KE CHEN

Audio Researcher · Software Engineer · Computer Musician

⌂ www.knutchen.com
✉ knutchen@ucsd.edu
in KeChen | ⦿ RetroCirce | 🎓 KeChen

EDUCATION

UC San Diego   *California, USA*

· Ph.D Candidate in Computer Science (Expected Graduation Year: 2024)   Sep 2019 - Present

· Ph.D Candidate in Music   Sep 2019 - Present

· M.A. in Music   Sep 2019 - Mar 2021

· Advisors: Prof. Shlomo Dubnov and Prof. Taylor Berg-Kirkpatrick

Fudan University   *Shanghai, China*

· B.E in Computer Science (Software Engineering Track), GPA 3.8/4.0   Sep 2015 - Jun 2019

· Advisors: Prof. Wei Li and Prof. Gus Xia

· Development Expertise: Front-End (Vue.js & React), iOS (Swift), Graphic (OpenGL & WebGL)

RESEARCH INTERESTS

Audio Representation Learning and **Music Information Retrieval**: including music generation, multimodal learning, audio source separation, audio classification, melody extraction, music recommendation, and speech enhancement.

SELECTED PROJECTS

- **Contrastive Language-Audio Pretraining (CLAP)** (⦿ 943 stars): A multi-modal framework that produces both audio and text corresponding representations for broad downstreaming applications.

- **HTS-Audio Transformer** (⦿ 260 stars): Hierarchical Token-Semantic Transformer (HTS-AT), an efficient and light-weight audio classification and event detection model, achieving the state-of-the-art (SOTA) results on AudioSet and ESC-50, and Speech Command V2.

- **POP909** (⦿ 249 stars): A dataset of multiple versions as the piano arrangements of 909 popular songs created by professional musicians, along with algorithms to extract partial labels for music generation and MIR communities.

- **Zero-shot Audio Source Separation** (⦿ 146 stars): A three-component pipeine that allows you to train a audio source separator to separate any source from the track by giving the source sample query. The model lies in a zero-shot setting as we never use the separation dataset but a general audio dataset AudioSet during the training stage.

- **MusicLDM** (⦿ 87 stars): A text-to-music generation model for investigating the novelty and plagiarism issue on the audio diffusion model via the latent mixup strategy.

- **Music SketchNet** (⦿ 76 stars): A controllable music generative model that allows you to specify your own music ideas, namely pitch contour and rhythm, in the monophonic music generation scenario.

- **LAION-Audio-630K** (⦿ 497 stars): An open-source large collection of audio dataset, each containing enormous amount of audio-text pairs, will be eventually processed and used for training CLAP model and others.

- **Selective Speech Enhancement** (Internal Project): A combined system of the speech enhancement model and the conditional source separator that is able to provide the high-fidelity speech track from the noised clip, while preserving the audio event of user specifications.

- **Choral Music Separation** (⦿ 28 stars): An automated pipeline for synthesizing choral music data from sampled instrument plugins within controllable options for instrument expressiveness.

- **TONet** (⦿ 38 stars): A plug-and-play model that improves both tone and octave perceptions by leveraging a novel input representation and a novel network architecture in the task of singing melody extraction.

PUBLICATIONS

🎓 Google Scholar Link | 🗃 DBLP Link | 🔖 ORCID Link | (* indicates the equal contribution)

Audio Representation Learning, Source Separation, Classification, and MIR

[1] **K. Chen**, J. Su, and Z. Jin, "MDX-GAN: Enhancing Perceptual Quality in Multi-Class Source Separation via Adversarial Training", in *ICASSP 2024 (oral presentation)*.

[2] **K. Chen***, Y. Wu*, T. Zhang*, Y. Hui*, T. Berg-Kirkpatrick, and S. Dubnov, "Large-Scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation", in *ICASSP 2023*.

[3] **K. Chen**, G. Wichern, F. G. Germain, and J. L. Roux, "Pac-HuBERT: Self-Supervised Music Source Separation via Primitive Auditory Clustering and Hidden-Unit BERT", in *ICASSP 2023 (SASE Workshop)*.

[4] K. Shao*, **K. Chen***, T. Berg-Kirkpatrick, and S. Dubnov, "Towards Improving Harmonic Sensitivity and Prediction Stability for Singing Melody Extraction", in *ISMIR 2023*.

[5] **K. Chen**, X. Du, B. Zhu, Z. Ma, T. Berg-Kirkpatrick, and S. Dubnov, "Zero-shot Audio Source Separation through Query-based Learning from Weakly-labeled Data", in *AAAI 2022 (oral presentation: 5%)*.

