Advances in Intelligent Systems and Computing 1455

Keikichi Hirose
Deepak Joshi
Shankha Sanyal *Editors*

Proceedings of 27th International Symposium on Frontiers of Research in Speech and Music

FRSM 2023

Springer

Advances in Intelligent Systems and Computing

Volume 1455

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by DBLP, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST).

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Keikichi Hirose · Deepak Joshi · Shankha Sanyal
Editors

Proceedings of 27th International Symposium on Frontiers of Research in Speech and Music

FRSM 2023

Springer

Editors
Keikichi Hirose
University of Tokyo
Tokyo, Japan

Shankha Sanyal
School of Languages and Linguistics
Sir C. V. Raman Centre for Physics
and Music
Jadavpur University
Kolkata, West Bengal, India

Deepak Joshi
Department of Electronics Engineering
Sardar Vallabhbhai National Institute
of Technology
Surat, Gujarat, India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-97-1548-0 ISBN 978-981-97-1549-7 (eBook)
https://doi.org/10.1007/978-981-97-1549-7

to the memory of

Late Prof. Asok Kumar Datta

and

Late Prof. Sanghamitra Mohanty,

visionaries behind the concept of FRSM

Organized by

Department of Electronics Engineering,
Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat

Sir C. V. Raman Centre for Physics and Music,
Jadavpur University, Kolkata, West Bengal

Message from the Organizing Chair

Prof. (Dr.) Shyam Sunder Agrawal
Director General, KIIT Group of Institutions.
Ex. Emeritus Scientist CSIR & Advisor to CDAC

It gives me immense pleasure to know that the 27th International Symposium on Frontiers of Research in Speech and Music (FRSM-2023) is being held at SVNIT, Surat in collaboration with Sir C. V. Raman Centre for Physics and Music, Jadavpur University on 4th and 5th August 2023 in the Diamond/ Silk city of our Country.

FRSM is a conference of unique nature focussed on highly specialized areas of research and study in Speech and Music, the two main forms of interpersonal communication among humans and use common apparatus of vocal system to produce them and hearing mechanism to listen to them and complex processing through brain. The study of these fields encompasses several disciplines such as physics, acoustics, mathematics, linguistics, phonetics, musicology, physiology, psychology, electronics, and computer science. The researchers in this area have made great contributions and many of them will be presented by researchers from India and abroad in this symposium.

FRSM has a long tradition and being held since 1991, in different institutions of higher learning and research in India with the initiatives and vision of Late Dr. Deepali

Nag and Late Prof. Asoke Datta in particular. A large number of distinguished scientists and researchers from many countries including Japan, Sweden, Spain, Australia, UK, etc., and many from India have personally participated and interacted with each other and have discussed state of art and emerging areas of research in this field. However, we have to still do a lot of work, create facilities, awareness, and importance of this research, and solve problems by integrating the tools of modern science, technology, AI, etc.

I would like to convey my great appreciation and gratitude to the organizers, Director SVNIT, Surat, and the Head Sir CV Raman Centre, Jadavpur, and participating scientists for sharing their knowledge.

My best wishes for its grand success.

Message from the General Chair

Dr. Suman Deb
SVNIT, Surat

The annual symposia on Frontiers of Research on Speech and Music (FRSM) have been organized in various locations across the country since 1990. These symposia serve as a crucial platform for scientists and technologists to engage in meaningful interactions. To date, a series of symposiums have been conducted in various locations including Kolkata (thrice), Varanasi, New Delhi, Santiniketan, Thiruvananthapuram, Kanpur (twice), Annamalai, Bhubaneswar (twice), Lucknow, Gwalior, Kharagpur, Gurgaon, Mysore (twice), Baripada, Rourkela, Silchar (in virtual mode), and Pune (in virtual mode). These symposiums aim to foster research interest among scholars from diverse fields such as physics, mathematics, musicology, linguistics, and computer science. The sustained progression of these occurrences would not have been feasible without the active collaboration of various universities and institutes in India.

The symposium (FRSM—2023) is taking place this year from August 4–6 at SVNIT in Surat, India. The Sir C. V. Raman Centre for Physics and Music at Jadavpur University in Kolkata and SVNIT in Surat are jointly organizing it. The goal of this symposium is to highlight the state-of-the-art and emerging applications and research in the field of speech, music, and related signal processing. In order for these emerging innovations in science and technology to develop via reciprocal interaction and be

critically analyzed by the individuals working in these sectors, the organizers wish to provide an annual venue for them. We accept and actively seek out novel concepts and methods.

