Le Mans Université

RESUMÉ et MOTS CLÉS

Pour la diffusion sur le *web*

TITRE EN FRANÇAIS : Traitement automatique de la parole en réunion par dissémination de capteurs

Résumé en français :

Ces travaux de thèse se concentrent sur le traitement automatique de la parole, et plus particulièrement sur la diarisation en locuteurs. Cette tâche nécessite de segmenter le signal afin d'identifier des évènements tels que la présence de parole, de parole superposée ou de changements de locuteur. Cette recherche se focalise sur le cas où le signal est capté par un dispositif placé au centre d'un groupe de locuteurs, comme lors de réunions. Ces conditions entraînent une dégradation de la qualité des signaux en raison de l'éloignement des sources sonores (parole distante). Afin de pallier cette dégradation, une approche consiste à enregistrer le signal à l'aide d'un ensemble de microphones formant une antenne acoustique. Le signal multicanal obtenu permet d'obtenir des informations sur la répartition spatiale du champ acoustique. Deux axes de recherche sont explorés pour la segmentation de la parole à l'aide d'antecnnes de microphones. Le premier axe introduit une méthode combinant des caractéristiques acoustiques avec des caractéristiques spatiales. Un nouveau jeu de caractéristiques, basé sur le formalisme des harmoniques circulaires, est proposé. Cette approche améliore les performances de segmentation en conditions distantes, tout en réduisant le nombre de paramètres des modèles et en garantissant une certaine robustesse en cas de désactivation de certains microphones. Le second axe propose plusieurs approches de combinaison des canaux en utilisant des mécanismes d'auto-attention. Différents modèles, inspirés d'une architecture existante, sont développés. La combinaison de canaux améliore également la segmentation en conditions distantes. Deux de ces approches rendent l'extraction de caractéristiques plus interprétable. Les systèmes de segmentation de la parole distante proposés améliorent également la diarisation en locuteurs. La combinaison de canaux montre une faible robustesse en cas de changement de géométrie de l'antenne en phase d'évaluation. Pour y remédier, une procédure d'apprentissage est proposée, qui améliore la robustesse en présence d'une antenne non conforme. Finalement, les travaux menés ont permis d'identifier un manque dans les jeux de données publics disponibles pour le traitement automatique de la parole distante. Un protocole d'acquisition est introduit pour l'acquisition de signaux en réunions et intégrant l'annotation de la position des locuteurs en plus de la segmentation. En somme, ces travaux visent à améliorer la qualité de la segmentation de la parole distante multicanale. Les méthodes proposées exploitent l'information spatiale fournie par les antennes de microphones en garantissant une certaine robustesse au nombre de microphones disponibles.

MOTS-CLÉS en français (8 maximum) :

1 parole distante
2 antennes de microphones
3 segmentation automatique de la parole
4 diarisation en locuteur
5 apprentissage profond
6
7
8

TITRE EN ANGLAIS : Automatic Speech Processing in Meetings using Microphone Arrays

Résumé en anglais :

(max 1700 words)

This thesis work focuses on automatic speech processing, and more specifically on speaker diarization. This task requires the signal to be segmented to identify events such as voice activity, overlapped speech, or speaker changes. This work tackles the scenario where the signal is recorded by a device located in the center of a group of speakers, as in meetings. These conditions lead to a degradation in signal quality due to the distance between the speakers (distant speech). To mitigate this degradation, one approach is to record the signal using a microphone array. The resulting multichannel signal provides information on the spatial distribution of the acoustic field. Two lines of research are being explored for speech segmentation using microphone arrays. The first introduces a method combining acoustic features with spatial features. We propose a new set of features based on the circular harmonics expansion. This approach improves segmentation performance under distant speech conditions while reducing the number of model parameters and improving robustness in case of change in the array geometry. The second proposes several approaches that combine channels using self-attention. Different models, inspired by an existing architecture, are developed. Combining channels also improves segmentation under distant speech conditions. Two of these approaches make feature extraction more interpretable. The proposed distant speech segmentation systems also improve speaker diarization. Channel combination shows poor robustness to changes in the array geometry during inference. To avoid this behavior, a learning procedure is proposed, which improves the robustness in case of array mismatch. Finally, we identified a gap in the public datasets available for distant multichannel automatic speech processing. An acquisition protocol is introduced to build a new dataset, integrating speaker position annotation in addition to speaker diarization. Thus, this work aims to improve the quality of multichannel distant speech segmentation. The proposed methods exploit the spatial information provided by microphone arrays while improving the robustness in case of array mismatch.

MOTS-CLÉS en anglais (8 maximum) :

1 distant speech
2 multichannel audio
3 automatic speech segmentation
4 speaker diarization
5 deep learning
6
7
8

NOM et Prénom du doctorant
MARIOTTE Théo

Visa du Directeur de thèse