CS 224S / LINGUIST 285
Spoken Language Processing

Andrew Maas
Stanford University
Spring 2017

Lecture 11: Emotion/Affect Extraction and Interpersonal Stance

Original slides by Dan Jurafsky

Outline

1. Theoretical background on emotion and smiles
2. Extracting emotion from speech and text: case studies and features
3. Interpersonal stance and speed dating case study

Social Signal Processing = Affect/Emotion Detection

- Detecting frustration of callers to a help line
- Detecting stress in drivers or pilots
- Detecting depression, intoxication
- Detecting interest, certainty, confusion in on-line tutors
  - Pacing/Positive feedback
- Hot spots in meeting summarizers/browsers
- Synthesis/generation:
  - On-line literacy tutors in the children's storybook domain
  - Computer games

Scherer's typology of affective states

Emotion: relatively brief episode of synchronized response of all or most organismic subsystems in response to the evaluation of an external or internal event as being of major significance

angry, sad, joyful, fearful, ashamed, proud, desperate

Mood: diffuse affect state …change in subjective feeling, of low intensity but relatively long duration, often without apparent cause

cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stance: affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange

distant, cold, warm, supportive, contemptuous

Attitudes: relatively enduring, affectively colored beliefs, preferences predispositions towards objects or persons

liking, loving, hating, valuing, desiring

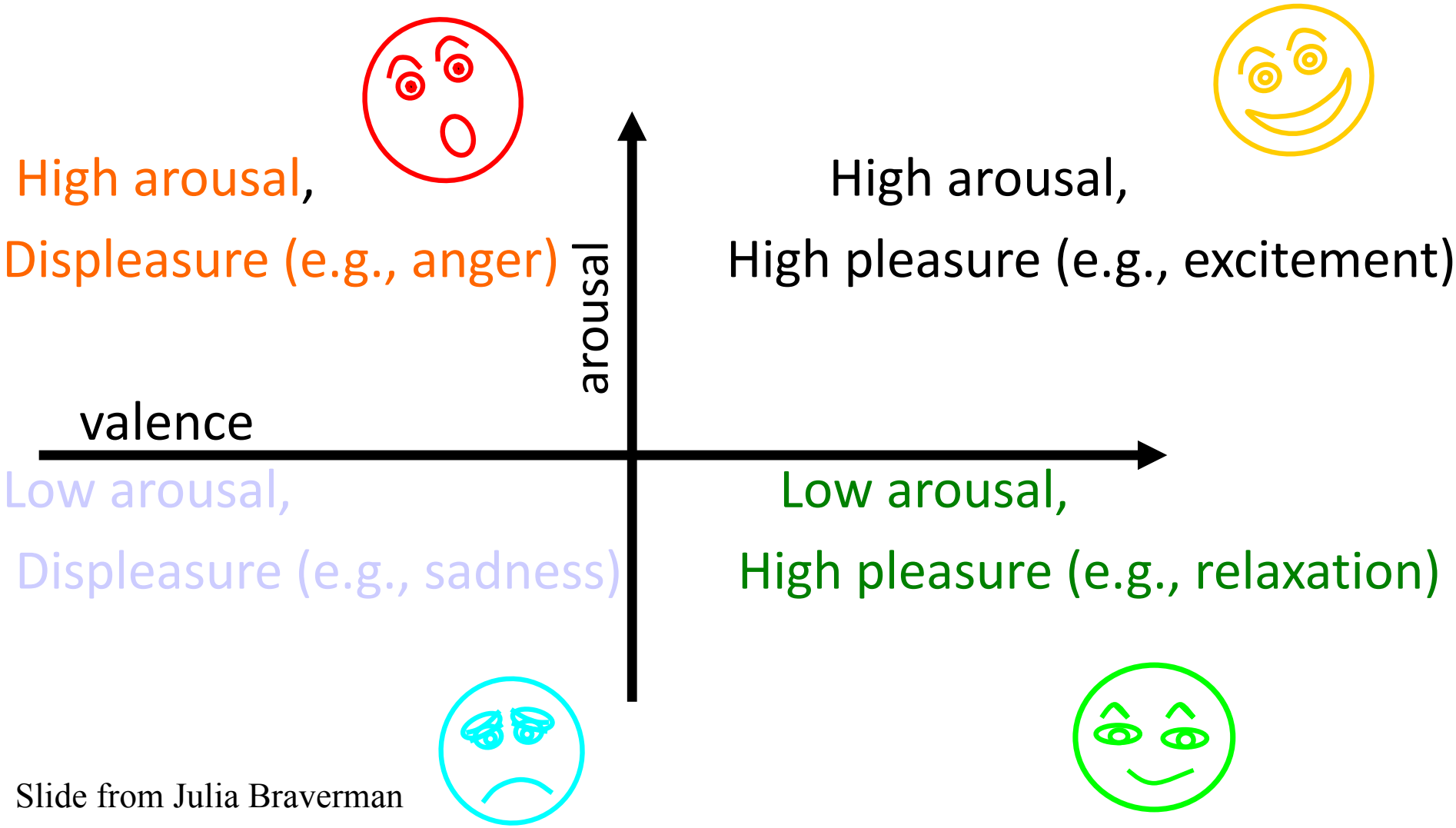
Personality traits: emotionally laden, stable personality dispositions and behavior tendencies, typical for a person

