

EXPLORING THE PROSODY OF FLOOR MECHANISMS IN ENGLISH USING THE FFV SPECTRUM

Kornel Laskowski^{1,2}, Mattias Heldner³ and Jens Edlund³

¹ Language Technologies Institute, Carnegie Mellon University, USA

² Institut für Anthropomatik, Universität Karlsruhe, Germany

³ Department of Speech, Music and Hearing, KTH, Sweden

Goal

Determine if fundamental frequency variation (FFV) discriminates between floor mechanisms and other dialog act (DA) types.

floor holder (f_h) hold floor within turn

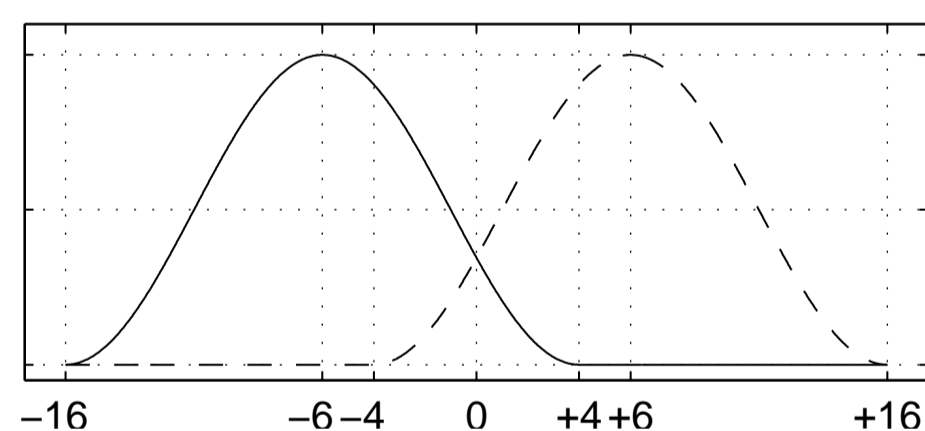
hold (h) secure offered floor at start of turn

floor grabber (f_g) secure unoffered floor at start of turn

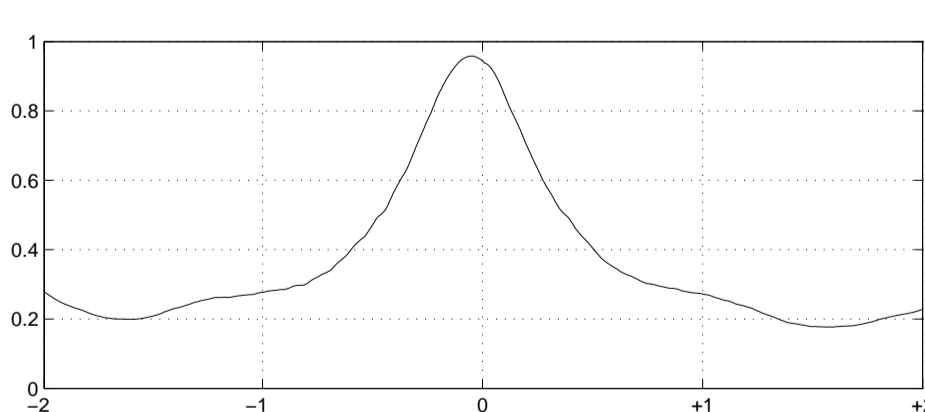
NB: Floor mechanisms lexically similar to other short DA types.

What Is the Fundamental Frequency Variation Spectrum?

- dot product of two spectra, F_L and F_R ...



- ... after dilating one of F_L and F_R by factor 2^ρ
- FFV spectrum is value of dot product over range of ρ



- 7-filter filterbank:
 - flat pitch
 - slowly falling (rising) pitch
 - quickly falling (rising) pitch
 - 2 normalization values

