

Evaluation of Paralinguistic-Aware Spoken Dialogue Systems using Next-Utterance Classification

Kouki Miyazawa and Yoshinao Sato



Fairy Devices

Motivation

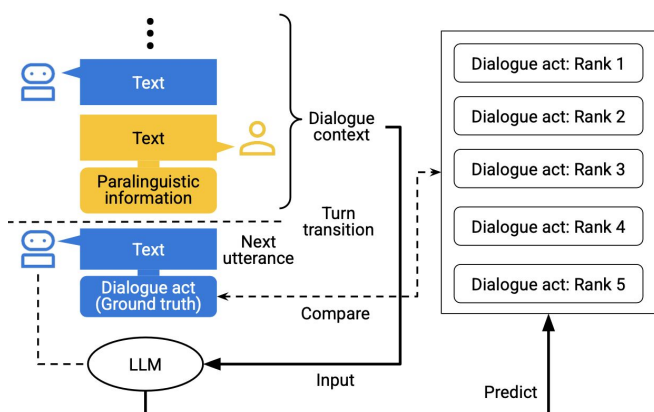
Conventional spoken dialogue systems (SDSs) rely solely on text, causing dialogue breakdowns due to the neglect of **paralinguistic cues**.

Paralinguistic modules is crucial, however their **effectiveness** is **costly to measure** and **hard to disentangle** from other factors.

Method

Corpus-based offline evaluation

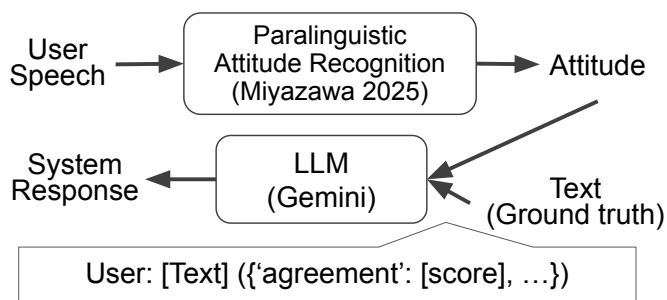
The system is evaluated via the dialogue act (DA) prediction for the next utterance.



DA rankings by the system are assessed based on

1. actual DAs in a spoken dialogue corpus
2. rankings by humans
3. preference test

System



Attitude	Definition
Agreement	in favor, accept to continue
Disagreement	against, dissatisfied, request to stop
Question	not understand, confirm facts, listen back
Stalling	thinking, worried, request to wait

Data

MapTask corpus: 23 h, 128 dialogues, 64 participants

DAs of 7,532 utterances were annotated using 11-class DA taxonomy by Araki (1999)

Experiments

We compared **three conditions**:

Paralinguistic-Agnostic	Paralinguistic-Perturbed	Paralinguistic-Aware
text only	+ random attitudes	+ inferred attitudes

Exp 1: Comparison with Actual DA

The actual DA in MapTask is regarded as correct.

	MRR (top-5)	Accuracy (top-1)
Agnostic	0.634	0.463
Perturbed	0.647	0.476
Aware	0.663	0.496

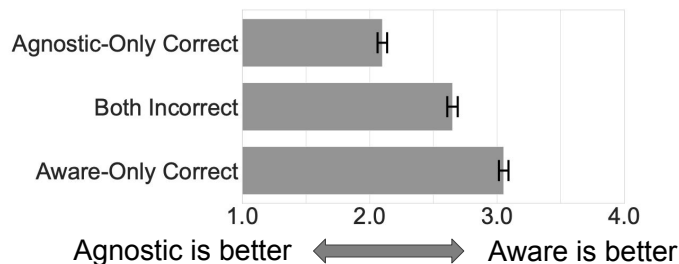
Exp 2: Comparison with Human Rankings

Three experts ranked the DAs for the next utterance based on the past dialogue history.

	NDCG	Kendall's Tau
Agnostic	0.829	0.482
Perturbed	0.844	0.503
Aware	0.860	0.531

Exp 3: Comparison with Preference Test

34 subjects rated which generated system response was more natural on a 1–4 scale.



Summary

We proposed **a corpus-based evaluation method via next-utterance DA prediction**.

The paralinguistic-aware system outperformed.

Our method

serves as a reliable proxy for human assessments
is applicable to other LLM-based SDSs.

Future Work

- Application to open-ended & chit-chat SDSs
- Exploration of better DA taxonomies
- Incorporation of diverse paralinguistic cues

Dialogue Acts

Suggestion/Request, Proposal/Invitation, Confirmation, Yes-No Question, Wh-Question, Promise/Offer/Wish, Inform, Gratitude/Apology, Positive/Accept, Negative/Reject, Answer