Advisory Committee

- Prof. Anupam Shukla, Director, SVNIT
- Prof. Amitava Dutta, Pro VC, Jadavpur University
- Prof. Tzung-Pei Hong, National University of Kaohsiung, Taiwan
- Prof. Solvi YSTAD, LMA-CNRS, France
- Prof. Xavier Serra, University of Pompeu Fabra, Barcelona, Spain
- Prof. Kristoffer Jensen, ad: mt, AAU, Denmark
- Prof. John Ohala, Berkeley, USA
- Prof. Aditi Lahiri, Oxford University, UK
- Prof. M. Pushpavathi, Director, AIISH, Mysore
- Prof. Hiroya Fujisaki, University of Tokyo
- Prof. Anupam Basu, Pro Vice Chancellor, Sister Nivedita University, Durgapur
- Prof. Phalguni Gupta, Vice Chancellor and Professor, Institute of Engineering & Technology, GLA University, Mathura
- Prof. Keikichi Hirose, University of Tokyo
- Prof. J. Sundberg, KTH, RIT, Sweden
- Prof. S. S. Agrawal, KIIT Group of Colleges, Gurgaon
- Prof. W. M. Hartman, Michigan State Univ., USA
- Prof. Partha Ghose, NAS, India
- Prof. S. G. Deshmukh, Director, ABV-IIITM, Gwalior
- Prof. Bhaskar Gupta, JU, Kolkata
- Mr. Javed Hashmi, Joint Secretary, AIT, Kanpur
- Dr. Carlos Guedes, NYU, Abu Dhabi
- Prof. Sanjay Kumar Singh, IIT BHU, India

Organizing Committee

Chief Patron

- Prof. Anupam Shukla, Director, SVNIT

Honorary Chairs

- Prof. S. S. Agarwal, KIIT Group of Colleges
- Prof. Dipak. Ghosh, SCVRCPM, JU

General Chairs

- Dr. Suman Deb, SVNIT, Surat
- Prof. Argha Deb, SCRCVM, JU, Kolkata
- Prof. R. Sengupta, SCVRCPM, JU, Kolkata

Program Chairs

- Dr. Anand D. Darji, SVNIT, Surat
- Dr. Kishor P. Upla, SVNIT, Surat
- Dr. Deepak Joshi, SVNIT, Surat
- Dr. Shyamal Kr. Das Mandal, IIT, KGP
- Dr. Shankha Sanyal, JU, Kolkata
- Dr. Archi Banerjee, IIT, KGP

Local Committee Members

- Prof. J. N. Sarvaiya
- Dr. Prashant K. Shah
- Dr. Abhilash S. Mandloi
- Dr. Kamal Captain
- Dr. Kirti Inamdar
- Dr. Raghavendra Pal

Technical Program Committee

- Prof. Mireille Besson, Research Director, CNRS, Aix-Mairseille University, France
- Prof. Kuldeep K. Paliwal, Griffith University, Australia
- Prof. Alicja Wieczorkowska, Polish-Japanese Academy of Information Technology, Warsaw, Poland
- Prof. Shahanum Mohd Shah, Universiti Teknologi MARA (UiTM), Malaysia
- Prof. Phalguni Gupta, Director, NITTTR Kolkata, India
- Prof. Sanghamitra Mohanty, Ex- VC, NOU, Odisha, India
- Prof. George Tzanetakis, University of Victoria, Canada
- Dr. Monojit Choudhury, Microsoft Research Lab, India
- Prof. Debashis De, Maulana Abul Kalam Azad University of Technology, India
- Prof. Adam Ockelford, Director, Applied Music Research Centre, University of Roehampton, UK
- Prof. Pabitra Pal Chowdhury, Indian Statistical Institute Kolkata, India
- Prof. Pradip Kumar Das, Indian Institute Technology Guwahati, India
- Dr. Asiya Khan, Plymouth University, UK
- Dr. Bhaskar Biswas, Indian Institute Technology (BHU) Varanasi, India
- Prof. Dr. Emile Wennekes, Chair Professor, Utrecht University, Netharland
- Dr. Ajay Srinivas Murthy, Amazon, India
- Dr. Kaustav Kanti Ganguly, New York University, Abu Dhabi
- Prof. C. T. Bhunia, Former Director, NIT Arunachal Pradesh, India
- Dr. Ferdous Ahmed Barbhuiya, IIT Guwahati, India
- Mr. Birjodh Tiwana, Linkedin, San Francisco, USA
- Prof. Nityananda Sharma, Tezpur University, India
- Dr. Vijay Bhaskar Semwal, MANIT Bhopal, India
- Prof. Alex Noel Joseph Raj, Shantou University, China
- Dr. Parismita Sarma, Gauhati University, India
- Dr. Swanirbhar Majumder, Tripura Central University, India
- Prof. Hima Bindu Maringanti, North Orissa University, India
- Prof. Omar Farooq, Aligarh Muslim University, India
- Prof. K. Samudravijay, TIFR Mumbai, India