nervous, anxious, reckless, morose, hostile, envious, jealous

Ekman’s 6 basic emotions

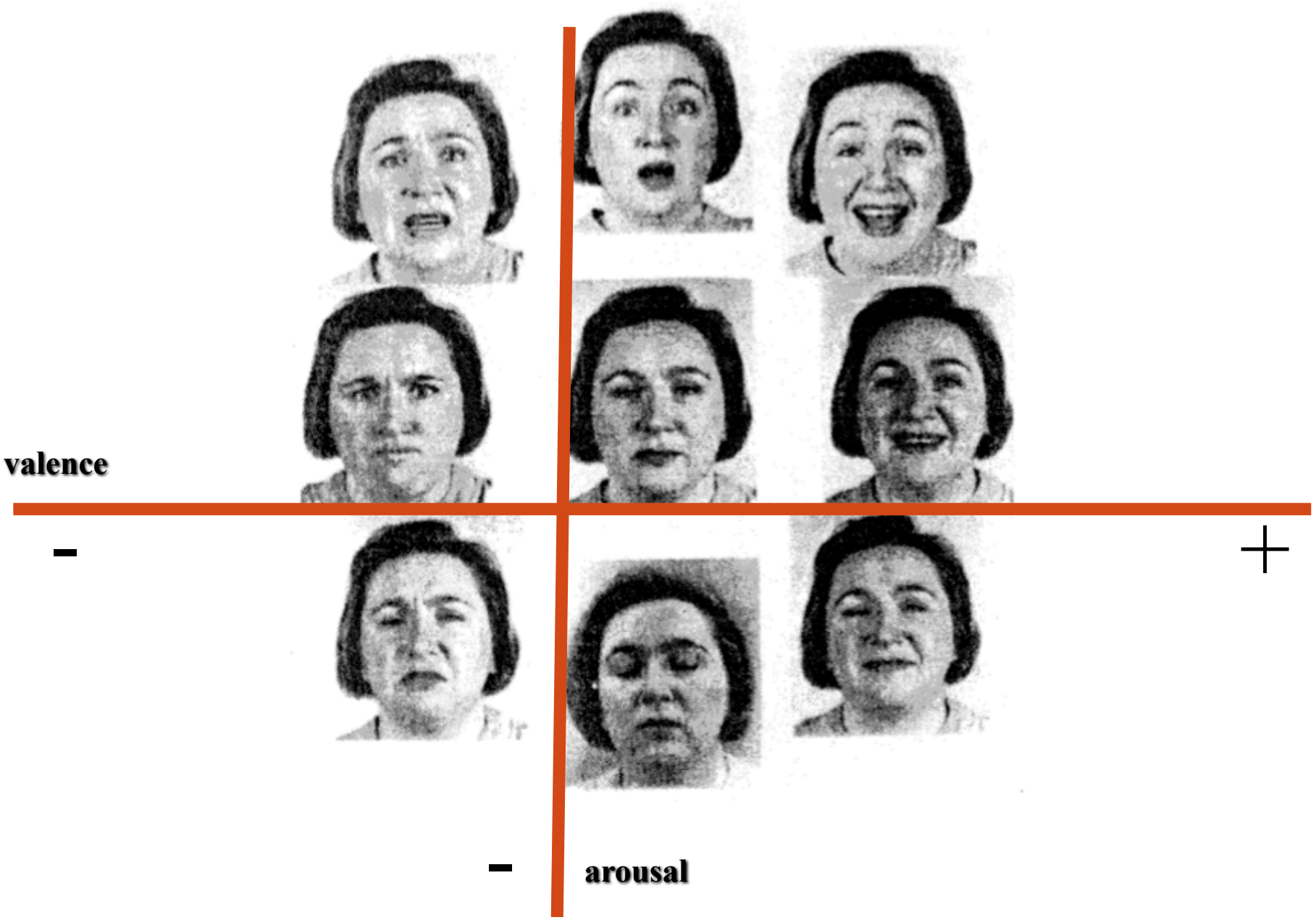
Surprise, happiness, anger, fear, disgust, sadness

Dimensional approach. (Russell, 1980, 2003)



Slide from Julia Braverman

Image from Russell 1997



Distinctive vs. Dimensional approach

Distinctive

- Emotions are units.
- Limited number of basic emotions.
- Basic emotions are innate and universal
- Methodology advantage
  - Useful in analyzing traits of personality.

