# Turntaking and Backchanneling

Linguistics 575
Shannon Watanabe
May 22, 2013

# Outline

- Turn-taking
  - Background
  - Early Research
  - Recent Efforts
    - A Bidding Approach to Turn-taking
    - A Finite-state Turn-taking Model
- Backchannels
  - Background
  - A Shallow Model of Backchannel Continuers in Spoken Dialogue

# Why do we care about turn-taking?

- It's a challenge
  - ASR and TTS perform satisfactorily (in general), but stilted turn changes keep the experience from feeling natural
- Many current systems: release-turn approach to turn-taking
  - System waits until user has completed utterance
  - Turn completion measured by pause threshold
    - Typically 500-1000ms
- Handling different turn options
  - Taking a turn
  - Keeping a turn
  - Releasing a turn

# Early Work

## Sacks et al. (1974)

- Most turn changes in dialog occur with little or no gap or overlap ("smooth switches")
- Turn changes can occur at **Transition Relevant Places (TRPs)**
  - TRPs have governing rules;

    (a) the current speaker (CS) can select someone to speak next, and this person must speak next.

    (b) if CS does not select the next speaker, then anyone may take the next turn;

    (c) if no one else takes the next turn, then CS may take the next turn.
  - TRPs are highly predictable by syntax.

# Early Work

## Duncan (1972-5), Duncan and Fiske (1977)

- Behavioral clues for turn endings:
  - Any phrase-final intonation other than a sustained, intermediate pitch level
  - A drawl on the final syllable of a terminal clause
  - The termination of any hand gesticulation - other work has extended this to cover gesture and gaze
  - A stereotyped expression like 'you know'
  - A drop in pitch and/or loudness in conjunction with a stereotyped expression
  - The completion of a grammatical clause
- Linear correlation between number of signals and likeliness of turn ending

# Early Work: issues

- Early studies looked at human dialogue, face-to-face
  - No gestures/gaze in most SDS
- Conclusions more observations and impressions than the result of objective analysis
- Small sample sizes - hard to get balanced set of utterances
- Nonetheless, springboards for many years of research

# A Bidding Approach to Turn-taking

Selfridge and Heeman (2009)

# A Bidding Approach to Turn-taking

Selfridge and Heeman (2009)

- Many systems: release-turn approach
  - Speaker controls and releases the turn
- But what about turn conflicts?

# A Bidding Approach to Turn-taking

Selfridge and Heeman (2009)

- Many systems: release-turn approach
  - Speaker controls and releases the turn
- But what about turn conflicts?
- Hypothesis:
  - People continually wish to speak, but limit utterance if it is insufficiently important to the conversation
    - Constant monitoring of utterance importance compared to current speaker's turn cues (turn-releasing or turn-taking)
  - If an utterance is deemed important, the person will interrupt the speaker regardless of release-turn cues
    - extreme example: "Your hair is on fire!"
  - In a turn conflict, whoever "bids" more turn-taking cues will win the turn

# A Bidding Approach to Turn-taking

Selfridge and Heeman (2009)

- Model:
  - Turn-bidding often happens at pauses
  - Speakers use utterance onset to bid for the turn at pauses
  - 5 bids: shorter, short, mid, long, longer
  - Based on importance, as determined through reinforcement learning
- Rationale:
  - Psycholinguistic evidence: Number of turn-conflicts increases under tighter time constraints, as utterances become more urgent

# A Bidding Approach to Turn-taking

Selfridge and Heeman (2009)

- Experiment:
  - Turn-bidding model vs keep-or-release model vs baseline (single utterance model)
  - System-system food ordering dialogue
  - Expert and novice users
  - Three environments: experts only, novices, mixed (unknown)
  - Dialog cost measured by number of actions, based on the belief that efficiency is the primary indicator of user satisfaction
- Results:

| Model | Novice | Expert | Both |
|---|---|---|---|
| Bidding | 9.0 | 4.0 | 6.5 |
| Keep-Or-Release | 9.0 | 4.0 | 7.5 |
| Single-Utterance | 8.7 | 6.0 | 7.4 |

# A Bidding Approach to Turn-taking

Selfridge and Heeman (2009)

- Issues
  - is efficiency really the best indicator of user satisfaction?
  - what about the other turn-taking and turn-releasing cues?
  - is utterance importance relative to the speaker?

