Escobar, I. A., Vallejo, E. E., & Taylor, C. E. (2006). Data Mining Applied to Acoustic Bird Species Recognition. In *18th International Conference on Pattern Recognition (ICPR'06)* (pp. 400–403). IEEE. https://doi.org/10.1109/ICPR.2006.426

Wan Ibrahim, W. K. (2016). *Kelestarian Usahawan Melayu Kelantan: Kajian Terhadap Industri Sarang Burung Walit*. Universiti Malaysia Kelantan.

Young, S., Kershaw, D., & Woodland, P. (1995). *The HTK Book*. Cambridge University Engineering Department.