Dimensional

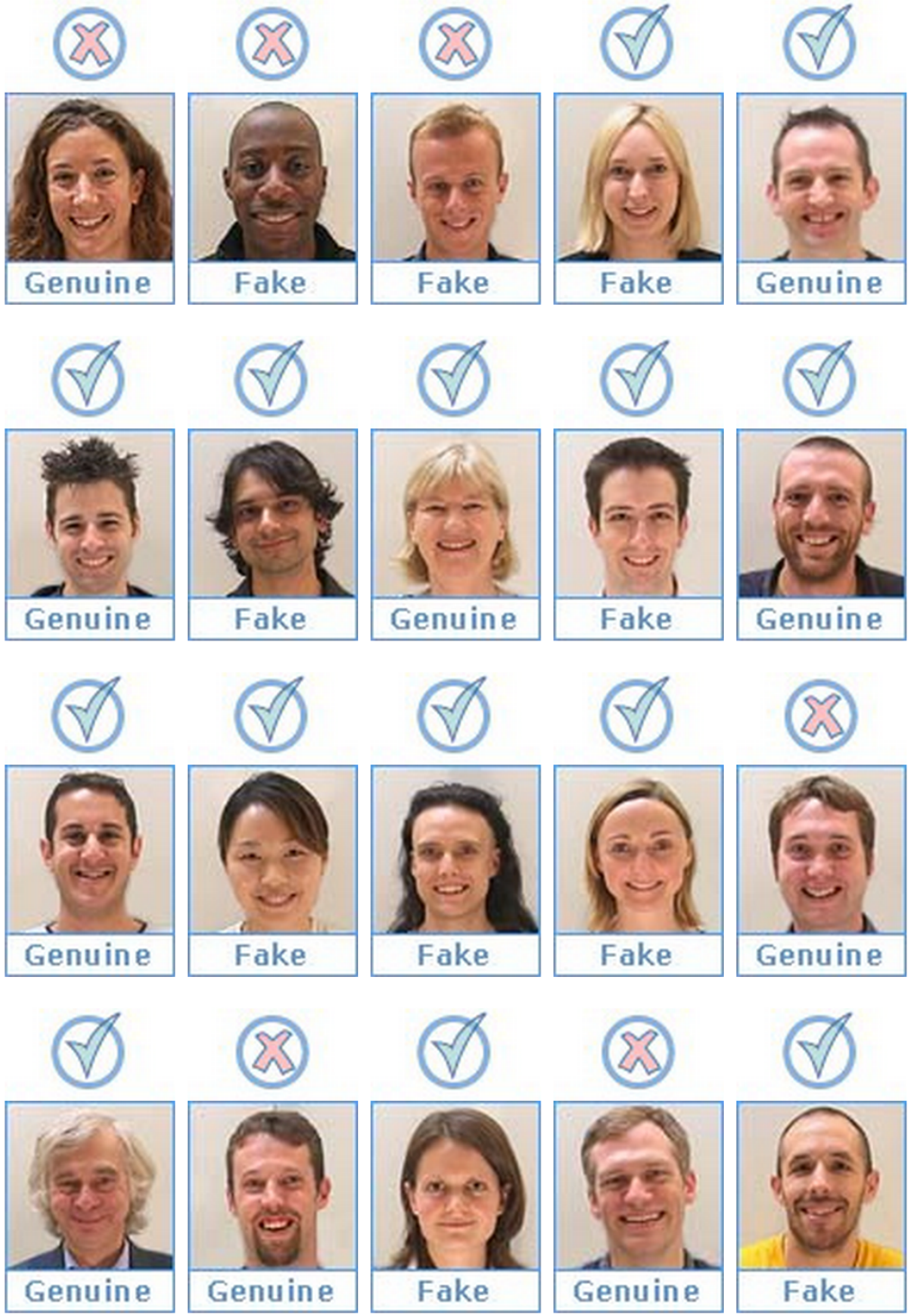
- Emotions are dimensions.
- Limited # of labels but unlimited number of emotions.
- Emotions are culturally learned.
- Methodological advantage:
  - Easier to obtain reliable measures.

Slide from Julia Braverman

Duchenne versus non-Duchenne smiles

- http://www.bbc.co.uk/science/humanbody/mind/surveys/smiles/
- http://www.cs.cmu.edu/afs/cs/project/face/www/facs.htm