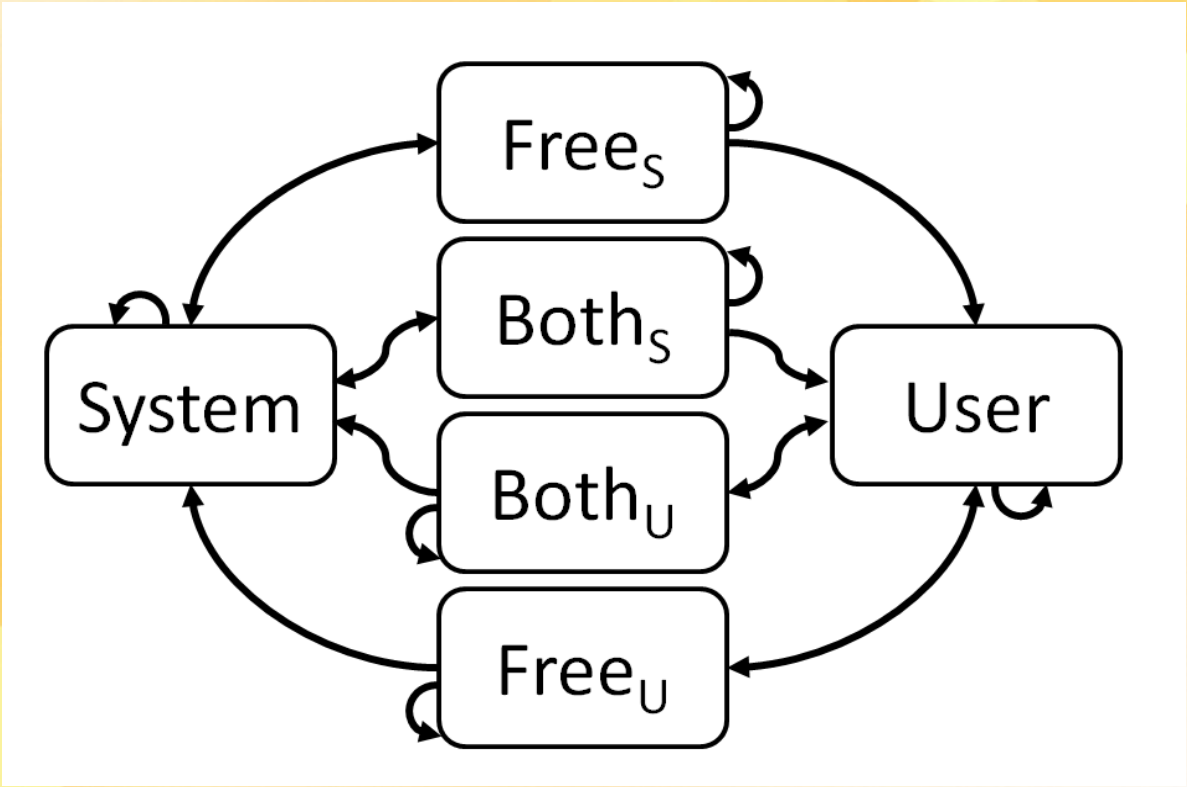
# A Finite-state Turn-taking Model for SDS

Raux and Eskenazi (2009)

- Release-turn system
  - More sophisticated model than the one outperformed by the bidding model
  - Based on predicting TRPs, thus allowing reduction of latency between turn changes
  - Other conversation models: deterministic FSMs with various states of speech and silence
- FSM
  - Proposed: six-state non-deterministic FSM modeling intention/obligation
  - Costs associated with transitions
  - "Decision theoretic action selection": equation to choose best system action given system's belief about current state of model (minimize cost)

# A Finite-state Turn-taking Model for SDS

Raux and Eskenazi (2009)




Finite-state Turn-taking Machine (FSTTM)

# A Finite-state Turn-taking Model for SDS

Raux and Eskenazi (2009)

- Four actions
  - Grab floor
  - Release floor
  - Wait without claiming
  - Keep floor
- Four two-step transitions from one-speaker state to another one-speaker state
  - Turn transitions with gap
  - Turn transitions with overlap
  - Failed interruptions
    - they include backchannels here, though they admit that backchannels do not have the intention of grabbing the floor
  - Time-outs: speaker releases and then grabs the floor

