Xia Lixun (夏 立勋)

Contact

江苏省昆山市玉山镇集街
江南新村 09-301

Zip Code: 215300
Mobile: 13013815427
Email: lixun.xia@outlook.com

Experience

- 2014.6 - present
  Principle Engineer, Tech. Lead, Digital Signal Processing, Harman Automotive Electronic System (Suzhou), Suzhou, China

  - Design and development of light-weight Key Word Spotter (KWS) based on deep neural networks for embedded audio products
  - Design and development of single-mic noise reduction based on deep models
  - Design and development of Voice Activity Detection (VAD) based on deep models for embedded audio products
  - Fully automatic test software for AI speaker products performance benchmarking
  - Development of classical signal processing modules for car audio and embedded audio
  - Acoustic measurement skill (Anechoic, Semi Anechoic, ETSI, ACQUA etc.)
  - Language/Framework skills include: Rust, Python, C++, Julia, Matlab, PyTorch

- 2010.9 - 2014.3
  R&D Engineer, Acosense AB, Göteborg, Sweden

  - Development of feature data collection for real-time online acoustic spectroscopic fluid chemical measurement based on Partial Least Square regression models
  - Development of signal processing algorithms for feature calculation
  - Development of FPGA for algorithm realizations
  - Development of instrument hardware platform
  - Development of industrial field bus communications with SCADA integration
  - In-depth knowledge of CE certification for industrial products

Education

2008.9 - 2010.9 *M.Sc.* Integrated Electronic System Design, Chalmers University, Sweden

2005.0 - 2008.6 *M.Sc.* Power Electronics, Central South University, China

2001.9 - 2005.6 *B.Sc.* Control Science and Engineering, Central South University, China

Projects

- ***Light-Weight Voice Activity Detection (VAD) for Harman Embedded Audio Products (2020 - present)***

  - For MIPS and memory constrained battery powered products
  - A deep model training framework
  - A work in progress project

Conventional VAD tool relies on calculations of speech features like MFCC, periodicity/harmonicity and band energies on frame basis, and apply boosting algorithms (AdaBoost for example) to find out the optimal thresholds for features sorted by importance. This approach is identical to single-layer neural net. Thus is often beat by deep model rivals. With the introduction of deep models and constraints of MIPS and memory cost, light weight models with superior performance would be developed. This is a on-going project and better performance over Web RTC VAD (based on Gaussian Mixture Model) can be expected.

- ***Light-Weight Key Word Spotter (KWS) for Harman Embedded Audio Products (2019 - 2020)***

  - Delivered a generic KWS deep model training framework aiming for portable audio products
  - Data collection and pre-processing for deep model training and validation
  - Feature calculation, deep model training, selection and validation
  - Model inference code for business units

Key word spotter of configurable target word with MIPS and memory constraints are getting more attractiveness to portable audio products, like TWS headphones. Performance per MIPS x Memory is the new ruler for the competition. Deep model training framework that fulfills our specific requirements has been carefully designed and development, in order to generate streamlined deep models with predefined MIPS and memory footprint, at acceptable precision and recall. The value of this framework is in how audio domain knowledge is combined with advanced software skills, on top of the tremendous data collected, to allow me to train a high performance model on an ordinary PC within reasonable amount of time.

- ***Performance benchmark of Samsung Galaxy Home AI Speaker (2018-2019)***

  - Objective measurement of KWS and ASR performance under varsious situations (noise, echo, orientation etc.)
  - Evaluation of product performance according to various standards (Samsung Certificate, MSFT Cortana and Skype, Amazon AVS and Google ART certificate)
  - Responsible for final performance certification

Latest project for Samsung AI group. Responsible for performance evaluations of the microphone preprocessing. This product is well known for its multi-channel AEC (M-AEC), which allows echo cancellation in multichannel playbacks. Its hex-tweeter configuration may render spatial sound effects. Thanks to the Universal AI Speaker Test Software, thorough tests on KWS and ASR accuracies, DOA properties, AEC as well as noise rejection performance are measured and full report can be generated without human in the loop.

- ***Universal AI Speaker Test Software (2018-2019)***

  - Automatic test of ASR, KWS, AEC, DOA and noise reduction performance of virtually any AI speaker products

- Test cases covering Quiet/Noise/Echo/Echo+Noise with arbitrary spatial configurations in anechoic/ETSI room
- Has been applied to numerous products: Alexa Echo/Echo Dot, Harman Kardon Invoke, Apple HomePod, Google Home, Samsung Galaxy Home
- Slim in size and almost no run-time dependencies
- Excellent scale-out capability
- Robustness for unsupervised operations

Build of AI speaker products involves fast iteration of verifications of different signal processing algorithms. Automatic objective measurement of the key performance of the engineering samples is the key. Written in Rust programming language, the universal test tool fulfills such requirements with a minimal run-time and with efficiency. For products without exposure of debug interface, the tool can even utilize camera-based approach to accomplish the measurement tasks!

