Titre de la thèse :

Traitement automatique de la parole en réunion par dissémination de capteurs

Mots –clés : Environnement réverbérant -- Antenne & Beamforming –Traitement du signal — Apprentissage profond – Transcription et reconnaissance du locuteur

Encadrement : Silvio Montrésor (LAUM), Anthony Larcher (LIUM), Jean-Hugh Thomas (LAUM)

Financement : LMAC (Paris scientifiques du LAUM)

Début : Septembre 2018

Contact : jean-hugh.thomas@univ-lemans.fr

Objectif de la thèse

Le sujet est porté par deux laboratoires de Le Mans Université : le laboratoire d'Acoustique (LAUM) et celui d'Informatique (LIUM). L'objectif est l'amélioration de procédés de traitement automatique de la parole en réunion, transcription et reconnaissance du locuteur, en recourant à un dispositif d'enregistrement et de traitement de signaux audio par l'intermédiaire d'une antenne de microphones.

Sujet de la thèse

Il s'agit de concevoir un système « mains libres » capable de localiser les locuteurs dans une salle, de séparer les signaux émis par ces locuteurs et d'améliorer le signal de parole et son traitement.

Les problématiques de la thèse sont les suivantes :

- Définir une géométrie d'antenne adaptée à la prise de son distante avec un nombre de microphones réduit.

- Proposer un traitement permettant d'une part d'exploiter la dimension multi-capteurs de l'acquisition des données et d'autre part de sélectionner les parties du signal audio (ordres de réflexion) les plus à même d'améliorer les performances du système de reconnaissance automatique de la parole du LIUM. Ce traitement devra tenir compte du confinement de l'environnement (salle de réunion). Il fera également appel à des algorithmes de séparation de sources afin d'identifier les différents locuteurs pendant la réunion.

- Proposer des évolutions aux schémas usuels d'extraction de paramètres dans le signal de façon à améliorer leur pertinence pour le réseau de neurones.

- Proposer une stratégie d'apprentissage du réseau de neurones permettant d'améliorer les performances en transcription.

Quelques références

[1] J. H. L. Hansen, T. Hasan, Speaker recognition by machines and humans, IEEE Signal Processing Magazine, 74, 2015.

[2] L. Deng, G. Hinton, B. Kingsbury, New types of deep neural network learning for speech recognition and related applications: An overview, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 8599-8603).

[3] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mhamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, B. Kingsbury, Deep neural networks for acoustic modelling in speech recognition, IEEE Signal Processing Magazine, 82, 2012.

[4] P Bell, MJF Gales, T Hain, J Kilgour, P Lanchantin, X Liu, A McParland, S Renals, O Saz, M Wester, et al.The MGB challenge : Evaluating multi-genre broadcast media recognition. Proc. of ASRU, Arizona, USA, 2015.

[5] T. B. Spalt, Background noise reduction in wind tunnels using adaptive noise cancellation and cepstral echo removal techniques for microphone array applications, Master of Science in Mechanical Engineering, Hampton, Virginia, USA, 2010.

[6] D. Blacodon, J. Bulté, Reverberation cancellation in a closed test section of a wind tunnel using a multi-microphone cepstral method, Journal of Sound and Vibration 333, 2669-2687 (2014).

[7] Q.-G. Liu, B. Champagne, P. Kabal, A microphone array processing technique for speech enhancement in a reverberant space, Speech Communication 18 (1996) 317-334.

[8] S. Doclo, Multi-microphone noise reduction and de-reverberation techniques for speech applications, S. Doclo, Thesis, Leuven (Belgium), 2003.

[9] Y. Liu, N. Nower, S. Morita, M. Unoki, Speech enhancement of instantaneous amplitude and phase for applications in noisy reverberant environments, Speech Communication 84 (2016) 1-14.

[10] Feng, X., Zhang, Y., & Glass, J. (2014, May). Speech feature denoising and dereverberation via deep autoencoders for noisy reverberant speech recognition. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1759-1763). IEEE.

[11] Kinoshita, K., Delcroix, M., Yoshioka, T., Nakatani, T., Sehr, A., Kellermann, W., & Maas, R. (2013, October). The reverb challenge: Acommon evaluation framework

for dereverberation and recognition of reverberant speech. In *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (pp. 1-4). IEEE.