Duchenne smiles

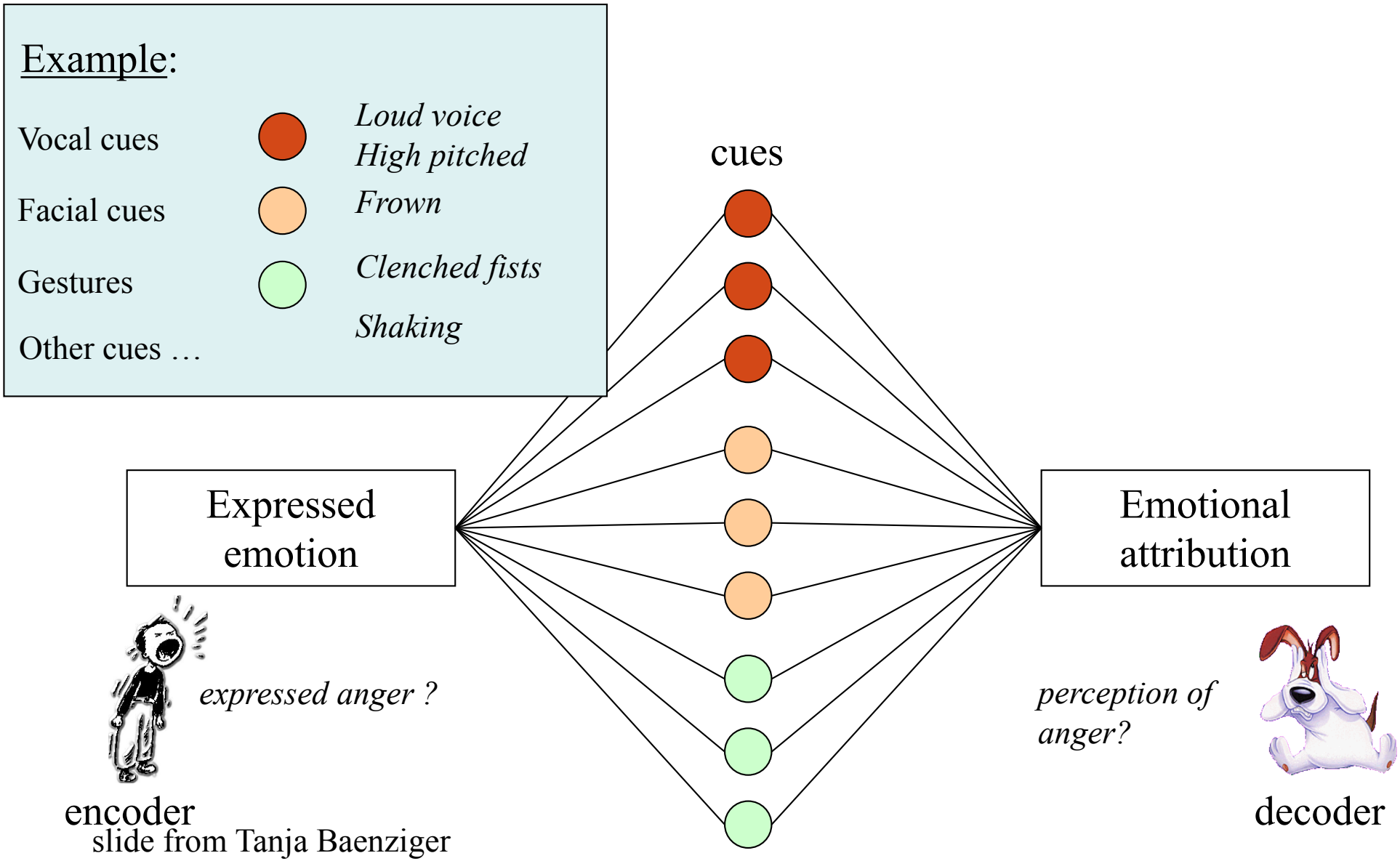


How to detect Duchenne smiles

- “As well as making the mouth muscles move, the muscles that raise the cheeks – the orbicularis oculi and the pars orbitalis – also contract, making the eyes crease up, and the eyebrows dip slightly.
- Lines around the eyes do sometimes appear in intense fake smiles, and the cheeks may bunch up, making it look as if the eyes are contracting and the smile is genuine.
- But there are a few key signs that distinguish these smiles from real ones. For example, when a smile is genuine, the eye cover fold - the fleshy part of the eye between the eyebrow and the eyelid - moves downwards and the end of the eyebrows dip slightly.”

BBC Science webpage referenced on previous slide

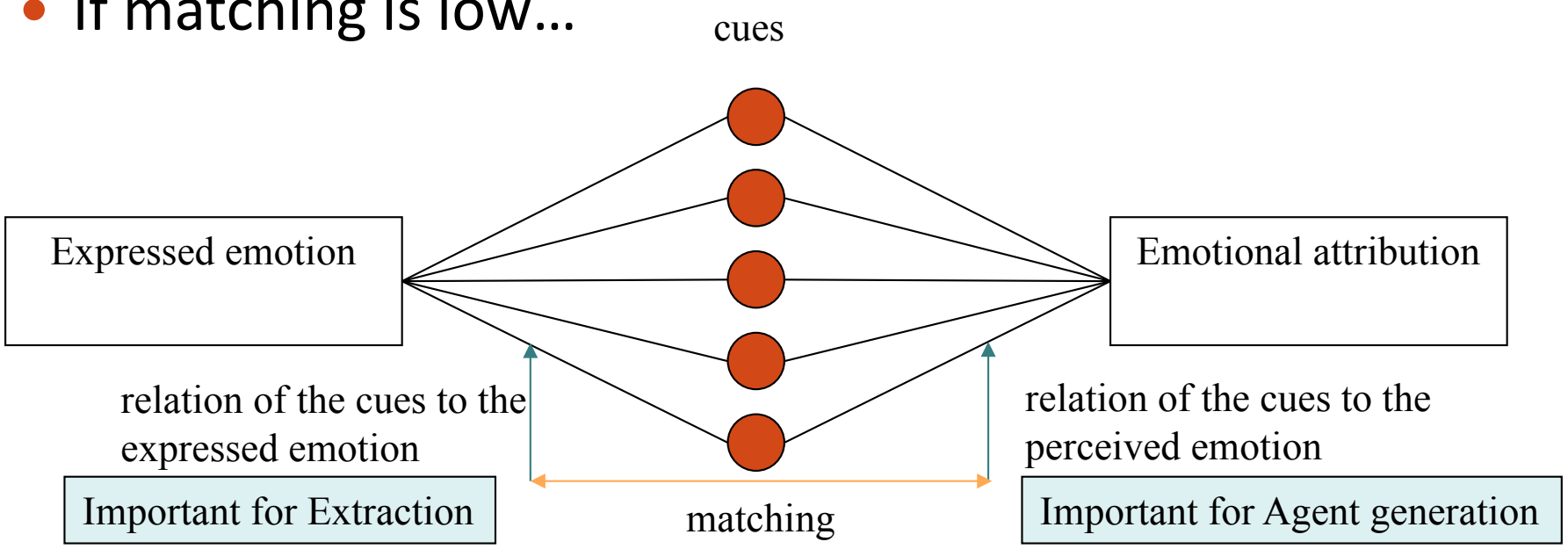
Emotional communication and the Brunswikian Lens



