<p style="text-align:center;">**Research Assessment Exercise 2020**
Impact Case Study</p>

University: The Chinese University of Hong Kong
Unit of Assessment (UoA): 12 Electrical & Electronic Engineering
Title of case study: Industrial and commercial impact of spoken language technology

(1)   Summary of the impact (indicative maximum 100 words)

The Chinese University of Hong Kong (CUHK) is inarguably the pioneer and leader in the research of multi-lingual speech technology for improving human-human and human-machine communications. In the past two decades, the CUHK team spearheaded forward-looking and frontier research that addressed diverse needs of the industry, supported technology commercialization, and benefited a large population of end users. The impact of our research encompasses the following aspects:

1) Infrastructure provision – an array of large-scale and multi-purpose speech and language resources that have been widely adopted by the industry;
2) Technology commercialization – award-winning hearing-enhancement devices that are available on commercial markets worldwide;
3) Entrepreneurship – spin-off artificial intelligence start-up focusing on language education.

(2)   Underpinning research (indicative maximum 500 words)

Key members of CUHK research team of spoken language technology include Pak-Chung Ching and Tan Lee from Department of Electronic Engineering, and Helen Meng and Xunying Liu from Department of Systems Engineering and Engineering Management.

Cantonese spoken language technology: Cantonese is a popular Chinese dialect. Cantonese speech has many distinctive characteristics, making it significantly different from Mandarin. For over two decades, the CUHK team has made sustained efforts on advancing Cantonese spoken language technology through fundamental and applied research. The most notable work by Ching, Lee and Meng led to a whole batch of large-scale Cantonese speech databases to support research and development of automatic speech recognition and text-to-speech systems, which were published in 2002 [R1] and made publicly available for technology licensing since then. These databases contain hundreds of hours of transcribed Cantonese speech and cover diverse acoustic conditions. The team's effort was extended to Cantonese-English mixed-language speech [R2], speaker recognition, speech perception, and pathological speech classification. These pioneer works make CUHK be widely recognized as the resources hub of Chinese spoken language research.

Signal processing for hearing enhancement: Hearing loss is one of the most common disorders concerning people of all ages. The direct consequence is to affect the ability of communication and that of appreciating beauty of sounds, i.e., music. Since 2004, Lee and Ching's team at the DSP and Speech Technology Laboratory has carried out inter-disciplinary studies on applying signal processing methods to hearing restoration and rehabilitation, through close collaboration with CUHK Department of Otorhinolaryngology, Head and Neck Surgery. Our research aimed to achieve better hearing experience with various kinds of hearing assistive technology, including cochlear implant, hearing aids and personalized hearing gadgets, which cover users with different degrees of hearing loss. Specifically, we focused on noise reduction, enhancement of language-specific cues, objective assessment of hearing loss, and self-administrated hearing test. During 2012 – 2015, we developed a variety of speech enhancement algorithms [R3] and new objective measures of speech quality. During 2012 – 2016, we investigated and demonstrated the feasibility of performing formal hearing test and personalized hearing enhancement on smartphones for users with early-stage hearing loss [R4].

Computer-assisted language learning: Meng and Liu's team at the Human-Computer Communications Laboratory has been focusing on the problems of recognizing and analysing non-native learners' speech for the purpose of mispronunciation detection and diagnosis (MDD) in computer-aided pronunciation training (CAPT). The research started with cross-language phonological comparison and prediction of common mispronunciation by non-native speakers. To support productive training (i.e., eliciting speech from the learner for analysis), automatic speech recognition techniques were applied to enable detection and diagnosis of targeted pronunciation errors. Specifically, a two-pass framework with discriminative acoustic modelling was developed to significantly improve the detection performance [R5]. Meng's team also presented the earliest attempt to deep learning based mispronunciation detection and successfully applied multi-distribution DNN to robust modelling of different types of mispronunciations [R6]. To support perceptual training, w applied automatic response generation to provide multimodal visualization of production process through text-to-audiovisual speech synthesis.