[12] Xiong X., Watanabe S., Erdogan H., Lu L., Hershey J., Seltzer M. L., Chen G., Zhang Y., Mandel M., Yu D., Deep Beamforming Networks for Multi-Channel Speech Recognition, Proceedings of ICASSP 2016, pp 5745-5749.

Le Mans Université

Title of the PhD thesis:

Automatic speech processing in meetings using microphone array

Key words : environment with reverberation– Array & Beamforming – Signal processing – Deep learning – Transcription and speaker recognition

Supervision : Silvio Montrésor (LAUM), Anthony Larcher (LIUM), Jean-Hugh Thomas (LAUM)

Funding: LMAC (Scientific bets of Le Mans Acoustique)

Beginning : September 2018

Contact : **jean-hugh.thomas@univ-lemans.fr**

Aim of the PhD thesis

The subject is supported by two laboratories of Le Mans – Université: the acoustics lab (LAUM) and the computer science lab (LIUM). The aim is to enhance automatic speech processing in meetings, transcription and speaker recognition, by using a recording device and audio signal processing from a microphone array.

Subject of the PhD thesis

It consists in implementing a hands-free system able to localise the speakers in a room, to separate the signals emitted by these speakers and to enhance the speech signal and its processing.

The thesis' issues are the following:

- Define an array geometry adapted to distant sound recording with few microphones.

- Propose processing able to take advantage of the acoustic data provided by the array and to select the parts of the audio signals (reflexion orders) the most relevant for enhancing the performance of the automatic speech recognition system of the LIUM. The process should take into account the confined environment (meeting room). It will also use source separation algorithms to identify the different speakers during the meeting.

- Propose new development to the usual methods to extract features from the signal to enhance the relevance for the neural network.

- Propose a learning strategy for the neural network to enhance the transcription performance.

Some références

[1] J. H. L. Hansen, T. Hasan, Speaker recognition by machines and humans, IEEE Signal Processing Magazine, 74, 2015.

[2] L. Deng, G. Hinton, B. Kingsbury, New types of deep neural network learning for speech recognition and related applications: An overview, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 8599-8603).

[3] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mhamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, B. Kingsbury, Deep neural networks for acoustic modelling in speech recognition, IEEE Signal Processing Magazine, 82, 2012.

[4] P Bell, MJF Gales, T Hain, J Kilgour, P Lanchantin, X Liu, A McParland, S Renals, O Saz, M Wester, et al.The MGB challenge : Evaluating multi-genre broadcast media recognition. Proc. of ASRU, Arizona, USA, 2015.

[5] T. B. Spalt, Background noise reduction in wind tunnels using adaptive noise cancellation and cepstral echo removal techniques for microphone array applications, Master of Science in Mechanical Engineering, Hampton, Virginia, USA, 2010.

[6] D. Blacodon, J. Bulté, Reverberation cancellation in a closed test section of a wind tunnel using a multi-microphone cepstral method, Journal of Sound and Vibration 333, 2669-2687 (2014).

[7] Q.-G. Liu, B. Champagne, P. Kabal, A microphone array processing technique for speech enhancement in a reverberant space, Speech Communication 18 (1996) 317-334.

[8] S. Doclo, Multi-microphone noise reduction and de-reverberation techniques for speech applications, S. Doclo, Thesis, Leuven (Belgium), 2003.

[9] Y. Liu, N. Nower, S. Morita, M. Unoki, Speech enhancement of instantaneous amplitude and phase for applications in noisy reverberant environments, Speech Communication 84 (2016) 1-14.

[10] Feng, X., Zhang, Y., & Glass, J. (2014, May). Speech feature denoising and dereverberation via deep autoencoders for noisy reverberant speech recognition. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1759-1763). IEEE.

[11] Kinoshita, K., Delcroix, M., Yoshioka, T., Nakatani, T., Sehr, A., Kellermann, W., & Maas, R. (2013, October). The reverb challenge: Acommon evaluation framework for dereverberation and recognition of reverberant speech. In *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (pp. 1-4). IEEE.

[12] Xiong X., Watanabe S., Erdogan H., Lu L., Hershey J., Seltzer M. L., Chen G., Zhang Y., Mandel M., Yu D., Deep Beamforming Networks for Multi-Channel Speech Recognition, Proceedings of ICASSP 2016, pp 5745-5749.