- Mr. Sauhaarda Chowdhuri, Resoniq Research, San Diego, USA
- Prof. Subhendu Ghosh, Delhi University, India
- Prof. Debasis Gangopadhyay, RKMVU, Belur, India
- Dr. Sudhakar Tripathi, REC, Lucknow, India
- Dr. Kaushik Banerjee, RBU, Kolkata, India
- Dr. Michał Kosiedowski, Poznań Supercomputing and Networking Centre, Poland
- Prof. Josef van Genabith, University of Saarland, Germany
- Prof. Alexander Gelbukh, Instituto Politecnico Nacional, Mexico
- Dr. Mridula Verma, IDRBT Heyderabad, India
- Prof. N. Parimala, Jawaharlal Nehru University, India
- Dr. Debasis Das, Indian Institute of Technology Jodhpur, India
- Prof. Rajni Jindal, Delhi Technical University, India
- Prof. R. K. Agrawal, Jawaharlal Nehru University, India
- Prof. Priyadarshi Patnaik, Indian Institute of Technology Kharagpur, India
- Dr. Vibhav Prakash Singh, MNNIT Allahabad, India
- Prof. Alain Trémeau, Université Jean Monnet, France
- Prof. N. Sreedevi, AIISH, Mysore, India
- Prof. Wai Ho Mow, Hong Kong University of Science and Technolog, Hong Kong
- Dr. Nandini Chatterjee Singh, NBRC Gurgaon, India
- Prof. Denis Trcek, University of Ljubljana, Slovenia
- Dr. Punita Singh, CSIR-NPL, New Delhi, India
- Mr. Abhinandan Bhunia, Facebook Inc., USA
- Dr. Sanasam Ranbir Singh, Indian Institute of Technology Guwahati, India
- Prof. Punam Kumar Saha, University of Iowa, USA
- Dr. Lairenlakpam Jaiprakash, North Eastern Hill University, India
- Prof. Paolo Rosso, Universitat Politècnica de València, Spain
- Dr. J. K. Mantri, North Orissa University, India
- Dr. Shantala Hegde, National Institute of Mental Health and Neuro Sciences (NIMHANS), Bengaluru, India
- Dr. Archi Banerjee, Indian Institute of Technology Kharagpur, India
- Dr. Ayush Dugra, IIT Roorkee, India

Preface

The works presented in this edited volume have the underlying motivation to look for objective data and corresponding categorization problems in different domains of speech, music, language, and allied interdisciplinary areas which were all presented in the 27th International Symposium on Frontiers of Research on Speech and Music (FRSM—2023), held at Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat from 4–5th August 2023. The Editorial Volume attempts to consolidate the substances of different interesting pieces of investigations in this diverse field carried out by different researchers across the globe. It may quite be possible that in spite of taking utmost care in choosing the selected and thoroughly revised versions of the manuscript, some inadvertent errors might have cropped in or there might be certain areas of research which we may have overlooked, if so, the Editors regret the same.

In the last two or three decades, the research on Speech, music, and the allied multidisciplinary domains have grown manifold with the advent of artificial intelligence and have become the two most prominent areas of research in the domain of audio signal processing. This Editorial Volume tries to bring together a number of frontier research works being made in the domains of language, speech, music, and their allied applications. In this Preface, let us give a brief overview into the various advances that have been taking place in these unique fields over the last few decades, the advances that have made these topics one of the most researched ones. We will round this Preface with a short overview of the various collected works that have found their place in this unique Editorial Collection.

The quest for the origin of speech and music is enigmatic. Holistic, manipulative, multimodal, musical, and mimetic language of the hominids also known as 'hmmmm' or 'musilanguage' refers to a pre-linguistic system of vocal communication from which both music and language are said to be derived later. Modern speech and music are forms of intentional emotional manipulation, and therefore not possible until the onset of intentionality—the ability to reflect on the past and the future. Between 60,000 and 30,000 years ago in the upper Palaeolithic age, humans started creating art in the form of paintings on cave walls, jewellery, and so on. Speech and music as we know now must also have emerged during that period. Speech and

Music are the two highest and most subtle and common forms of human communication. The understanding of these oldest and most fundamental faculties in human beings needs the scientific and technological resources of Language Processing.