# A Finite-state Turn-taking Model for SDS

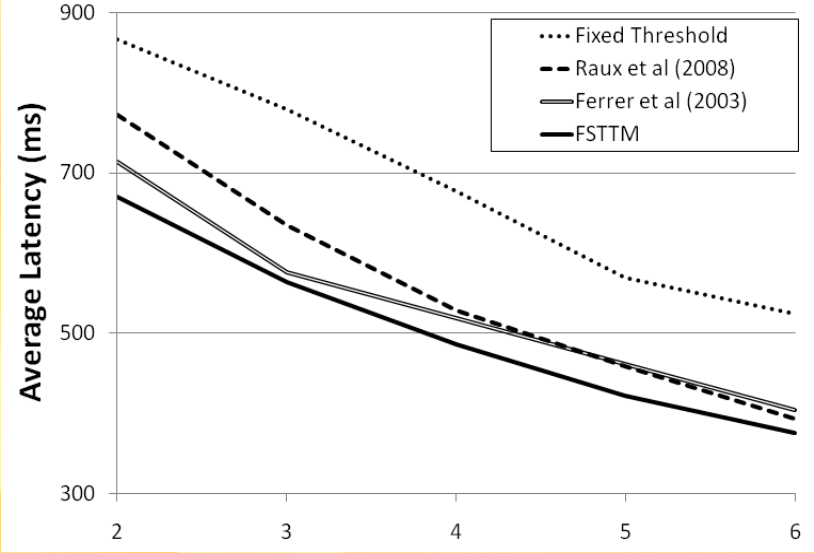
Raux and Eskenazi (2009)

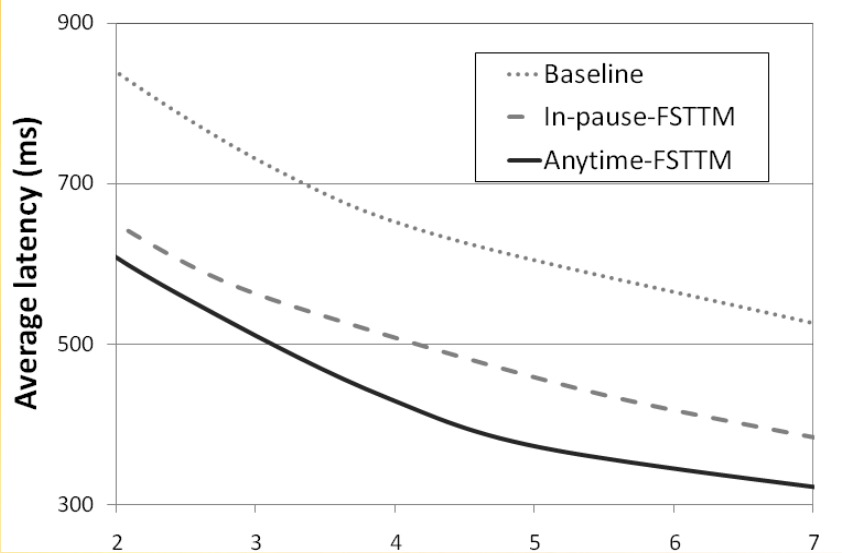
- Examples
  - Turn transitions with gap
    - most common type of transition
    - SYSTEM --(R,W)--> FREE_s --(W,G)--> USER
  - Turn transitions with overlap
    - barge-in
    - SYSTEM --(K,G)--> BOTH_s --(R,K)--> USER
- Why non-deterministic?
  - System doesn't know intention of the user; thus, it cannot know for certain which state it is in.
- Goal: Endpointing
  - Determine whether a pause is turn-final or turn-internal
  - System grabs floor when cost of waiting exceeds cost of grabbing

# A Finite-state Turn-taking Model for SDS

Raux and Eskenazi (2009)

## Results



# Outline

- Turn-taking
  - Background
  - Early Research
  - Recent Efforts
    - A Bidding Approach to Turn-taking
    - A Finite-state Turn-taking Model
- **Backchannels**
  - Background
  - A Shallow Model of Backchannel Continuers in Spoken Dialogue

# Backchannels

- **Backchannel**: signal that communication is working
  - **Continuers**: short utterances indicating that the speaker should continue with his/her turn
    - e.g. "right", "okay", "mm-hmm"
  - Backchannels can also be longer utterances, repeating parts of a speaker's utterance

# Backchannels

- **Backchannel**: signal that communication is working
  - **Continuers**: short utterances indicating that the speaker should continue with his/her turn
    - e.g. "right", "okay", "mm-hmm"
  - Backchannels can also be longer utterances, repeating parts of a speaker's utterance
- Why do we care about backchannels?

# Backchannels

- **Backchannel**: signal that communication is working
  - **Continuers**: short utterances indicating that the speaker should continue with his/her turn
    - e.g. "right", "okay", "mm-hmm"
  - Backchannels can also be longer utterances, repeating parts of a speaker's utterance
- Why do we care about backchannels?