[6] **K. Chen**, X. Du, B. Zhu, Z. Ma, T. Berg-Kirkpatrick, and S. Dubnov, "HTS-AT: A Hierarchical Token-Semantic Audio Transformer for Sound Classification and Detection", in *ICASSP 2022 (oral presentation)*.

[7] **K. Chen**, S. Yu, C. Wang, W. Li, T. Berg-Kirkpatrick, and S. Dubnov, "TONet: Tone-Octave Network for Singing Melody Extraction from Polyphonic Music", in *ICASSP 2022 (oral presentation)*.

[8] **K. Chen**, H. Dong, Y. Luo, J. McAuley, T. Berg-Kirkpatrick, M. Puckette, and S. Dubnov, "Improving Choral Music Separation through Expressive Synthesized Data from Sampled Instruments", in *ISMIR 2022*.

[9] J. Jiang, **K. Chen**, W. Li, and G. Xia, "Large-vocabulary Chord Transcription via Chord Structure Decomposition", in *ISMIR 2019 (oral presentation)*.

[10] Q. Kong*, **K. Chen***, H. Liu, X. Du, T. Berg-Kirkpatrick, S. Dubnov, and M. Plumbley, "Universal Source Separation with Weakly Labelled Data", in *submission for TPAMI*.

[11] N. Srivatsan, **K. Chen**, S. Dubnov, and T. Berg-Kirkpatrick, "Retrieval-Augmented Music Captioning via Multimodal Prefixes", in *submission for a double-blinded conference*.

Music Generation and Recommendation

[12] **K. Chen***, Y. Wu*, H. Liu*, M. Nezhurina, T. Berg-Kirkpatrick, and S. Dubnov, "MusicLDM: Enhancing Novelty in Text-to-Music Generation Using Beat-Synchronous Mixup Strategies", in *ICASSP 2024 (oral presentation)*.

[13] H. Dong, **K. Chen**, S. Dubnov, J. McAuley, and T. Berg-Kirkpatrick, "Multitrack Music Transformer", in *ICASSP 2023 (oral presentation)*.

[14] S. Dubnov, **K. Chen**, and K. Huang, "Deep Musical Information Dynamics: Novel Framework for Reduced Neural-Network Music", in *Journal of Creative Music Systems, JCMS 2022*.

[15] **K. Chen**, B. Liang, X. Ma, and M. Gu, "Learning Audio Embeddings with User Listening Data for Content-based Music Recommendation", in *ICASSP 2021*.

[16] **K. Chen**, C. Wang, T. Berg-Kirkpatrick, and S. Dubnov, "Music SketchNet: Controllable Music Generation via Factorized Representations of Pitch and Rhythm", in *ISMIR 2020*.

[17] **K. Chen***, Z. Wang*, J. Jiang, Y. Zhang, M. Xu, S. Dai, X. Gu, and G. Xia, "POP909: A Pop-song Dataset for Music Arrangement Generation", in *ISMIR 2020*.

[18] **K. Chen**, G. Xia, and S. Dubnov, "Continuous Melody Generation via Disentangled Short-Term Representations and Structural Conditions", in *International Conference on Semantic Computing, ICSC 2020*.

[19] **K. Chen**, W. Zhang, S. Dubnov, and G. Xia, "The Effect of Explicit Structure Encoding of Deep Neural Networks for Symbolic Music Generation", in *Inter. Workshop on Multi-Music Representation and Processing, MMRP 2019*.

Others

[20] H. Liu, **K. Chen**, Q. Tian, W. Wang, and M. Plumbley, "AudioSR: Versatile Audio Super-resolution at Scale", in *ICASSP 2024 (oral presentation)*.

[21] M. Chemillier, **K. Chen**, M. Malt, and S. Dubnov, "A Posthumous Improvisation by Toots Thielemans", in *Toots Thielemans (1922-2016). A Century of Music across Europe and America, 2022*.

[22] X. Du, **K. Chen**, Z. Wang, B. Zhu, and Z. Ma, "ByteCover2: Towards Dimensionality Reduction of Latent Embedding for Efficient Cover Song Identification", in *ICASSP 2022*.

[23] X. Du, H. Liang, Y. Wan, Y. Lin, **K. Chen**, B. Zhu, and Z. Ma, "Latent Feature Augmentation for Chorus Detection", in *ISMIR 2022*.

[24] H. Dong, **K. Chen**, J. McAuley, and T. Berg-Kirkpatrick, "MusPy: A Toolkit for Symbolic Music Generation", in *ISMIR 2020*.

Patents

[25] W. Li and **K. Chen**, "A Robust Digital Audio Watermark Embedding System Based on Constant Watermark", in *Chinese Patent: 2018104457495*.