The physical data in research in speech and music being of natural origin, a significant amount of mathematical modelling is non-deterministic in nature. Except for the primary mandatory signal processing aspects to extract parameters, the rest of the analysis has to rely largely on stochastic as well as non-stochastic random mathematical approaches like chaos and fractals. In fact, it often happens that new paradigms of mathematical approaches have to be resorted to, which often brings clarity to the mathematical non-deterministic approaches. The same is true for music. In fact, in a multi-categorical musical country like India, where one has classical music at one end and numerous ethnic and folk music at the other end, it is only the objective analysis of music which can bring out the cross-cultural intercourses taking place. Unfortunately, little attention has been paid to this subject by main national knowledge institutions. The fact that in India objective research in these two fields is never considered seriously by the academicians was also an important additional factor. Keeping the level of the interaction accessible to all, novice to erudite scholars was the modus operandi.

The language we speak in our everyday lives propagates as sound waves through various media and allows communication or entertainment for us, humans. The music we hear or create can be perceived in different aspects as rhythm, melody, harmony, timbre, or mood. Like speech, singing can also be a way to communicate. Both speech and music can be described using the four basic parameters of sound: pitch (how high or low the note is), loudness, duration, and timbre (the quality or tone of a sound; put simply, it is what makes one musical sound different from another). Speech and music use these parameters in different unique ways. In singing, the two are brought together. This may seem simple when we listen to music, but in fact, reaching an accommodation between speech and music is a complex human skill. The multifaceted nature of speech or music information requires algorithms, systems using sophisticated signal processing, machine learning, and deep learning techniques to better extract useful information. A systematic scientific investigation into the production, perception, and cognition of speech and music, thus requires a multidisciplinary approach, involving Physics, Computer Science, Phonetics, Physiology, Psychology, and Musicology to name a few. Even if one puts aside aesthetic appreciation, the task becomes formidable. To name a few aspects from the multifaceted nature of the study, viz, the purity of pitch, the purity of vowel articulation, the beauty of voice, and that of rhythm, opens up a vast area of study. Music Information Retrieval (MIR) has been one of the most researched fields in the last decade which deals with different aspects of searching and organizing large collections of music, or music information, according to their relevance to specific queries, i.e., involving different classification and feature extraction algorithms. In this collection, we have a few chapters which lie on the periphery of MIR, but cannot be said to be directly part of MIR. This has been done consciously as there exists already a number of books in the field of MIR and we wanted to keep the focus of this volume primarily

on the classical studies of speech, music, and their associated application-oriented fields.

Almost about a century ago, the great Indian philosopher and poet Rabindranath Tagore wrote "Enough study has been done about the musical structure of different ragas. I do not think more study is important in this domain. What is important is that scientists should endeavor to find the reasons why different musical structures evoke different emotions?" In the last few decades, the advent of robust bio-sensors has also provided us with a medium to look into the intricacies and neuronal connections present in the human brain and a way to model the brain. These bio-sensors have enabled scientific endeavors to progress in the path envisaged by Tagore and a number of studies have shown that listening to music as well as receiving formal music training, creates an effect of plasticity from the cochlea to the auditory cortex in the human brain. Since the auditory path of musical sounds overlaps functionally with that of speech path, music actually aids in the perception of speech too. Both perceptual and cognitive functions are involved in this process. Music engages a large area of the brain, so music can be used as a supplement in rehabilitation programs and helps the improvement of speech and language skills. For example, studies show that for people with Alzheimer's, music can aid in the process of helping patients access memories that were previously lost. There's also evidence of patients who have suffered brain damage and lost the ability to speak but that can still sing when music is played. As elucidated earlier, many excellent studies have shown similar mechanisms between speech and music across many levels. However, a fundamental question, often overlooked, is what makes the brain treat music and speech signals differently, and why do humans need two distinct auditory signals. A number of new studies involving advanced neural network theories and computational advances, are pointing toward differences in pitch and rhythm as key factors that enable people starting in infancy to distinguish speech from music, as well as how the predictive capabilities of the brain underlie both speech and music perception.

In India, systematic scientific research on music acoustics started after 1920 when Sir C. V. Raman did some pioneering work on Indian string and percussion instruments. After that few Indian scientists namely Ramakrishna and Kar did some systematic research on musical acoustics. In 1978, ITC Sangeet Research Academy was formed for the teaching and research of Indian music, it had an equipped Scientific Research Department; probably the only of its kind in India, and an academia of traditional gurus to impart training in Gurukul form. They carried out research in various fields of vocal and instrumental music and published their work in various journals. In the year 1990, it was decided at this institute to organize a yearly symposium with an objective to bring researchers in the field of speech and music together. This symposium is held in different parts of the country. The objective of the symposium has always been to encourage interactions from physical, biological, psychologica,l and neuro-sciences in the field of speech and music to provide new ideas and directions. This is the only symposium of its kind, till date, in India in the field. After the closure of the Scientific Research Department of ITC Sangeet Research Academy in 2015, this symposium is continued by Sir C. V. Raman Centre for Physics and Music,

Jadavpur University in different places of India with the help of other Universities/ Institutes. It has completed its 27th year.