# Backchannels

- Whether through gesture or utterance, we constantly seek feedback and confirmation from our audience
  - Lack of backchannels often cause speaker to elicit explicit acknowledgements (e.g. "Does that make sense?")
- Do we need them for SDS?
  - May not be as necessary for information-seeking systems, with short prompts and commands
  - Important for other tasks, where user must give longer, more complex input (e.g. tutoring system)
  - Done wrong, can be unnatural and disruptive
- What does it mean when a system is silent?
  - System is listening (user should speak)
  - System is processing (user should not speak)

# A Shallow Model of Backchannel Continuers in Spoken Dialogue

Cathcart et al (2003)

- Goal: Low-cost method of adding continuers to SDS
- Hypothesis:
  - Backchannel continuers (bcs) occur at TRPs
    - TRP identified by a grammatical completion (the syntactic approach of Sacks et al)
    - cTRP identified by grammatical completion, intention and intonation
- HCRC Map Task Corpus
  - bcs occur as subset of *acknowledge* moves in annotated dialog
  - filtered by content words, conveyed acceptance
- Three models:
  - Pause-duration model
  - N-gram POS model
  - Combination model

# A Shallow Model of Backchannel Continuers in Spoken Dialogue

Cathcart et al (2003)

- Baseline model
  - Insert bc after every n words
  - Rationale: expect bcs at intonational phrase boundaries (TRP indicator)
  - Low-cost - no pitch tracker. In spoken English, phrase boundaries known to occur every 5-15 syllables

# A Shallow Model of Backchannel Continuers in Spoken Dialogue

Cathcart et al (2003)

- Baseline model
  - Insert bc after every n words
  - Rationale: expect bcs at intonational phrase boundaries (TRP indicator)
  - Low-cost - no pitch tracker. In spoken English, phrase boundaries known to occur every 5-15 syllables
- Pause-duration model
  - Rationale: continuers often occur at TRPs, and TRPs often contain pauses
    - 50% of pauses w/o continuers are < 500ms, and only 11% of these pauses have continuers
  - Automatically produce a continuer when pause reaches a certain threshold

# A Shallow Model of Backchannel Continuers in Spoken Dialogue

Cathcart et al (2003)

- N-gram POS model
  - Find POS trigrams most likely to contain bc
  - Nouns before pauses are good indicators (nine of top ten contain pause)
  - Continuer inserted after likely trigrams
    - issue: probability for top trigram is 0.26, meaning 3/4 of the insertions would be erroneous

# A Shallow Model of Backchannel Continuers in Spoken Dialogue

Cathcart et al (2003)

- N-gram POS model
  - Find POS trigrams most likely to contain bc
  - Nouns before pauses are good indicators (nine of top ten contain pause)
  - Continuer inserted after likely trigrams
    - issue: probability for top trigram is 0.26, meaning 3/4 of the insertions would be erroneous
- Combined model
  - Most of the sequences predicted by the LM contain pauses
  - Pauses also indicate end of move
  - Solution: use pause-threshold to eliminate some end-of-move pauses

# A Shallow Model of Backchannel Continuers in Spoken Dialogue

Cathcart et al (2003)

- Evaluation
  - Sticking with low-cost, compared model to annotated corpus (previously unseen)
  - Bcs are optional, so human speakers may choose to forgo a bc opportunity

- Results

| Model | Precision | Recall | F-measure |
|---|---|---|---|
| Baseline (7 words) | 4 | 13 | 7 |
| Pause-Duration (.9s) | 22 | 58 | 32 |
| *n*-gram POS | 22 | 50 | 30 |
| Combined (3 tri, .6s) | 29 | 43 | **35** |
| Combined (10 tri, .9s) | 25 | 51 | 33 |

# Outline

- Turn-taking
  - Background
  - Early Research
  - Recent Efforts
    - A Bidding Approach to Turn-taking
    - A Finite-state Turn-taking Model
- **Backchannels**
  - Background
  - A Shallow Model of Backchannel Continuers in Spoken Dialogue

# References


Ethan O. Selfridge and Peter A. Heeman. 2009. A Bidding Approach to Turn-Taking. In *1st International Workshop on Spoken Dialogue Systems*.

A. Raux and M. Eskenazi. 2009. A Finite-State Turn-Taking Model for Spoken Dialog Systems. In *Proceedings of HLT-NAACL* 2009.

Nicola Cathcart; Jean Carletta; Ewan Klein. 2003. A shallow model of back-channel continuers in spoken dialogue. In *Proceedings of EACL-2003*.

Starkey Duncan. 1972. Some signals and rules for taking speaking turns. In *Conversations Journal of Personality and Social Psychology*, Vol 23(2): 283-292.

Harvey Sacks, Emanuel A. Schegloff and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50(4):696-735.

Questions?