Are You Talking to Me? Determining Conversation Partners from Voice Activity

Eline Borch Petersen* and Asger Ougaard

*WS Audiology, Lynge, Denmark. *Eline.Petersen@WSA.com*

Introduction

Conversations consist of rapid and fluent turns, exchanged with precise timing [1].
However, conversing in noise or having impaired hearing (HI) reduce the turn taking precision [2-3].

It has been proposed to use the timeliness of turn takings to determine who, among multiple potential partners, the conversation partner of a target talker is (2-talker conversations) using the minimum overlap-gap method [4].

The minimum overlap-gap (MOG) method:
A speaker has higher probability of being the conversation partner of a target talker if:

I) The combined (binary) voice activity has minimum overlap and gap and

II) All other potential partners exhibit more overlap and gap

Using voice activity of conversations from six datasets, the current study aims to:

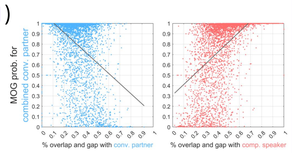
1) Explore if the MOG method is applicable beyond 2-talker conversations.
2) Investigate if noise, number of conversation partners, HI and hearing aid amplification affects the MOG probability.
3) Examine the performance of an automatic implementation of the MOG method.

Table 1: Overview of conversational data.

Dataset		Talkers x groups		Conv. Type	Conditions	Duration
Danpass	[5]	NH-NH	x 11	Map trail	Quiet	616 min
ConvDTU	[6]	NH-NH	x 19	Diapix	Quiet and noise, L1 and L2	2736 min
ConvDyn	[2]	NH-HI	x 11	Diapix	Quiet and noise, HA on and off	1056 min
ConvQ	[7]	NH-HI	x 16	Diapix	Quiet, HA on and off	1344 min
Triad	[3]	NH-NH-HI	x 25	Free	Low noise(quiet), HA on and HA off, High noise, HA in omni- and directional mode	4500 min
Quartet	[8]	NH-NH-HI-HI	x 18	Free	Quiet, middel, and high noise, HA on and HA off	2592 min

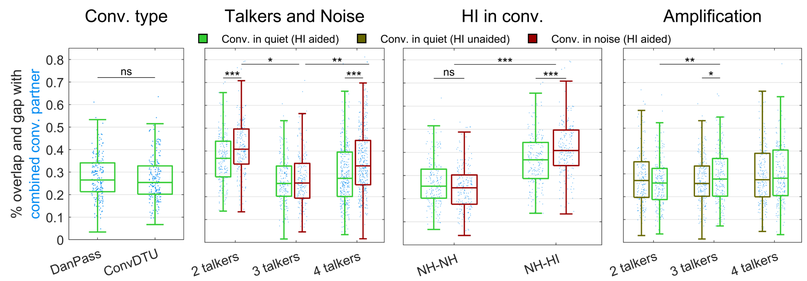
2) Effects of Conditions on Speech Overlap and Gap

Data was extracted for the relevant datasets and conditions (see Table 1)



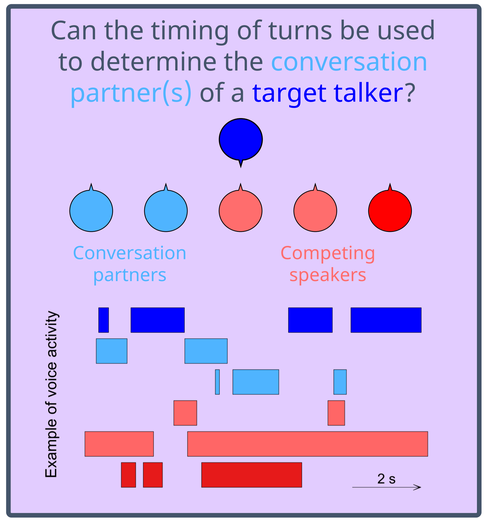
The MOG probability depends on the choice of competing speakers:

Condition effects focus on the % overlap and gap between the target talker and the combined conversation partners



↑ % overlap and gap ← ↑ in noise relative to quiet (2 and 4 talker conversations)
↑ when a HI interlocutor is participating in the conversation
↑ when the HI interlocutor is aided relative to unaided (3 talkers)

Performance of MOG approach could be affected by listening conditions and talkers



Method

- Unlike [4], no hearing-aid system was simulated. Ideal voice activity detection (VAD), from manual of automatic detection, was used for all datasets.
- All VADs were down-sampled to 100 Hz and analyzed in segments of 10 s.
- In 1) and 2) competing speakers were extracted from another segment of the conversation partners VAD to ensure similar speech activity between the conversation partner and the competing speaker.
- All speakers were active within the time window (no silent/inactive speakers).
- When combining the speech of all conversation partners, the resulting VAD was binarized.
- All bar graphs indicate the mean result with lines showing the variance.