slide from Tanja Baenziger

Implications for HCI

- If matching is low…



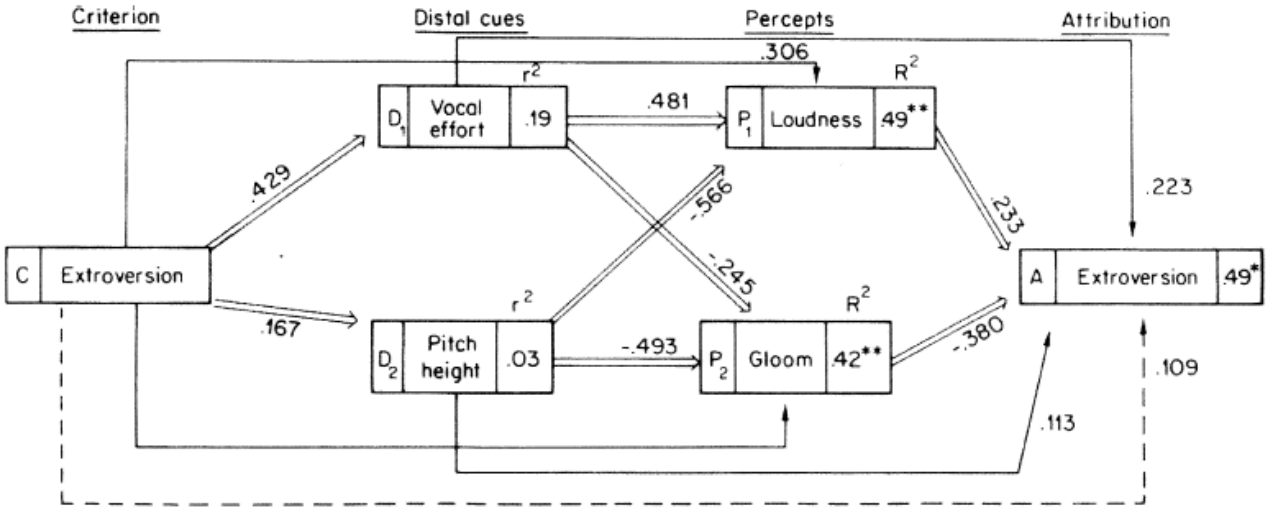
- Generation (Conversational agents): relation of cues to **perceived** emotion
- Recognition (Extraction systems): relation of the cues to **expressed** emotion

slide from Tanja Baenziger

Extroversion in Brunswikian Lens

- Simulated jury discussions in German and English
  - speakers had detailed personality tests
- Extroversion personality type accurately identified from naïve listeners by voice alone
- But not emotional stability
  - listeners choose: resonant, warm, low-pitched voices
  - but these don't correlate with actual emotional stability

I



Acoustic implications of Duchenne smile

Amy Drahota, Alan Costall, Vasudevi Reddy. 2008.
The vocal communication of different kinds of smile. Speech Communication

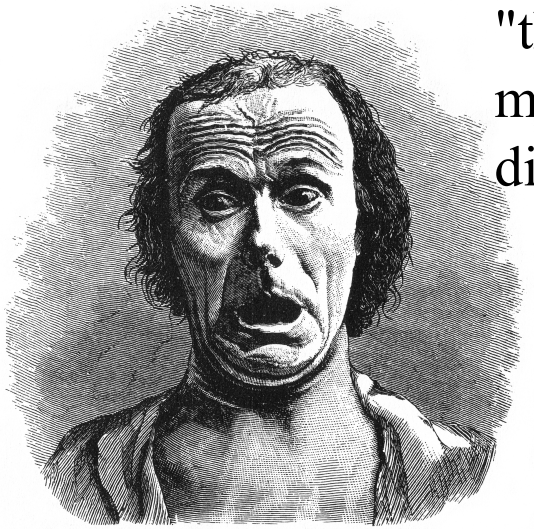
- "Asked subjects to repeat the same sentence in response to a set sequence of 17 questions, intended to provoke reactions such as amusement, mild embarrassment, or just a neutral response."
- Coded and examined Duchenne, non-Duchenne, and "suppressed" smiles".
- Listeners could tell the differences, but many mistakes
- Standard prosodic and spectral (formant) measures showed no acoustic differences of any kind.
- Correlations between listener judgments and acoustics:
  - larger differences between f2 and f3-> not smiling
  - smaller differences between f1 and f2 -> smiling

Evolution and Duchenne smiles

- “honest signals” (Pentland 2008)
- “behaviors that are sufficiently expensive to fake that they can form the basis for a reliable channel of communication”

Four Theoretical Approaches to Emotion: 1. Darwinian (natural selection)

- Darwin (1872) The Expression of Emotion in Man and Animals. Ekman, Izard, Plutchik
  - Function: Emotions evolve to help humans survive
  - Same in everyone and similar in related species
    - Similar display for Big 6+ (happiness, sadness, fear, disgust, anger, surprise) → 'basic' emotions
    - Similar understanding of emotion across cultures



The particulars of fear may differ, but "the brain systems involved in mediating the function are the same in different species" (LeDoux, 1996)

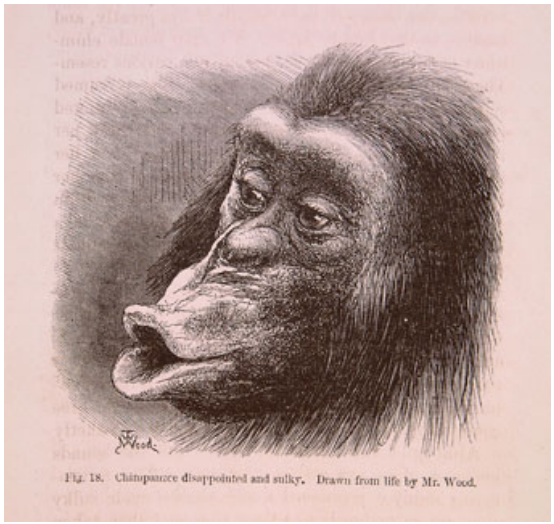


Fig. 18. Chimpanzee disappointed and sulky. Drawn from life by Mr. Wood.