- ***Single-Channel Noise Reduction Based on Deep Models (2017 - 2018)***

  - Build of high-quality noise data set for training of deep models
  - Delivers state-of-the-art signal-distortion ratio (SDR > 13.5dB)
  - Model aided labelling of extended noise data

Compared to traditional noise reduction algorithms, deep model based approach shows superior performance on non-stationary and transient noise types, for instance, key-board strokes. The existence of single channel also indicates that the NR performance shall solely rely on nothing but knowledge of speech and noise. Therefore a high quality noise data set are carefully collected and pre-processed for training of deep neural nets. Mel-spectrogram as feature input is used to infer the binary/fractional masks from the noisy speech. The masked spectrogram will be synthesized back to time domain as the clean speech. The signal-to-distortion ratio is comparable to state-of-the-art value in industry. The model trained can also be used as a VAD tool for speech/noise labelling tasks.

- ***Algorithm Tuning & Performance Measurement of Microphone Processing for Harman Kardon Invoke (2016 - 2017)***

  - Microsoft far field AI speaker (Cortana) for home use
  - Tuning parameters of mic frontend processing (AEC/BF/BS/ABF/NR/Leveller) for Cortana (KWS/ASR) and Skype standards
  - Hands on ACQUA system for measurements
  - Harman global reward for successful project accomplishment

Microsoft AI speech products are tier one in industry. Skype and Cortana specifications for speakerphone is the de-facto standards for nearly all AI speakers. Insights into the certification have been obtained by means of parameter tuning and performance measurement during the project. ACQUA system was also used for Skype standard certification. Thorough understanding of the certifications urged me to go for an universal solution of AI speaker test and benchmark, which is listed as a subsequent project.

- ***Code Porting (Matlab -> C++): VAD based on Adaoost (2015 - 2016)***

  - Realizations of features (MFCC, Band Energy, Harmonicity etc.) that can separate speech from noise content
  - Realizations of VST for real time showcasing

- Trained with AdaBoost ML algorithm

- ***Code Porting (Matlab -> C++): Voice De-reverberant (2015 - 2016)***

  - Fast online estimation of RT60
  - Determine 'dry'/'wet' energy with short time window and hop size
  - Development of VST plugins for realtime application

- ***Code Porting (Matlab -> C++): Voice Leveller Based on Simple VAD (2015 - 2016)***

  - A simple VAD tool provides frame level voice signal
  - Dynamic gain control to augment speech frames without introducing much distortion
  - Development of VST plugins for realtime application

- ***Code Porting: Engine Order Cancellation --- ANC for Cars (2014-2015)***

  - Implementation of in-vehicle engine noise cancellation
  - Multi-speaker multi-mic configuration
  - Control state machine to allow disturbing events for example window opening
  - Deploying in many commercial car products

Engine noise is mixture of harmonics of the rotor base frequency or RPM Adaptive filters are used to track the transfer function from the noise source to passengers, the estimated paths will be used to generate compensatory sounds from the speakers. The overall effect to every passenger is zones without low-frequency noises.

- ***Power Supply Envelop Control Without DAC (2014 - 2015)***

  - creative method for power control of class-H PSU
  - save the cost of expensive DAC
  - used in amplifier products

- ***ACOspector(TM) – non-invasive fluid measurement in real-time (2010 - 2014)***

Lead the development of the industrial fluid property measurement sensor based on active acoustic spectroscopy measurement. Products have been deployed among pulp and paper, chemical factories in Sweden. Some information can be found in *www.acosense.com*. Contributions:

  - made the first prototype that delivers to customers
  - improved the product to its second generation

- ***Matrix Power Converter Design and Implementation (2007 - 2008)***

Matrix converter can be used as motor drivers(PSM, Induction Motor) thanks to its high power density. This is an academic project for my first graduate thesis. Matrix power converter has the merit of higher energy density compared to traditional back-to-back rectifier-inverter topology. Nevertheless the synchronized control of all semiconductor switchers poses harder control problem against the old. In this National Science Foundation project, theory of matrix converters has been studied and a prototype has been engineered. The novel prototype is based on four-leg structure thus allows for more intuitive implementation in a carrier-modulation style.

This is a national science foundation project thus the deliverables are publications below:

1. Su Mei, Xia Lixun, Sun Yao, et al "*Carrier modulation of four-leg matrix converter based on FPGA*", Electrical Machines and Systems, 2008. ICEMS 2008. IEEE International Conference on. pp.1247-1250.

2. Yao Sun, Mei Su, Lixun Xia, et al "*Randomized carrier modulation for four-leg matrix converter based on optimal Markov chain*", Industrial Technology, 2008. ICIT 2008. IEEE International Conference on. pp.1-6.

3. Hengsi Qin, Mei Su, Lixun Xia, et al "*A novel controller design method for power converters*", IEEE 11th Workshop on Control and Modeling for Power Electronics, 2008. COMPEL 2008. IEEE International Conference on

Language Skills

1. Chinese - native
2. English - proficient

Driver's License - C1