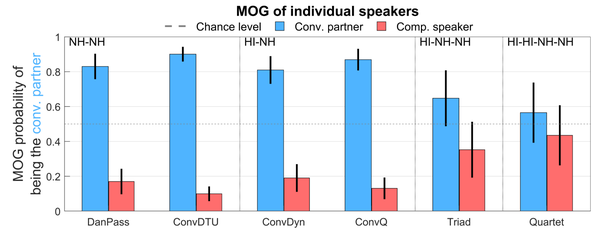
Conclusion

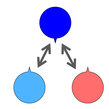
- The MOG approach works well for 2-talker conversations, but the performance decrease with a higher number of conversation partners (Figures in 1A+1B) and when communication difficulties are experienced (Figure in 2).
- An automatic MOG procedure result in low detection of the correct speaker combination beyond NH-NH conversation (Figure in 3)
- More work is needed to understand how overlaps and gaps manifests in different conversations depending on the type, communication difficulties and number of conversation partners participating.

1) Are you Talking to Me or Us ?

1A Can the conversation partner(s) be determined by using MOG on individual speakers?

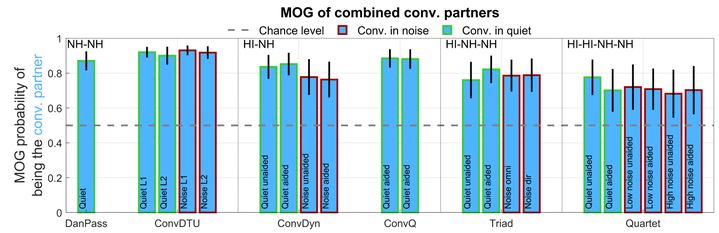
- 1000 10-s VAD segments from a target talker and one conversation partner across conditions
- One competing speaker segment chosen from the same conversation partner at another time

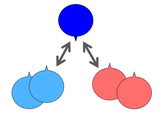




1B Can the conversation partners be determined when combining their VAD?

- 200 10-s segments of a target talker and all conversation partners summed from each condition
- One competing speaker segment chosen from same combined speech signal

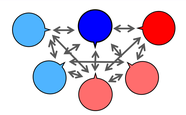




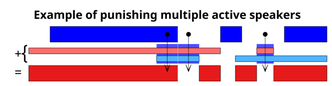
MOG does not work for individual speakers beyond 2-talker conversations. When combining the convonversation partners, the MOG probabilty improves, but differacross datasets and conditions.

3) Automatic Detection of Conversation Partners

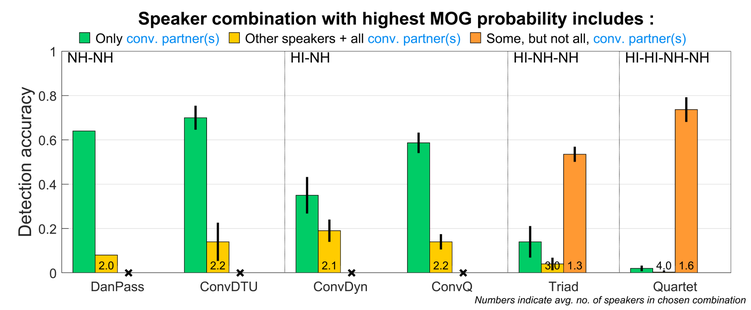
Automatic MOG approach when the no. of conversation partners are unknown :



- Include 5 potential conversation partners selected from different datasets, but always including all speakers within that conversation if possible.
- Create all 31 speaker combinations by adding VADs.
- Multiple active speakers in the combined VAD was punished by setting the resulting VAD = target talker, as illustrated here:



- MOG probability computed for all 31 speaker combinations.
- The combination with highest MOG prob. is considered the conversation partner(s).



Detection accuracy is surprisingly low except for NH-NH conversations.

Explanations include speakers not talking much, backchannels creating overlaps, and that overlaps between multiple speakers were not punished appropriately.

[1] Levinson and Torreira (2015): *Timing in turn-taking and its implications for processing models of language*. Front. Psychol. 10.3389/fpsyg.2015.00731. [2] Petersen et al (2022): *The Effects of Hearing-Aid Amplification and Noise on Conversational Dynamics Between Normal-Hearing and Hearing-Impaired Talkers*. Trends in Hearing. 10.1177/23312165221103340. [3] Petersen (2024): *Investigating conversational dynamics in triads: Effects of noise, hearing impairment, and hearing aids*. Front. Psychol. 10.3389/fpsyg.2024.1289637. [4] Hoang et al (2022): *The Minimum Overlap-Gap Algorithm for Speech Enhancement*. IEEE Access. 10.1109/ACCESS.2022.3147514 [5] Grønnum (2009): *A Danish Phonetically Annotated Spontaneous Speech Corpus (DanPASS)*. Speech Communication 10.1016/j.specom.2008.11.002. [6] Sørensen (2021): *PDF and PSD files of DiapixDK picture materials - adapted Danish version*. 10.5281/zenodo.4638693 [7] Petersen (in press): *Speak Up: How Hearing Loss and the Lack of Hearing Aids affect Conversations in Quiet*. JSLHR. [8] Nicoras et al (2023) *Small Group Conversations: Communication Behaviour and Success*. Proceed. ForumAcusticum 2023. 10.61782/fa.2023.797

WSAudiology