extended from Julia Hirschberg's slides discussing Cornelius 2000

Four Theoretical Approaches to Emotion: 2. Jamesian: Emotion is experience

- William James 1884. What is an emotion?
  - Perception of bodily changes → emotion
    - “we feel sorry because we cry… afraid because we tremble"’
    - “our feeling of the … changes as they occur IS the emotion"
  - The body makes automatic responses to environment that help us survive
  - Our experience of these responses consitutes emotion.
  - Thus each emotion accompanied by unique pattern of bodily responses
    - Stepper and Strack 1993: emotions follow facial expressions or posture.
    - Botox studies:
      - Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic use of botulinum toxin-A affects processing of emotional language Psychological Science, 21, 895-900.
      - Hennenlotter, A., Dresel, C., Castrop, F., Ceballos Baumann, A. O., Wohlschlager, A. M., Haslinger, B. (2008). The link between facial feedback and neural activity within central circuitries of emotion - New insights from botulinum toxin-induced denervation of frown muscle Cerebral Cortex, June 17.

extended from Julia Hirschberg’s slides discussing Cornelius 2000

Four Theoretical Approaches to Emotion: 3. Cognitive: Appraisal

- An emotion is produced by appraising (extracting) particular elements of the situation. (Scherer)
  - **Fear**: produced by the appraisal of an event or situation as obstructive to one's central needs and goals, requiring urgent action, being difficult to control through human agency, and lack of sufficient power or coping potential to deal with the situation.
  - **Anger**: difference: entails much higher evaluation of controllability and available coping potential
- Smith and Ellsworth's (1985):
  - **Guilt**: appraising a situation as unpleasant, as being one's own responsibility, but as requiring little effort.

Adapted from Cornelius 2000

Four Theoretical Approaches to Emotion: 4. Social Constructivism

- Emotions are cultural products (Averill)
- Explains gender and social group differences
- **anger** is elicited by the appraisal that one has been wronged intentionally and unjustifiably by another person. Based on a moral judgment
  - don't get angry if you yank my arm accidentally
  - or if you are a doctor and do it to reset a bone
  - only if you do it on purpose

Adapted from Cornelius 2000

Link between valence/arousal and Cognitive-Appraisal model

- Dutton and Aron (1974)
- Male participants cross a bridge
  - sturdy
  - precarious
- Other side of bridge: female asks participants to take part in a survey
  - willing participants were given interviewer's phone number
- Participants who crossed precarious bridge
  - more likely to call and use sexual imagery in survey
- Participants misattributed their arousal as sexual attraction

