Speech Technology Seminar - 1 June 2016

TURN-TAKING IN A TWO PERSON DIALOGUE SYSTEM

Khooshal Saurty

OUTLINE

- Introduction
- What is Turn-Taking?
- Factors affecting turn-taking
  - Cues, Pauses, timing of turn taking, overlapping
- The oscillator model of the timing of turn-taking
- Conclusion and References

INTRODUCTION

- No specific set of rules in a two person dialogue system.
- Different from other forms of speech exchange
  - Debates / Ceremonies
  - Rules for
    - Order of speaking
    - Length of a turn
    - (some cases) Context of the speech.

WHAT IS TURN-TAKING?

- Participants take turns to express themselves.
- At any point : 1 speaker - 1 Listener
- Speaker: Right / Obligation to speak
- Listener: Right / Obligation to attend to the speaker

SACKS' IDEA OF TURN-TAKING

- Turns consist of
  - Turn-Constructional Units (TCU)
    - Part of sentence / A sentence
  - Transition Relevance Place (TRP)
    - Located at the end of the TCU that completes an action.
    - Place where turn-taking can properly occur

TWO PERSON DIALOGUE SYSTEM

- Conversation managed locally by participants
- Participants set their own properties of turn-taking
- Each conversation has different properties.
    - The length of the turn
    - Who speaks when
    - What should be spoken
    - etc…

FACTORS AFFECTING TURN-TAKING

- Timing of turn-taking
- Between speaker silences
- Cues
  - Visual (gestures, facial expressions)
  - Audible (audible inbreath, interjected words)
  - Adjustment to cues
- Overlapping / Simultaneous talk

TIMING OF TURN-TAKING

- Short duration of transition between turns
  - Quite common to have no-gap transitions.
- Temporary hesitation to take a turn in quite short gaps
- Quite clearly, turn-taking does not proceed by the listener waiting for the speaker to be completely silent before taking the turn.
- Listener already estimates the upcoming TRP and make the required preparations for speech well before the TRP occurs.

BETWEEN SPEAKER SILENCES

- occurs at the TRP
- different possibilities to allocate next turn.
- Sacks et al. model of allocation of turn
  - Current speaker selects next speaker
  - Listener selects himself
  - Current speaker selects himself

CUES

- Produced by Speaker
- to indicate an upcoming TRP
- Visual Cues:
  - Eye Gaze, Body Movement
- Audible Cues:
  - semantic, syntactic and prosodic

- Produced by Listener
- to indicate intention to take turn
- Visual Cues:
  - Movements
- Audible Cues:
  - audible inbreath, interjected words

CUES

- adjustment to cues from listener
- Speaker adjusts speech production
  - Rush-through - End the current turn
  - Interjection of words - To retain the turn

OVERLAPPING / SIMULTANEOUS TALK

- Not a common occurrence.
- More than 1 participant taking a turn
    - Solution : One participant drops out.

THE OSCILLATOR MODEL OF THE TIMING OF TURN-TAKING

- Endogenous oscillators (in human brain) are involved in turn-taking

- Property of oscillators:

  - When allowed to influence each other -> Phase locked

- Periodicity of oscillator in listener synchronises with that of speaker.

ASSUMPTIONS

- Timing of turn-taking based on oscillating function of readiness to take the turn (in both speaker and listener).
- Frequency of oscillation determined by speaker's syllable rate
- Listener also engages in oscillator based cycle of readiness to initiate a syllable
- Listener's cycle is counter phased to that of speaker.
- If listener does not speak in first cycle following previous speaker's completion, oscillators continue to be in sync for short time then drift apart

MODEL: SPEAKER

- After the last syllable in a TCU, speaker can still want to convey more ideas - (introduce new TCUs)

- If brief moment of silence (no initiation of further syllables)
  - Speaker in “Mid-Syllable” state
  - Wait until next peak of readiness cycle to initiate speech

MODEL: LISTENER

- If an approaching TRP is detected -> probability of listener initiating speech is maximal when
    - Speaker produces final syllable
        - or in first half cycle after speaker has finished.
- Gap during turn switching is brief (Can sound like no-gap)
- Example: In a model where the rate of syllable is 150ms
    - Initiating speech at 75ms before/after completion of final syllable

CONCLUSION

- Various factors affect turn-taking
- Some quite debatable factors
  - e.g Visual cues - Not present in telephone conversations but still turn taking occurs
- Wilson & Wilson base themselves on a combination of syllable production and oscillators in participants to determine TRP.

REFERENCES

[1] M. Wilson, and T. P. Wilson, “An oscillator model of the timing of turn-taking.” in *Psychonomic bulletin & review*, vol. 12, no. 6, pp. 957-968, 2005.

[2] D. Jurafsky, and J. H. Martin, “Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition.” Prentice-Hall, 2000.

[3] H. Sacks, E. A. Schegloff, and G. Jefferson, “A simplest systematics for the organization of turn-taking for conversation.” in *Language*, vol. 50, no. 4, pp. 696-735, 1974.

[4] T. P. Wilson, and D. H. Zimmerman, “The structure of silence between turns in two-party conversation.” in *Discourse Processes*, vol. 9, no. 4, pp. 375-390, 1986.