The book contains thirty chapters spanning over the following broad areas:

In the area of Speech and Hearing-

(a) Proposal of an Automated Hate speech identification system in different social media platforms using Deep Learning methods.
(b) Compares traditional methods of emotion classification with latest Deep Learning techniques using MFCC features for emotion classification.
(c) ML Based Application for Enhanced Communication with Specially Abled Children.

The results from this interesting paper show that the application has the potential to improve the interaction experience for children with physical limitations and enhance their ability to communicate with the world around them.

(d) Acoustic Phonetic Analysis of Vowels of Native Bengali Speakers with Parkinson's Disease where the findings reveal significant changes in the acoustic properties of vowels in PD patients compared to the control group, which can have implications for speech therapy.
(e) Accent conversion aims to synthesize a new voice that contains the voice quality of the source speaker but accent of the target speaker. The researchers present an overview of the accent conversion techniques and their performance evaluation methods based on deep learning.

In the area of Music-

(a) Ongoing research and continuation of the CompMusic project in the tradition of Carnatic music, introducing new methodologies for dataset gathering and Music Information Retrieval tasks.
(b) An Ethnomusicological Study of Darpawngi's Mizo Folksongs where the aim is to determine the social positions of women through the songs she/they sing/ s and attempt to seek her/their individualities in a cultural/musical setting, how the songs serve as a role in reflecting or protesting the various gender arrangements, ranging from male dominance to female separation.
(c) An approach of similarity measure in Hindustani music where the authors propose a quantitative method for the assessment of closeness of the performance of a student with respect to his/her teacher/idol in Hindustani classical music.
(d) A search for different ornamentations in Hindustani classical music using different acoustical parameters.
(e) Comparison of nasality in female vocalists of Hindustani classical music using basic acoustical parameters like fundamental frequency and formants.
(f) Use of a Regression Approach for Identification of Shruthi from Indian Classical Instrumental Polyphonic Audio.
(g) Use of Convolutional Neural Networks (CNNs) for classification of music induced emotions generated from EEG signals while listening to Western and Indian music clips which have been psychologically rated earlier.

(h) Studying the acoustical and perceptual correlates of two Indian spiritual music traditions—(a) Sikh Gurbani/ Sabad Kirtan, (b) Bangla Vaishnav Kirtan using different linear (pitch, entropy, etc) and nonlinear parameters like Detrended Fluctuation Analysis (DFA).
(i) A novel approach for classification and characterization of emotions (happy-sad) and audience categories (Musician-Non Musician) while listening to Hindustani music depending upon the nature of both audio-EEG cross correlation as well as inter-lobe EEG cross-correlation.
(j) Proposal of a methodology to extract loudness descriptor from audio that can further be encoded in the extended framework to map to dynamics in automatic music learning and assessment for Hindustani music.
(k) Learning representations from audio data in Indian musical instruments, and using the transformer models to classify instrument types and emotions.
(l) Studying the styles and rhythms in Hindustani Raga pairs (each pair consisting of same musical notes)—a) Bhairav and Kalingda, b) Marwa and Sohini using various linear spectral features like spectral skewness, centroid, and pitch contour etc.
(m) Analysis of the timbre characteristics of nine strokes from five different tablas from the LTAS. Signals and consequently the study of statistical correlations among timbre parameters.

In the area of Languages and Linguistics-

(a) A comparative study of recitation vs singing in Bengali literature identical literary texts by forms, mark the difference during utterance because of prosodic variation. Also analyzing the fact that structural modifications are triggered by the change in prosodic pattern while transforming a poetic form into a song and vice versa.
(b) Acoustic and neuro-cognitive classification of multi-verb constructions—complex predicate constructions (CPC) and serial verb constructions (SVC), constructions which are very much predominant in languages of South-East Asian family.
(c) Conceptualization of *sprachbund* 'linguistic area' as a metric space corresponding to the Federation of Indian languages belonging to the new Indo-Aryan language family.
(d) A study on the unique, unexplored genre of elementary Bengali nursery rhymes (chhara) and explores the contribution of the underlying music that cooperates with motion, helping the children to retain and recall the verses.
(e) An in-depth investigation into the recognition of speech acts as orders or requests in Bengali, emphasizing the role of intonation using robust behavioral and EEG experiments.

In the area of Multidisciplinary applications-

(a) A Case Study in Turn-Taking while understanding and developing the Musical Syntax and Sonification of Voice & Speech Interfaces.

