Remeeting — Get More Out of Meetings

Arlo Faria, Korbinian Riedhammer
Remeeting, Inc.
{arlo,korbinian}@remeeting.com

Abstract
Remeeting is a tool that helps you get more out of in-person meetings. Calendar integration and a special email address allow users to email agenda items prior to a certain meeting. A discrete notification at the time of the meeting reminds the user to start the recording. During the meeting, the user focuses on the conversation, or can add notes and photos if desired. After the meeting, every participant gets notified by an automated email that lists the participants along with automatically extracted keywords, notes and photos. This stimulates collaboration, and keeps follow-up contributions at a central place: Just reply to add further notes to the meeting. The resulting meeting “document” can be shared with others and reviewed using a web app that acts as a visual index to the meeting. This makes Remeeting the perfect tool for regular group meetings, standups and interviews, where people typically track progress and follow up on. Remeeting is leveraging, promoting and contributing to open source projects including kaldi and docker.

Index Terms: speaker identification, speech recognition, automatic summarization, productivity.

1. Mobile Apps
The mobile app (see Fig. 1 for Android) is designed for easy-to-use productivity and offers basic recording and note taking functionality. It can record a new meeting, or list previously recorded conversations. While recording, users can mark an interesting point in the conversation, add a textual note, or take a photo (such as of the current whiteboard in a room). Depending on the device, the audio is recording using dual microphones, which is useful for computing time-delay of arrival for speaker separation.

2. Web App
2.1. Cloud-Based Service
Meetings are stored on secure cloud storage (AWS S3), including audio recordings and metadata associated with meetings such as notes, documents or photos. This storage also contains user-specific speech processing parameterizations, including speaker-specific acoustic models and topic-adapted language models.

A pool of compute nodes is allocated dynamically in accordance with current demand (meetings tend to occur predictably, scheduled during regular business hours). In response to subsequent user-generated updates, such as labeling an unknown speaker or uploading relevant documents, a meeting may be enqueued for reprocessing to improve the accuracy of speaker identification and speech recognition algorithms. (See also Sec. 2.6)

2.2. Audio Processing
Following the upload to the cloud, audio recordings from a meeting are automatically enhanced using state-of-the-art denoising and patent-pending audio post-processing techniques. Depending on the capabilities of the smartphone and the number of devices present in a meeting, a combination of noise filtering, automatic gain control, and signal beamforming (from both stereo recordings as well as ad-hoc arrays of multiple devices) can be used to extract high-fidelity audio. The resulting audio not only benefit the user’s listening experience during playback, but also improves the performance of the downstream speech processing technologies.

2.3. Speaker Identification and Linking
Speaker identification (SID) helps to determine who spoke when, and what was contributed. In contrast to the “classic” diarization problem, we assume a much more realistic scenario where a lot of prior knowledge can be leveraged: The calendar typically provides guesses on the participants, which can be further augmented by the users’ social circal and professional organization. It is worth mentioning that the error metric for this problem is actually not DER: for human dialog analysis, it is much more valuable to find coherent dialog acts of certain speakers as it helps to catch the keyphrases reliably.
2.4. Automatic Speech Recognition and Indexing

Remeeting does not use or provide 1-best transcription output; our initial studies suggest that users are unable to leverage verbatim transcripts, mostly due to the high error rates for meeting speech recognition as well as the disfluent nature of conversations. Instead, a keyword search index is constructed from the lattice data structures. Acoustic rescoring and and semantic query expansion are used to improve the performance of keyword search. Implementing search as an interactive process can significantly improve the user experience: results are returned with low-latency response, exhibiting higher recall relative to precision, and displaying contextual “snippets” around potential search hits to guide users through the recording.

In total, a multi-pass decoding strategy is applied, with several adaptation steps that incorporate channel, speaker and topic information. A tremendous improvement in search accuracy is achieved by biasing language models to a known topic, and incorporating otherwise out-of-vocabulary search terms in the recognition grammar and indexed lattices.

2.5. Take a Tour

While the speech processing is not yet enabled by default, feel free to visit https://remeeting.com/tour to explore the review interface.

2.6. Orchestration

Remeeting leverages the open source projects kaldi\(^1\) and docker\(^2\). While kaldi provides state-of-the-art speech recognition algorithms, docker allows to deploy worker nodes with complex workflows to AWS’s elastic cloud. We are working towards a public release of a kaldi docker image that allows to easily train and deploy a kaldi system.

3. Outlook

Remeeting is exploring ways to provide privacy preserving services to customers: Using patent-pending technologies, Remeeting will provide automatic transcriptions without knowledge of the individual spoken word [1, 2].

4. References


\(^{1}\)http://kaldi.sf.net

\(^{2}\)https://www.docker.com