(3)   References to the research (indicative maximum of 6 references)

[R1]   Tan Lee, W.K. Lo, P.C. Ching and Helen MENG, "Spoken language resources for Cantonese speech processing," Speech Communication 36(3-4), pp.327 – 342, March 2002.
[R2]   Joyce Y. C. Chan, P. C. Ching, Tan Lee and Houwei Cao, "Automatic speech recognition of Cantonese-English code-mixing utterances," Proceedings of INTERSPEECH 2006, pp.113 – 116, September 2006.
[R3]   Feng Huang, Tan Lee, W. Bastiaan Kleijn and Ying-Yee Kong, "A method of speech periodicity enhancement using transform-domain signal decomposition," Speech Communication, vol. 67, pp.102-112, March 2015.
[R4]   Anna Chi Shan Kam, John Ka Keung Sung, Tan Lee, Terence Ka Cheong Wong and Andrew van Hasselt, "Improving mobile phone speech recognition by personalized amplification: application in people with normal hearing and mild-to-moderate hearing loss," Ear and Hearing 38(2), No.2, 2017.
[R5]   Xiaojun Qian, Helen M. Meng, Frank K. Soong, "A two-pass framework of mispronunciation detection and diagnosis for computer-aided pronunciation training," IEEE/ACM Trans. Audio, Speech & Language Processing 24(6), pp.1020-1028, June 2016.
[R6]   Kun Li, Xiaojun Qian, and Helen Meng, "Mispronunciation detection and diagnosis in L2 English speech using multi-distribution deep neural networks," IEEE/ACM Trans. Audio, Speech & Language Processing, vol. 25, Issue 1, pp.193-207, January 2017.

(4)   Details of the impact (indicative maximum 750 words)

Technology Licensing: With 60 million native speakers all around the world, Cantonese is one of the most influential Chinese dialects. In all the best-known speech AI products, e.g., Google Cloud Speech-to-Text, Apple Siri, Microsoft Cortana, the second Chinese dialect being supported after Mandarin must be Cantonese. The Cantonese speech databases developed by CUHK have been a significant driving force in the advancement of Cantonese spoken language technology. Virtually any technology companies, research groups or individuals need to turn to the CUHK speech databases and relevant research papers if they wish to do serious work on Cantonese speech. During 1999-2013, there were 28 industrial user licenses granted to companies locally and internationally, including IBM, Microsoft, Nuance, Philips, Sony, Nokia, Motorola, Toshiba, SmarTone, PCCW, and many others [S1].

In the past five years, there were 6 user licenses granted [S1]. Nexidia Analytics was granted a commercial user license of CUCall (Cantonese telephone speech) in July 2015. Nexidia was a pioneer and market leader in multilingual speech analytics. In January 2016, Nexidia was acquired by NICE Systems (https://www.nice.com/), an NASDAQ listed company and recognized leader in analytics-

based solutions. According to the NICE Nexidia published in 2017: "*Nexidia supports phonetic language packs in more than 25 languages*", which includes Cantonese [S2].

In 2017, Hong Kong Applied Science and Technology Research Institute Company Ltd (ASTRI) licensed our Cantonese speech databases for the development of automatic speech recognition and speaker identification systems. In 2018, ASTRI launched a few new projects on Artificial Intelligence Chatbot, and announced related technologies that support Cantonese/English mixed-language speech interaction (http://astri.dev.onederfo.com/tdprojects/natural-language-processing/). According to Dr. Christina Chan, Director of AI and Big Data Analytics in ASTRI, "*… the lack of Cantonese training data makes it difficult to model our speech recognition engine (SRE). We are fortunate enough to find out about the speech and language resource provided by CUHK*" [S3]. Other recent licensees include Motorola Solutions and NEC. Dr. Kam Hong Shum, Project Director in Cybersecurity, Cryptography and Trusted Technology in ASTRI, expressed that "*the research outputs of the research team of Dr. Tan Lee is very helpful in supporting our research, especially the databased of voice in Cantonese is a good data source for us to train our voice recognition models*."

Commercialization: The collaboration between Lee's team at Electronic Engineering and CUHK Faculty of Medicine led to the invention of a series of software applications and digital gadgets for customized hearing enhancement. The commercialization of these technology products was done through a local company named ACE Communications Limited. In 2014, ACE Hearing, a smartphone apps for customized hearing enhancement, was selected as the Community Impact Winner of the Talent Unleashed Awards (https://www.talentunleashedawards.com/winners/2014-winners/), by a panel of judges comprising Sir Richard Branson and Mr. Steve Woziack. The technology allows smartphone users to self-administer professional-grade assessment of hearing ability and use the assessment results to optimize output sound on their personal devices. It aims to achieve better hearing experience in daily communication and music appreciation for everyone in need, especially those who are in the early stage of hearing impairment.

