ACTIVITIES on JAPANESE PROSODIC CORPUS

Shigeyoshi Kitazawa, Shizuoka University, Hamamatsu, Japan (kitazawa@cs.inf.shizuoka.ac.jp)

MULTEXT: Parallel Corpus

	Japanese-MULTEXT Prosodic Corpus	MULTEXT Prosodic database
Language	Tokyo dialect	French, English, Italian, German, Spanish
Speech Duration	3 hours and 37 minutes	4 hours and 20 minutes
Speakers	3 male and 3 female	50 (5 male and 5 female per lang.)
Recording Unit	Translated EUROM1 Passage	EUROM1 Passage
Spline Stylised	Yes	Yes
Symbolic Coding	108,500 J-ToBI	50,000 pitch movements
Orthographic transcription	Given	Given
Phoneme transcription	Yes	No
Manual correction	Yes	Yes
Re-synthesis	No	Yes
EGG waveform	Yes	No
EGG analysis	Will be available soon	No

Japanese Prosodic Corpuses

	Priority Area Project	ASJ Guide-task	Multimodal Dialogue	Map Task	Weather Forecast
Speaking style	Played dialogue	Read speech	Debate task	Natural dialogue	Read speech
Speakers	83	18 male, 18 female	12 male, 12 female	16 males	2 male, 2 female
Duration	6400 sentences	10 hours	360 minutes	68 minutes	1027 sentences
Annotation	Sentence, pause, J-ToBI*	Sentence, pause, J-ToBI*	Words, gaze, nod, J-ToBI*	J-ToBI	Sentence J-ToBI*
Dialogues	64	202	72	8	
F0	Min. Max.	Min. Max.	No	Yes	Yes
Phoneme	No	No	No	Yes	Yes
Contact	S. Itahashi	S.Itahashi	Y. Horiuchi	Y. Horiuchi	K. Hirose

J-ToBI*: planned

Prosodic Unit

Accent Kernels

Phoneme Labels

Extracted F0 over the Narrow Band Spectrpgram

J-ToBI

Stylized F0 over the Narrow Band Spectrpgram
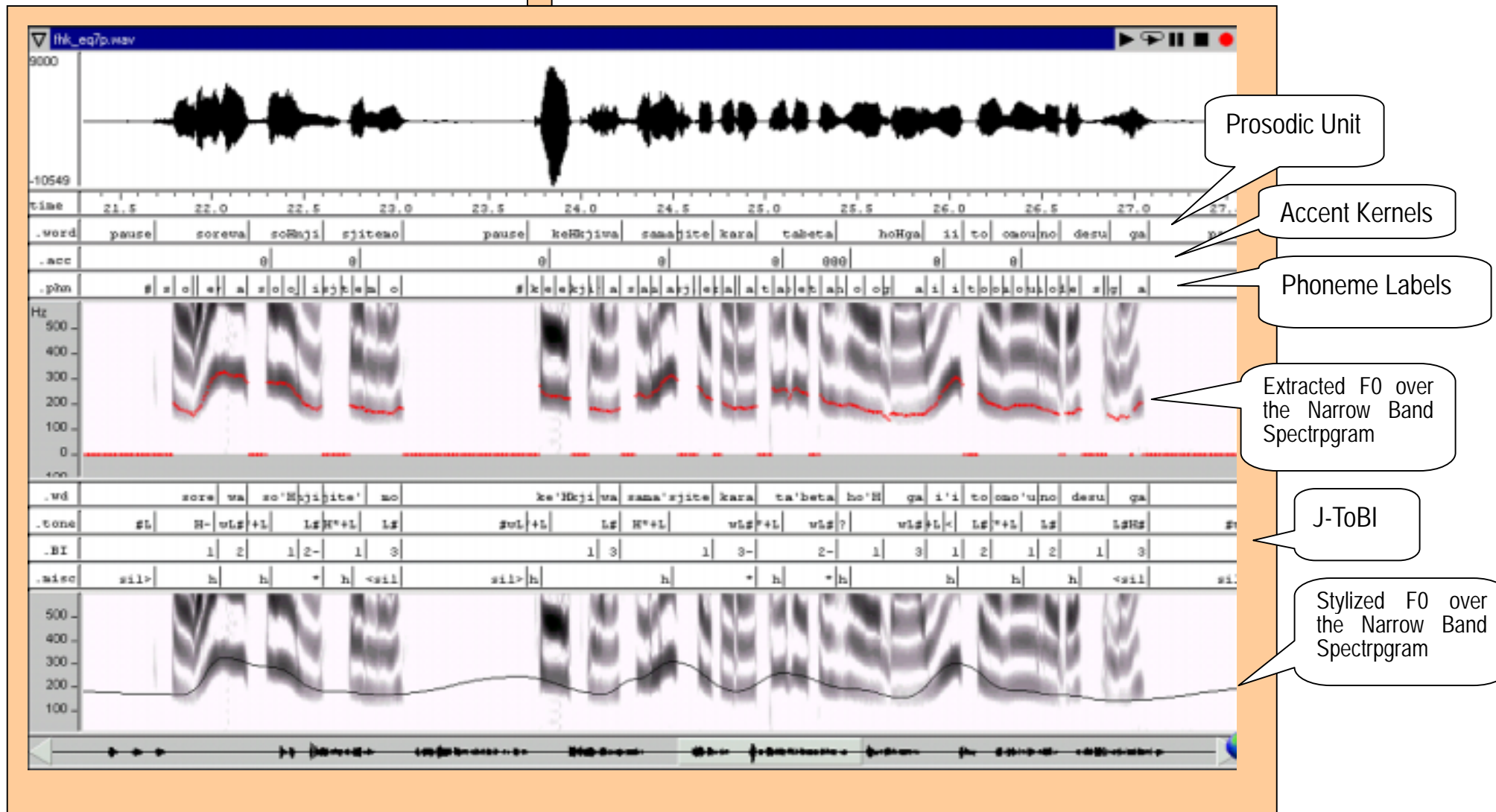
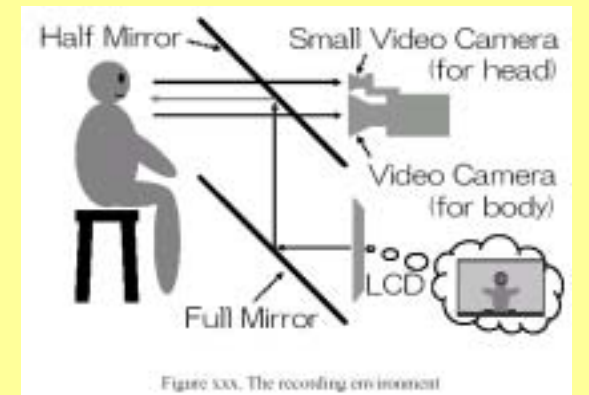### Annotation of Gesture in Speech Dialogue
(Yasuo Horiuchi, Chiba University)

Figure xxx. The recording environment