**Literature Review**

**General Papers**

Yusuke Fujita et al., End-to-End Neural Speaker Diarization with Self‑attention, 2019. *(Proposes EEND, a diarization architecture that handles speaker overlap using self-attention rather than clustering)*

Masahiro Makishima et al. / Microsoft, Continuous Speech Separation: Dataset and Analysis, 2020. *(Defines CSS task and introduces LibriCSS dataset, which simulates realistic partial-overlap meetings)*

Soumi Maiti et al., EEND‑SS: Joint End‑to‑End Neural Speaker Diarization and Speech Separation for Flexible Number of Speakers, 2022. *(A joint framework combining diarization, separation, and speaker counting)*

Thilo von Neumann et al., Meeting Recognition with Continuous Speech Separation and Transcription‑Supported Diarization, 2023. *(Describes modular pipeline (TF‑GridNet → ASR → diarization) evaluated on LibriCSS benchmarks)*

Peter Vieting et al., Combining TF‑GridNet and Mixture Encoder for Continuous Speech Separation for Meeting Transcription, 2023.

Expands TF‑GridNet with a mixture encoder to improve long-meeting separation and ASR accuracy.

Dongmei Raj et al., Integration of Speech Separation, Diarization, and Recognition for Multi‑Speaker Meetings, SLT 2021. *(Presents a diarization-first pipeline (e.g. CHiME‑6) that guides separation with diarization cues. referenced in plenty of literature on hybrid pipelines)*

**Tech-Stack Specific Papers**

TF‑GridNet & CSS Models

Zhong‑Qiu Wang et al., TF‑GridNet: Integrating Full‑ and Sub‑Band Modeling for Speech Separation, IEEE/ACM TASLP 2023. *(The core two‑speaker (and multi-channel) separator architecture powering current CSS pipelines)*

Modular Pipeline on Real Meetings (LibriCSS)

Thilo von Neumann et al., Meeting Recognition with Continuous Speech Separation ..., 2023.

Demonstrates TF‑GridNet + diarization + ASR pipeline achieving SOTA cpWER on LibriCSS.

Peter Vieting et al., Combining TF‑GridNet and Mixture Encoder ..., 2023.

Enhances modular pipelines with a mixture-encoder to mitigate separation artifacts in long meetings.

End-to-End/Hybrid Model Approaches

Soumi Maiti et al., EEND‑SS: Joint End‑to‑End Neural Speaker Diarization and Speech Separation ..., 2022.

Integrates diarization, separation, and speaker counting in a unified model for flexible speaker count.

Overlapped Speech Detection / Speaker Extraction

Though not explicitly covered earlier, consider adding works on Target Speaker Extraction (e.g. SpeakerBeam, VoiceFilter, TF‑GridNet conditioned variants), and overlap detection via pyannote’s overlapped-speech-detection model.