The software-based technology was soon transformed into a series of hardware-based hearing enhancement devices, including: (i) *AumeoAudio*: electronic gadget for sound customization (https://shop.aumeoaudio.com/products/aumeo-headphones-personalizer); (ii) *Heari*: outdoor-use ear phone for enhanced hearing (https://www.heariaudio.com/); (iii) *MyAudioSession*: headphone for individualized music experience (https://theaudiosession.com/). The products were/are available for purchase online through Amazon and Fortress and retail through HMV and Broadway. In the past five years, over 20,000 devices were sold all over the world. There have been countless reviews and press on AUMEO (see a partial compilation at https://aumeoaudio.com/press/). One of the reviews, written by Glenn Zorpette, executive editor of IEEE Spectrum magazine, was published on IEEE Spectrum Audio/Video Gadget column in July 2016: "*I am perhaps the ideal Aumeo customer. Three decades of scuba diving, and two injuries to my left eustachian tube, have left me with hearing that is idiosyncratic and asymmetric. And, indeed, I did find that the Aumeo let me listen to music more comfortably than I have in many years. With Aumeo, I had the odd and pleasurable sensation of listening to my favorite songs almost as if through young ears again*" [S5]. Since 2015, the AUMEO related technologies were granted 7 patents worldwide.

According to Paul Lee, Co-Founder and CEO of ACE Communications Ltd.: "*ACE Communications was founded with a major technology contribution from the research from Dr. Tan Lee's group on hearing rehabilitation. The contribution propelled ACE into the forefront of democratizing hearing health care for the masses*" [S6].

Spin-off company: Leveraging the research of Meng's team in the area of computer-assisted language learning, SpeechX Limited was founded in 2016. All founding members of SpeechX were former PhD students of CUHK Human-Computer Communications Laboratory (HCCL) directed by Meng. SpeechX is focused on language education powered by artificial intelligence, in order to make language learning more efficient, productive and enjoyable. The company's core technologies include Mispronunciation Detection and Diagnosis (MDD), Text-to-Speech (TTS), dialogue system, which

are largely inspired by the HCCL's research work. According to Kun Li, Founder and CEO of SpeechX, "*The skills acquired and research conducted in their (company founders) PhD thesis research projects have been instrumental in inspiring further development of deployable technologies which form the unique core capability of SpeechX Ltd.*"

SpeechX was recognized as the Top 5 of Internet Industry in China Innovation & Entrepreneurship Competition (out of 28,147 entries) in 2017. SpeechX also received the 1st Prize in Zhongguancun Talent Maker Competition of HK Division and award of the Cyberport Creative Micro Fund and Incubation Program. The company has successfully obtained angel investment of 10 million RMB and pre-A round funding of 40 million RMB. It currently has 32 employers located in Shenzhen and Hong Kong. The major customers include Samsung, Huawei, Roobo, KooLearn, WEBI English, Coelitus, Baicizhan-Xiandan and Baicizhan-zhishipai. The number of daily API calls exceeds 1 million.

Professional take-up: The research by Meng's team on computer-aided pronunciation training (CAPT) has been influential internationally in the area of language education and assessment. In particular, the research has made significant impact on the technology development work at the Educational Testing Service (ETS), the world's largest and most reputable educational testing and assessment organization. According to Dr. Yao Qian, Senior Research Scientist of ETS, "*Professor Meng's research, public talks and publications have provided inspiration in the development of our research prototype, that is currently used by our research work on human-machine interactive language learning.*"

(5) Sources to corroborate the impact (indicative maximum of 10 references)

[S1]   Record of granted user licenses of Cantonese speech databases;
[S2]   Phonetic Search Technology, a Whitepaper by NICE Nexidia;
[S3]   Letter from Dr. Christina Chan, ASTRI
[S4]   Letter from Dr. Kam Hong Shum, ASTRI
[S5]   Review on AumeoAudio by Glenn Zorpette, executive editor of IEEE Spectrum magazine
[S6]   Letter from Mr. Paul Lee, Founder and CEO, ACE Communications Limited
[S7]   Letter from Kun Li, Founder and CEO, SpeechX Limited
[S8]   Letter from Yao Qian, Senior Research Scientist, ETS