(b) Focusing on the research within the broad framework of cognitive sciences that investigate how the cognitive and neural processing of language and music are related.
(c) Study of the intermediality of musical emotions from the perspective of audio visual and (AV) and audio-only (AO) stimulus and their corresponding neural manifestations.
(d) Proposal of a novel method for bird species identification from the Timbral attributes of a sound produced by the bird.
(e) Proposal of two new features, residual harmonic amplitude (RHA) and residual harmonic frequency (RHF) using Fourier model-based analysis of the linear prediction (LP) residual signal for diagnosing Parkinson's disease (PD) from the speech signal.
(f) Proposal of a neural network based detection of common cold disease through speech signal processing.

The above mentioned 30 selected works presented in this Editorial Volume have been taken from eminent researchers in the allied disciplines of Speech and Music across the globe. We have put special emphasis on collecting works related to language, speech, and musical paradigms of the Indian subcontinent. This Editorial Volume is primarily meant for serious researchers in the pristine field of speech and music, though it is curated in a way such that it reaches the larger community of budding academicians, students, music learners, industry personnel as well as technical and research institutions in general.

Tokyo, Japan — Keikichi Hirose
Surat, India — Deepak Joshi
Kolkata, India — Shankha Sanyal

Acknowledgements

The list of acknowledgements would be quite large as this Volume is ultimately a culmination of curated works from the huge volume of works presented in the 26th International Symposium on Frontiers of Research in Speech and Music, hosted jointly by the Indian Institute of Information Technology, Pune and Sir C.V. Raman Centre for Physics and Music, Jadavpur University, India. The Editors would like to extend their gratitude to Prof. Anupam Shukla, Director, SVNIT, Surat, and Prof. Suranjan Das, Vice Chancellor, Jadavpur University for being the main pillar of strength behind the organization of the Symposium. Professor S. S. Agrawal and Dipak Ghosh have been instrumental in providing valuable inputs whenever required, during the Editorial and Review process. Professors Priyadarshi Patnaik, Argha Deb, and Dr. Samir Karmakar have been the most important persons behind the teams that worked together from IIT Kharagpur and Jadavpur University, respectively. Professors Preeti Rao, Xavier Serra, Indranil Dutta, Dr. Vinoo Alluri, Prof. Anders Friberg, Dr. Vipul Arora, Prof. Alicja Wieczworska, Bruce Balentine have to be acknowledged for their continuous support and companionship during the conference as well as during the Editorial Process of this Volume. Professor Subhendu Ghosh, Soubhik Chakraborty, Phalguni Gupta, H. B. Maringanti, Dr. Kaustuv Kanti Ganguly and Dr. S. K. Das Mondal have been associated with the FRSM community for a long time and played key roles during this year's Conference and the post collection process of manuscripts for this Volume. The review and selection process of the manuscripts involved a long list of reviewers, who have gratefully responded to our call and reviewed the manuscripts, the Editors are indeed thankful to them. In the end, the Editors would like to thank all the faculties, colleagues, and students associated with Sir C. V. Raman Centre for Physics and Music, School of Languages and Linguistics, Jadavpur University, and Indian Institute of Information Technology, Pune without whose continuous support and encouragement the culmination of this book would not have been possible. Lastly, the acknowledgement would remain incomplete without the mention of the continuous and prompt support that we received from the Editorial Team of Springer during the Editorial process.

Contents

Speech and Hearing (SH)

Framework for Detecting Toxic Speech Using BERT and Deep Learning 3
Ankit Barai, Pooja Jain, and Tapan Kumar

Speech Emotion Classification Using Deep Learning 19
Siba Prasad Mishra, Pankaj Warule, and Suman Deb

ML-Based Application for Enhanced Communication with Specially Abled Children 33
Eeshan Wairagade, Dhananjay Mishra, Anushka Chauhan, and Pooja Jain

Acoustic Phonetic Analysis of Vowels of Native Bengali Speakers with Parkinson's Disease 47
Himabindu Matta, Puja Bharati, Sabyasachi Chandra, G Satya Prasad, Shyamal Kumar Das Mandal, and Nirmalya Ghosh

Recent Advancement in Accent Conversion Using Deep Learning Techniques: A Comprehensive Review 61
Sabyasachi Chandra, Puja Bharati, G. Satya Prasad, Debolina Pramanik, and Shyamal Kumar Das Mandal

Speech Enhancement: Traditional and Deep Learning Techniques 75
Satya Prasad Gaddamedi, Anuj Patel, Sabyasachi Chandra, Puja Bharati, Nirmalya Ghosh, and Shyamal Kumar Das Mandal

Music (M)

Continuing CompMusic: New Approaches in the Computational Analysis of Carnatic Music 89
Genís Plaja-Roglans, Thomas Nuttall, Xavier Serra, and Marius Miron