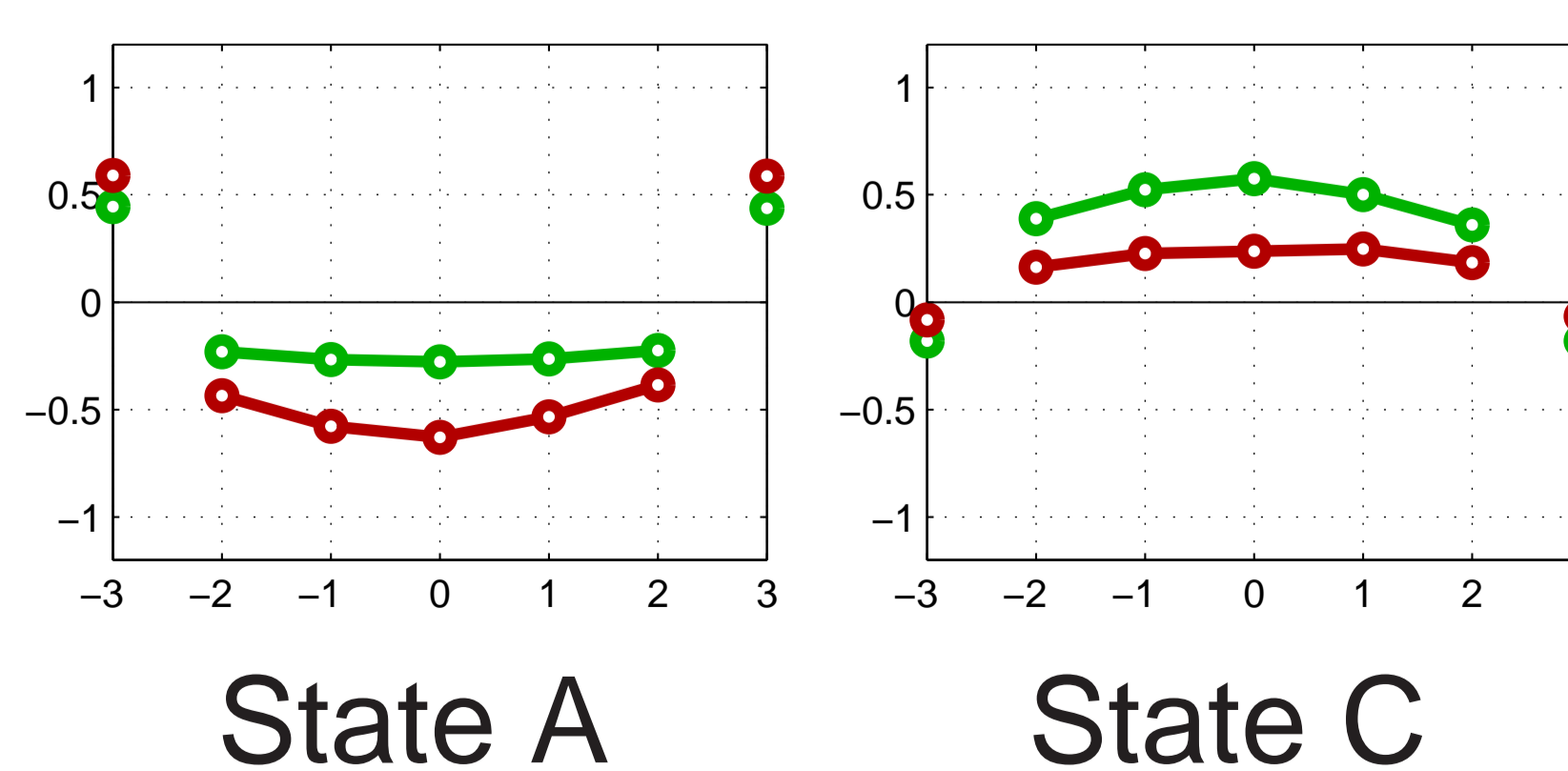
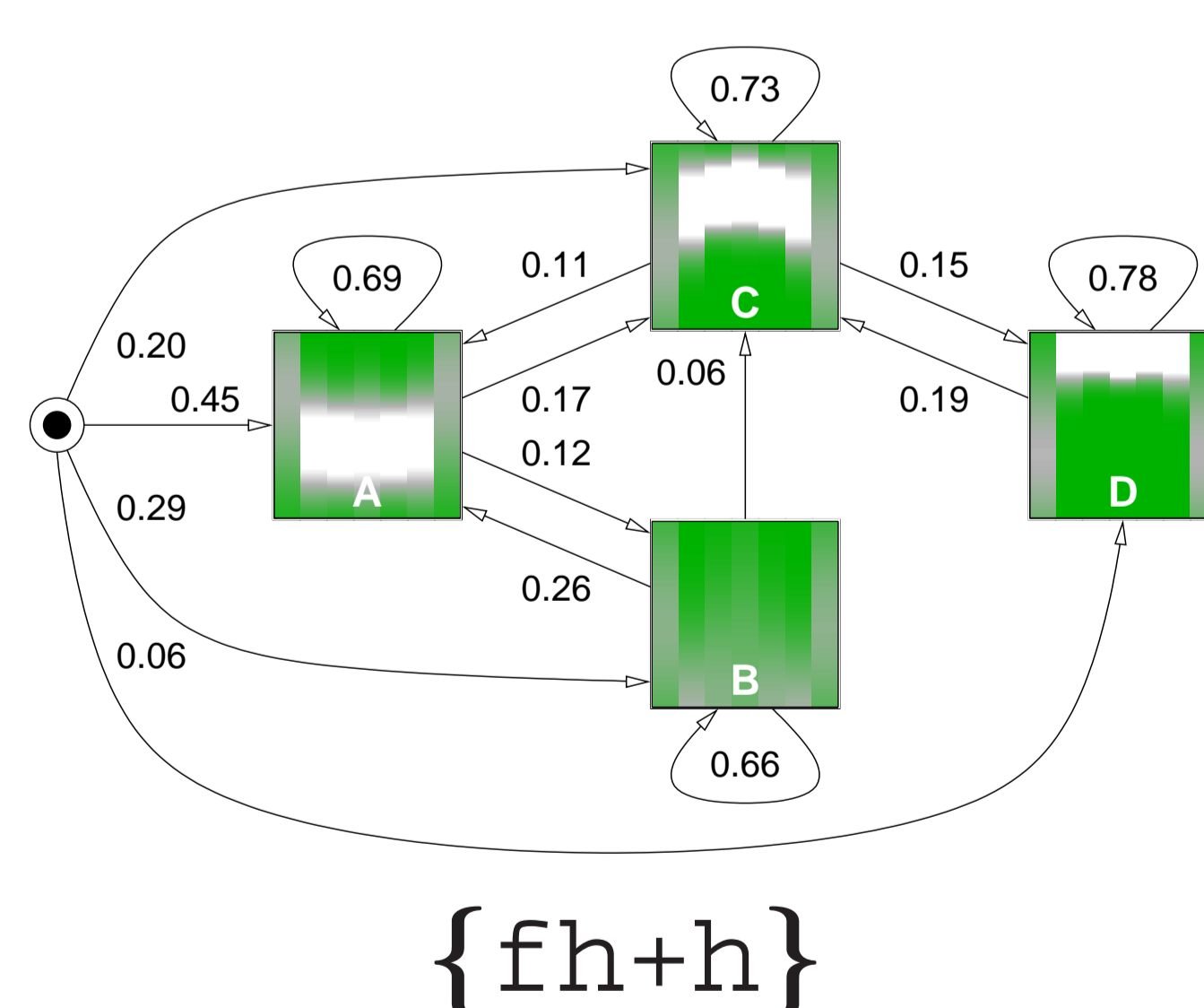
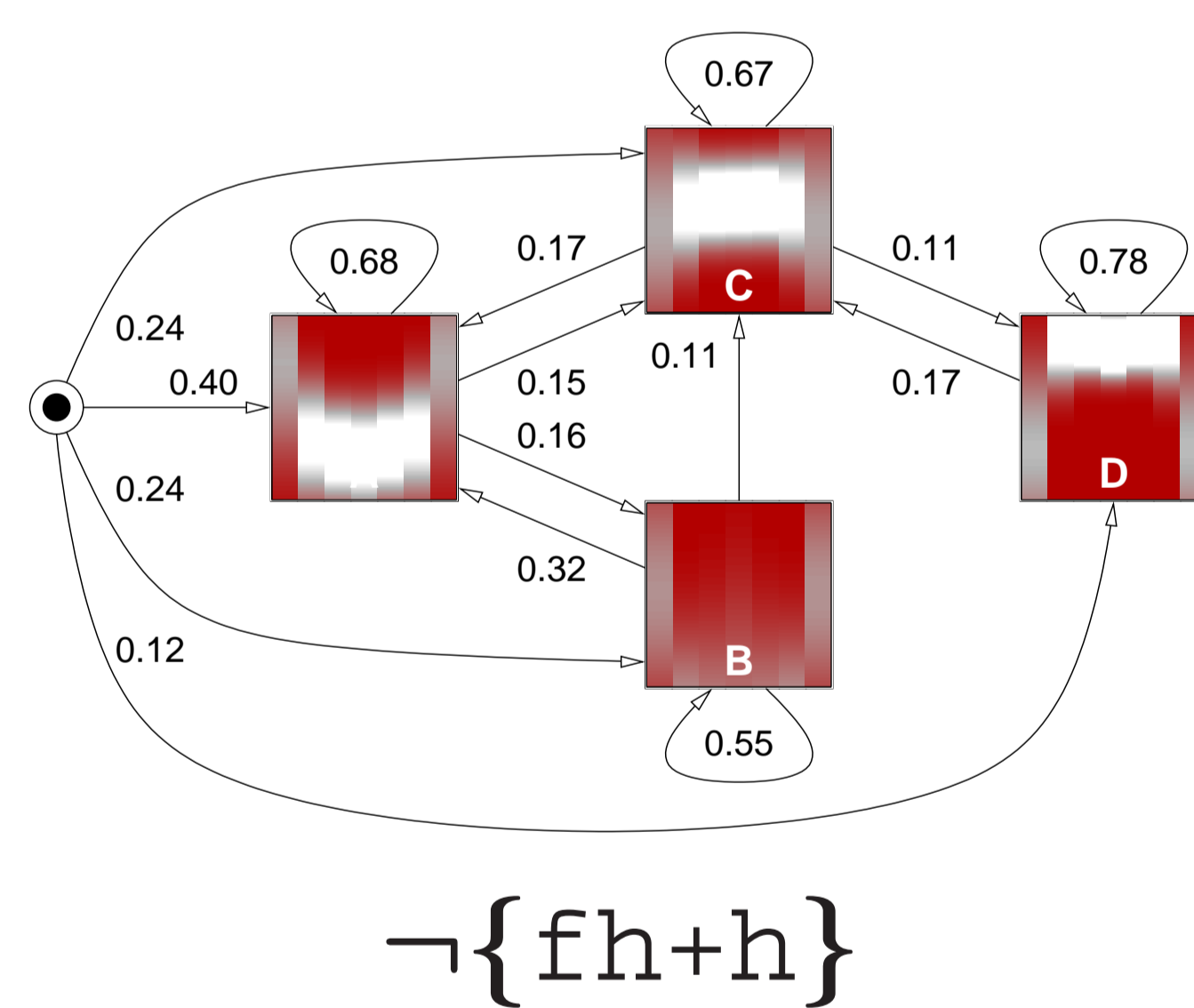
Experiments

- ICSI Meeting Corpus
 - naturally occurring, spontaneous multiparty speech
 - close-talk-mic 16kHz audio
 - training: 2×7000 talkspurts
 - testing: 2×1000 talkspurts
- compute HMM density models
- classify: $\neg\{f_h + h\}$ or $\{f_h + h\}$
 - first 500ms of every talkspurt
 - last 500ms of every talkspurt
- test accuracy: **64.1–65.3%**
- test ROC area: **70.6–72.2%**

Findings

- FFV discriminates between floor mechanisms $\{f_h + h\}$ and other DA types $\neg\{f_h + h\}$
- consecutive frames more similar for $\{f_h + h\}$ than for $\neg\{f_h + h\}$
 - first 500ms: effect more pronounced during flat pitch
 - last 500ms: effect independent of pitch trajectory
- FFV representation useful in:
 - multiparty dialogue scenarios
 - non-anechoic-chamber recording scenarios

Inferred Models of Talkspurt Initiation



Inferred Models of Talkspurt Termination