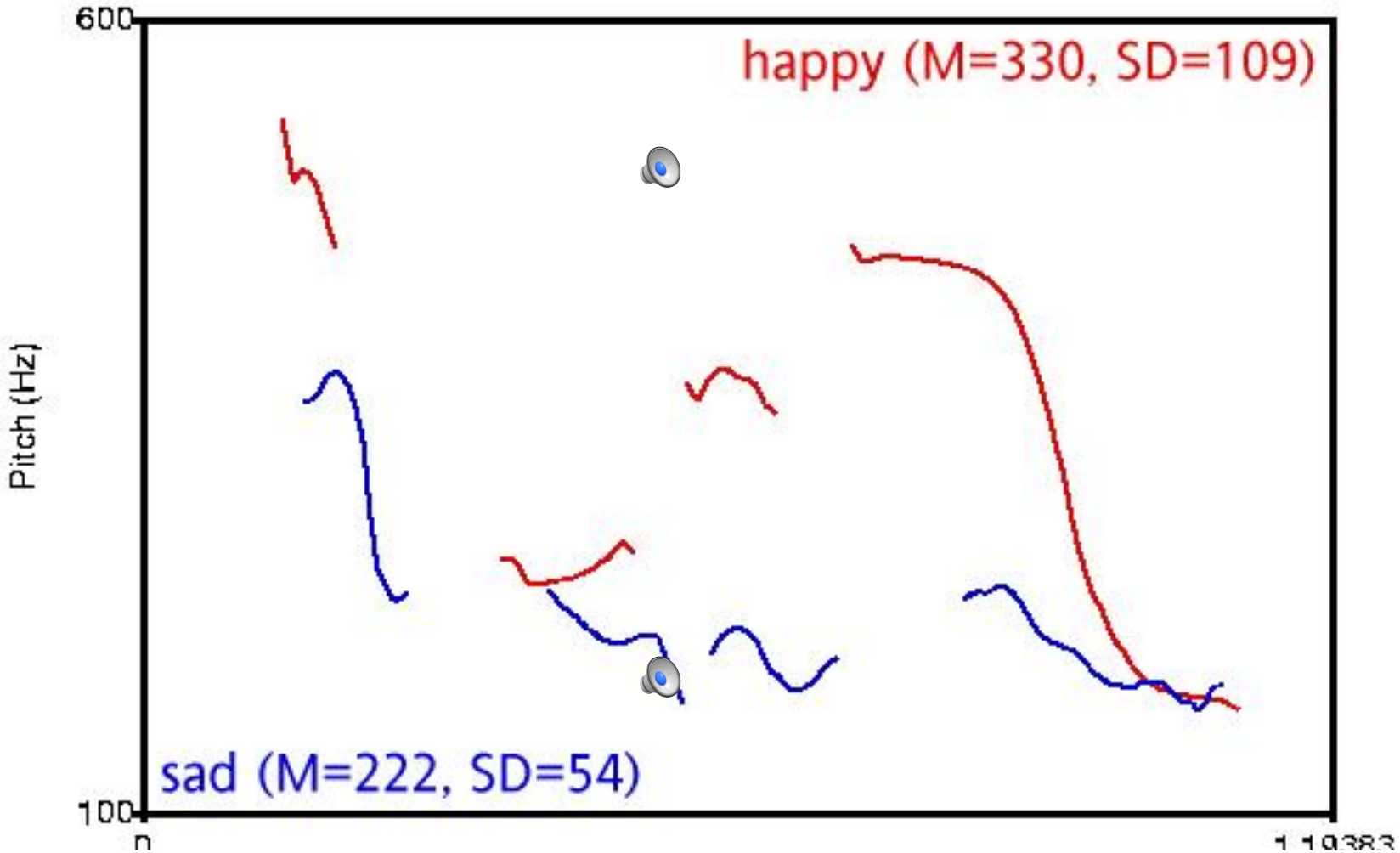
Part II: Case studies and features

Hard Questions in Emotion Recognition

- How do we know what emotional speech is?
  - Acted speech vs. natural (hand labeled) corpora
- What can we classify?
    - Distinguish among multiple 'classic' emotions
    - Distinguish
      - Valence: is it positive or negative?
      - Activation: how strongly is it felt? (sad/despair)
- What features best predict emotions?
- What techniques best to use in classification?

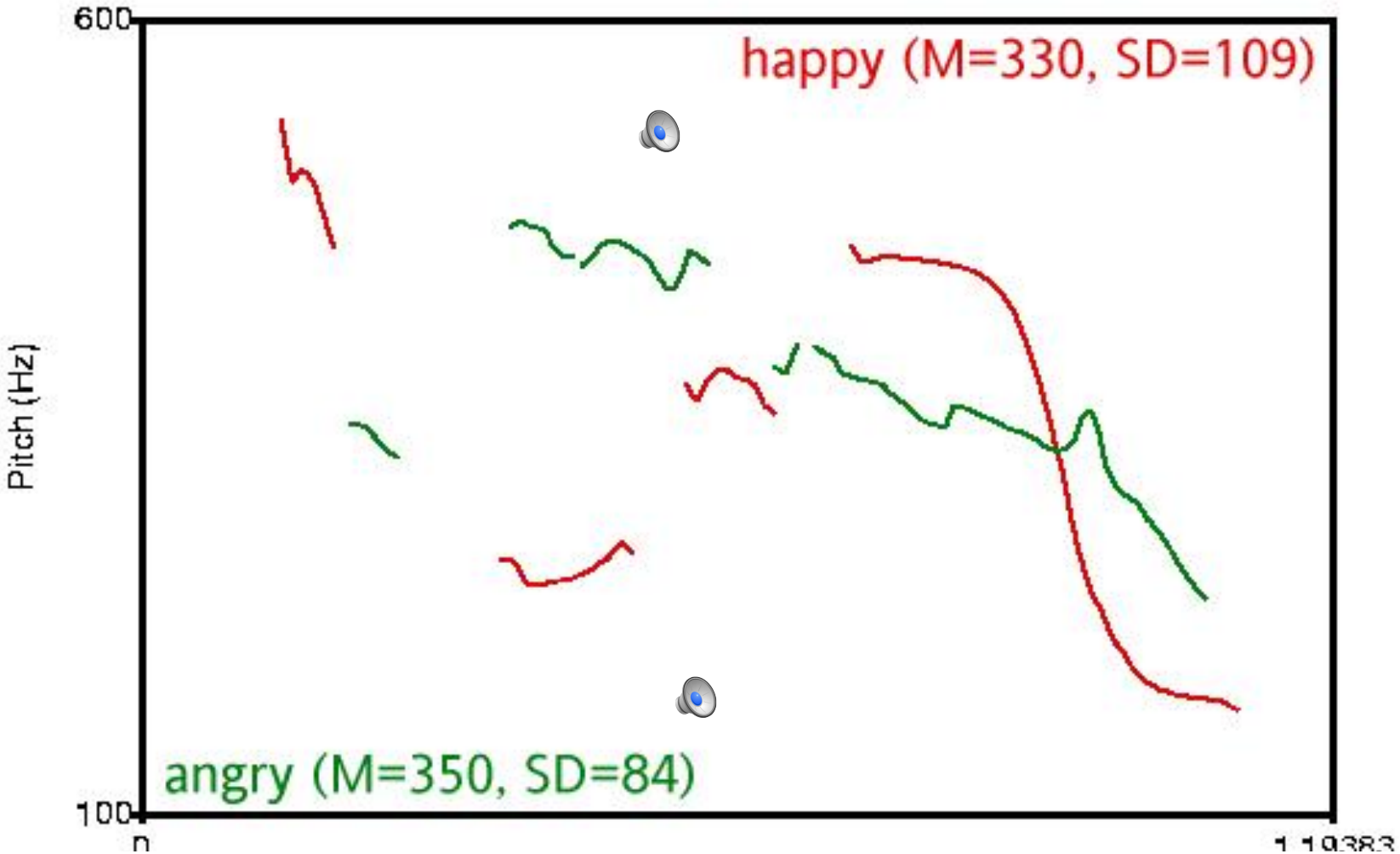
Slide from Julia Hirschberg

Major Problems for Classification: Different Valence/Different Activation



slide from Julia Hirschberg

But....
Different Valence/ Same Activation



slide from Julia Hirschberg

Data and tasks for Emotion Detection

- Scripted speech
  - Acted emotions, often using 6 emotions
  - Controls for words, focus on acoustic/prosodic differences
  - Features:
    - F0/pitch
    - Energy
    - Speaking rate
- Spontaneous speech
  - More natural, harder to control
  - Kinds of emotion focused on:
    - frustration,
    - annoyance,
    - certainty/uncertainty
    - “activation/hot spots”