An Ethnomusicological Study of Darpawngi's Mizo Folksongs: A Cantometric Experiment 101
C. Lalnunhlui

An Approach of Similarity Measure in Hindustani Music 113
Anirban Patranabis, Kaushik Banerjee, Ranjan Sengupta, and Dipak Ghosh

Ornamentation in Hindustani Vocal Music 125
Kaushik Banerjee, Anirban Patranabis, Ranjan Sengupta, and Dipak Ghosh

A Comparative Study of the Non-nasal and Nasal Voice in Hindustani Vocal Music 137
Aditi Mandal, Anirban Patranabis, Kaushik Banerjee, Argha Deb, Ranjan Sengupta, and Dipak Ghosh

Regression Approach for Shruthi Identification from Indian Classical Instrumental Polyphonic Audio 153
Pratap A. Kygonahalli, Vaibhav Katariya, and B. Ashwini

Music-Evoked Emotion Classification from EEG: An Image-Based CNN Approach 169
Bommisetty Hema Mallika, Junmoni Borgohain, Archi Banerjee, and Priyadarshi Patnaik

Perception of Devotion and Happiness in Indian Spiritual Music: An Acoustical and Audience Response Exploration 191
Archi Banerjee, Medha Basu, Shankha Sanyal, Junmoni Borgohain, and Priyadarshi Patnaik

Identifying Correlations Between Hindustani Music and the Brain: A Nonlinear EEG-Based Exploration 217
Medha Basu, Shankha Sanyal, Archi Banerjee, Sayan Nag, Ranjan Sengupta, Kumardeb Banerjee, and Dipak Ghosh

Representation and Analysis of Dynamics for Automated Music Assessment in Hindustani Vocal Music 229
Jyoti Narang, Ajay Srinivasamurthy, and Xavier Serra

Emotion and Instrument Recognition from Indian Classical Music Recordings Using Transformers 239
Ahana Deb, Ayan Sinha Mahapatra, Shankha Sanyal, Archi Banerjee, Sayan Nag, Medha Basu, and Dipak Ghosh

Styles and Rhythms of Musical Transitions in Indian Ragas: An Acoustical Exploration 251
Medha Basu, Archi Banerjee, Shankha Sanyal, Kumardeb Banerjee, and Dipak Ghosh

A Statistical Approach to the Acoustical Analysis of Harmonics and Timbre of Tabla Strokes 267
Anirban Patranabis, Kaushik Banerjee, Vishal Midya, Sneha Chakraborty, Shankha Sanyal, Archi Banerjee, Ranjan Sengupta, and Dipak Ghosh

Language and Linguistics (LL)

Recitation Versus Singing: A Prosodic Intervention 289
Rimi Ghosh Dastidar

Acoustic and Neural Signatures of Event Construals in Bengali 301
Shankha Sanyal, Aniket Goswami, Namrata Paul, Shahbaz Hoque, Rajanya Purkait, Pijush Kanti Gayen, Samir Karmakar, and Dipak Ghosh

***Sprachbund* as Metric Space: Quantifying Linguistic Traces of Convergences and Diffusions in Synchrony** 323
Anagh Chattopadhyay, Soumya Sankar Ghosh, and Samir Karmakar

Understanding the Interplay of Music, Motion, and Memory in Bengali Elementary *Chhara* (Nursery Rhymes) 337
Rajoshree Chatterjee

Classification of Speech Rhythm Using Brain Rhythm: An EEG Study on Natural Bengali Speech Prosody 349
Pijush Kanti Gayen, Medha Basu, Uddalok Sarkar, Shankha Sanyal, Archi Banerjee, Sayan Nag, Samir Karmakar, and Dipak Ghosh

Multidisciplinary Applications (MA)

Musical Syntax and Sonification of Voice and Speech Interfaces: A Case Study in Turn-Taking 369
Bruce Balentine

Speech Based Meaning of Music 385
Alice Karbanova

Intermediality of Musical Emotions in a Multimodal Scenario: Deep Learning-Aided EEG Correlation Study 399
Shankha Sanyal, Archi Banerjee, Sayan Nag, Medha Basu, Madhuparna Gangopadhyay, and Dipak Ghosh

Identification and Classification of Singing Birds Using Machine Learning 415
Shobha Lolge and Saurabh Deshmukh

Fourier Model-Based Analysis of LP Residual for Diagnosing Parkinson's Disease Using Speech 429
Pankaj Warule, Siba Prasad Mishra, Sudhansu Sekhar Nayak, and Suman Deb

LSTM- and GRU-Based Common Cold Detection from Speech Signal 439
Snigdha Chandratre, Pankaj Warule, Siba Prasad Mishra, and Suman Deb