[26] W. Li, Y. Wu, and **K. Chen**, "An Automatic Detection System for Audio Signal Similarity", in *Chinese Patent: 2018104457546*.

EXPERIENCE

Academic

Center for Research in Entertainment and Learning, UC San Diego — *La Jolla, California, USA*

Music X Lab, NYU Shanghai — *Shanghai, China*

Audio and Music Technology Laboratory, Fudan University — *Shanghai, China*

Center for Computer Research in Music and Acoustics, Stanford University — *Palo Alto, California, USA*

Industry

Adobe Inc. — *California, USA*
Research Intern by Zeyu Jin and Jiaqi Su | Adobe Research — Jun 2023 - Present
· Design high-fidelity audio event and speech separation models for speech-centered acoustic scenarios.
· **Selected in Adobe Max Sneak Japan 2023.**

Mitsubishi Electric Research Laboratories — *Massachusetts, USA*
PhD Intern by Gordon Wichern | Speech & Audio Team — Sep 2022 - May 2023
· Design machine learning models on self-supervised music source separation.

Apple Inc. — *California, USA*
PhD Intern by Venkatraman Atti | Interactive Media Group — Jun 2022 - Sep 2022
· Design machine learning models on audio codecs.
· **Selected as the intern presenter to Craig Federighi, the Apple's senior vice president.**

Bytedance Ltd. — *Shanghai, China*
Research Intern by Xingjian Du and Bilei Zhu | AI Lab Speech & Audio Team — Jun 2021 - Sep 2021
· Design hierarchical audio transformer for audio classification and zero-shot audio source separation.

Tencent Music Entertainment — *Shenzhen, China*
Research Intern by Beici Liang — Jul 2020 - Sep 2020
· Design machine learning models on music recommender systems across QQ Music Platform.

ACADEMIC SERVICE

Invited Talks & Presentations

· Text-to-Music Generation via Latent Diffusion Models | Spotify, USA	Oct 2023
· Contrastive Language-Audio Pretraining | Harmonica.ai, Canada	May 2023
· Contrastive Language-Audio Pretraining | Music X Lab, NYU Shanghai, Shanghai	Mar 2023
· Hierarchical Token-Semantic Audio Transformer | ICASSP 2022, Singapore	May 2022
· Singing Melody Extraction with Tone-Octave Network | ICASSP 2022, Singapore	May 2022
· Zero-shot Audio Source Separation through Query-based Learning | Mila x Vector Institute, Canada	Mar 2022
· Zero-shot Audio Source Separation through Query-based Learning | Bytedance (Tiktok), USA	Jan 2022
· Zero-shot Audio Source Separation through Query-based Learning | AAAI 2023, Online	Jan 2022
· Singing Melody Extraction with Tone-Octave Network | Computer Music Group, CMU, USA	Aug 2021
· Controllable Music Generation via Latent Variable Disentanglement | ISMIR 2020, Online	Nov 2020

Reviewing

· Conference: Neurips 2023, ICML 2022 & 2023, AAAI 2022 & 2023, ICASSP 2022 & 2023, ISMIR 2022 & 2023
· Journal: IEEE JSTSP, EURASIP

Teaching

Teaching Assistant | Music Department, UC San Diego *La Jolla, California, USA*

· Mus 5: Sound in Time	Fall 2021
· Mus 170: Music Acoustics	Spring 2021
· Mus 173: Electronic Music Production	Winter 2021; Fall 2020
· Mus 172: Computer Music II	Spring 2020
· Mus 111: Topics/World Music Traditions	Winter 2020
· Mus 171: Introduction to Computer Music	Fall 2019

SCHOLARSHIPS AND AWARDS

- **National Scholarship: 2016 - 2017** (Top 1 in terms of Overall Evaluation, 0.2% in China)
- **National Scholarship: 2015 - 2016** (Top 1 in terms of Overall Evaluation, 0.2% in China)
- **Golden Award** in *Shanghai A Cappella Competition*: Winter 2016
- First Runner-Up in Fudan iShamrock Software Competition: Mar 2018
- First Prize in Shanghai "Ideal Cup" Modeling Competition: Summer 2017
- Freshman Scholarship: Fall 2015

SKILLS

- **Language**: Native in Chinese; Fluent in English
- **Programming**: Python, JavaScript, Swift-C, C, HTML, Java, PureData, and Matlab
- **Framework**: PyTorch, TensorFlow, Vue.js, and React
- **Digital Audio Workstation**: Cubase, FL-Studio, and Ableton Live
- **Instrument**: Piano (20 years)
- **Creative Production**: Adobe Premiere, Adobe AfterEffect, Adobe Audition, and Adobe Photoshop
- **Hobbies**: composition, video clipping, table tennis, and volleyball