Author Index 449

About the Editors

Dr. Keikichi Hirose received Ph.D. degree in Electronic Engineering in 1977 from the University of Tokyo. He was a professor at the University of Tokyo from 1994. In 2015, he retired from the University of Tokyo, and received Professor of Emeritus title. He also has been a visiting professor of National Institute of Informatics since 2015, and served as a project professor from June 2019—March 2023. From March 1987 to January 1988, he was a Visiting Scientist at the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, U.S.A. He has been engaged in a wide range of research on spoken language processing, including analysis, synthesis, recognition, dialogue systems, and computer-assisted language learning. From 2000 to 2004, he was a principal investigator of the national project "Realization of advanced spoken language information processing utilizing prosodic features," supported by the Japanese Government. He served as the general chair for INTERSPEECH 2010, Makuhari, Japan. During 2010–2018, he served as the chair of ISCA Special Interest Group on Speech Prosody (SProSIG). Also, he served as a board member of ISCA during 2009–2017. He has been a member of International Advisory Council of ISCA since January 2021. He received ISCA fellow grade in 2018. He became an honorary member of, Polish Phonetic Association, in 2013. For his long-term and remarkable contribution to spoken language processing, he received Achievement Award from Acoustical Society of Japan in 2020. In 2015, he was honored as a Named Person of Merit in Science and Technology by the Mayor of Tokyo. He published more than 100 journal papers and more than 380 conference papers.

Dr. Deepak Joshi is currently working as an assistant professor in the Department of Electronics Engineering, at Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, Gujarat, India. He obtained his doctoral degree from IIT Guwahati in the area of Microelectronics and VLSI Design. His areas of research include the design and optimization of analog circuits, CAD for VLSI Design, and Metaheuristics.

Dr. Shankha Sanyal is a Ph.D. in the field of Cognitive Neurophysics from Sir C. V. Raman Centre for Physics and Music, Jadavpur University. He completed Graduation and Post-Graduation in Physics from Jadavpur University, Kolkata. He is currently a post-doctoral research associate at the School of Languages and Linguistics, Jadavpur University. His primary research interest currently revolves around studying the acoustical and neural manifestations of speech and music signals using state-of-the-art signal processing algorithms. He has published in more than 25 peer-reviewed international journal papers and more than 100 conference papers including a book from Springer. He is a member of the Acoustical Society of America, Asiatic Society, Kolkata as well as a reviewer of several international journals such as Frontiers in Neuroscience and Cognitive Neurodynamics. He has been part of the Organizing and Program of FRSM conference series since 2015.

Speech and Hearing (SH)

Framework for Detecting Toxic Speech Using BERT and Deep Learning

Ankit Barai, Pooja Jain and Tapan Kumar

Abstract We have witnessed a big surge in the use of online social media platforms in the past few years with the widespread availability of mobile phones and cheaper data rates. This has also led to an increase in misuse of these platforms to spread hatred in the society. The stats show that most of the cyberbullying happens on social media platforms like Facebook, WhatsApp, etc. The Hatred speech was used for spreading hatred against specific persons, communities, etc. It is necessary to check for the hate speech on social media platforms before it can lead to damaging the peace situation. Automating the task of hate speech detection is a very important step in mitigating this issue as such contents can be quickly acted upon before it spreads hatred. In this paper, we are proposing an Automated Hate speech identification system. The datasets are obtained from HASOC which stands for Hate Speech and Offensive Content Identification in Indo-European Languages, for sub-task b to classify into one of the four categories, Hate, Profane, Offensive, or None for English and Hindi languages. The experiments are carried out to compare the performance of Bert-based architecture with conventional deep learning architectures including CNN (Convolution Neural Network) architecture for English subtask and for Hindi Subtask we fine-tuned the Bert model which gave better results than conventional solutions. We obtained 75.41% accuracy with BERT-based fine tuning better than baseline methods LSTM (Long Short-Term Memory) with 41% accuracy for Hindi dataset and for English dataset, we achieved an accuracy of 82.92% compared to 81.57% with baseline models.

Keywords Hate speech · Toxic speech · BERT · LSTM · Cyberbullying · CNN

A. Barai · P. Jain (✉) · T. Kumar
Indian Institute of Information Technology, Nagpur, India
e-mail: poojajain@iiitn.ac.in

T. Kumar
e-mail: tapan.jain@iiitn.ac.in

K. Hirose et al. (eds.), *Proceedings of 27th International Symposium on Frontiers of Research in Speech and Music*, Advances in Intelligent Systems and Computing 1455,
https://doi.org/10.1007/978-981-97-1549-7_1