Four quick case studies

- Acted speech:
  - LDC's EPSaT
- Annoyance/Frustration in natural speech
  - Ang *et al.* on Annoyance and Frustration
- Basic emotions cross linguistically (read on your own)
  - Braun and Katerbow, dubbed speach
- Uncertainty in natural speech:
  - Liscombe et al's ITSPOKE

Example 1: Acted speech: Emotional Prosody Speech and Transcripts Corpus (EPSaT)

- Recordings from LDC
  - http://www.ldc.upenn.edu/Catalog/LDC2002S28.html
- 8 actors read short dates and numbers in 15 emotional styles

Slide from Jackson Liscombe

EPSaT Examples

happy

sad

angry

confident

frustrated

friendly

interested

anxious

bored

encouraging

Slide from Jackson Liscombe

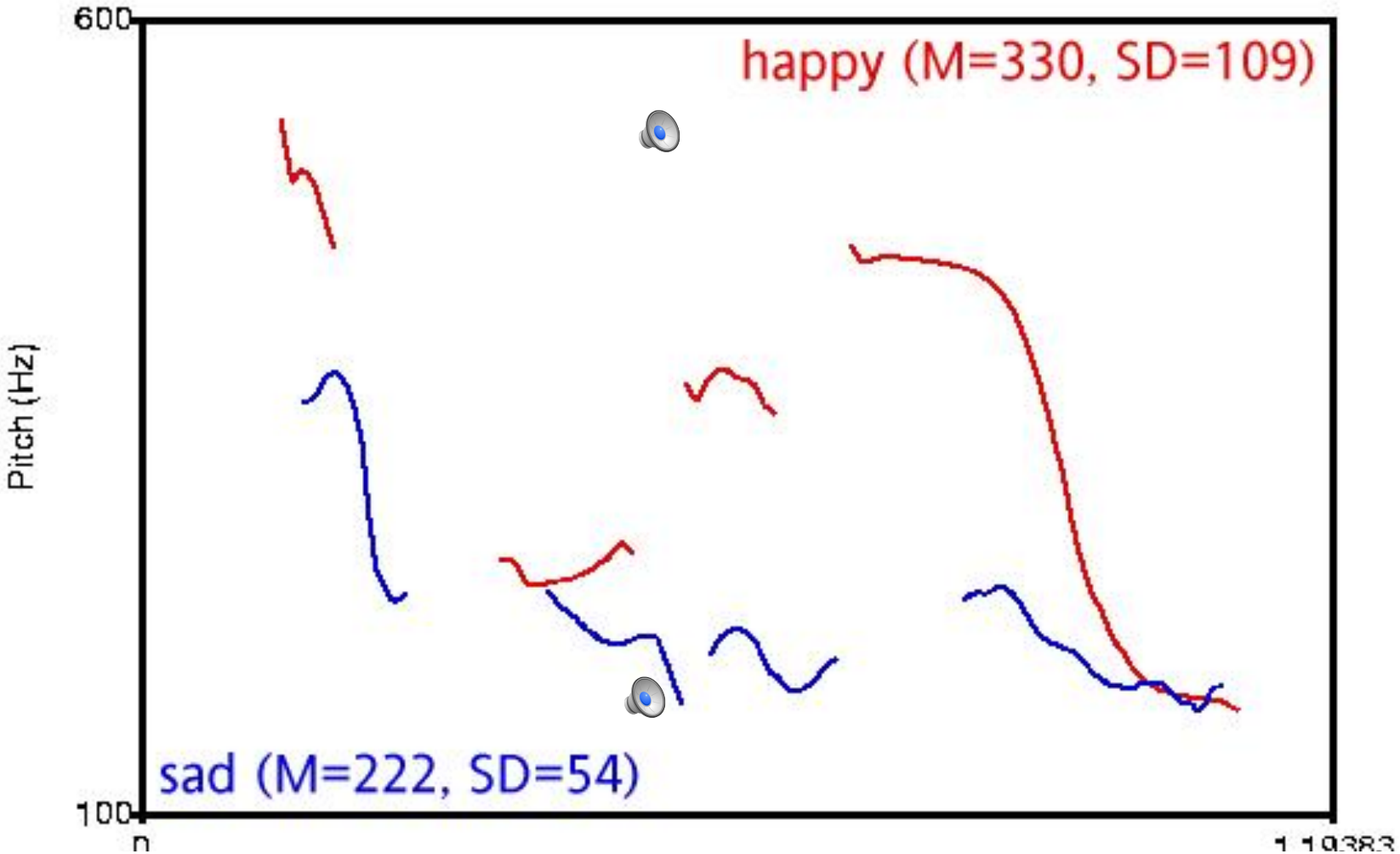
Liscombe et al. 2003 Features Automatic Acoustic-Prosodic

- **F0**
  - min, max, mean, range, stdev, above
- **Energy [RMS]**
  - min, max, mean, range, stdev, above
- **Voicing**
  - vcd (percentage of voiced frames)

Liscombe et al. 2003 Features Semi-Automatic Acoustic-Prosodic

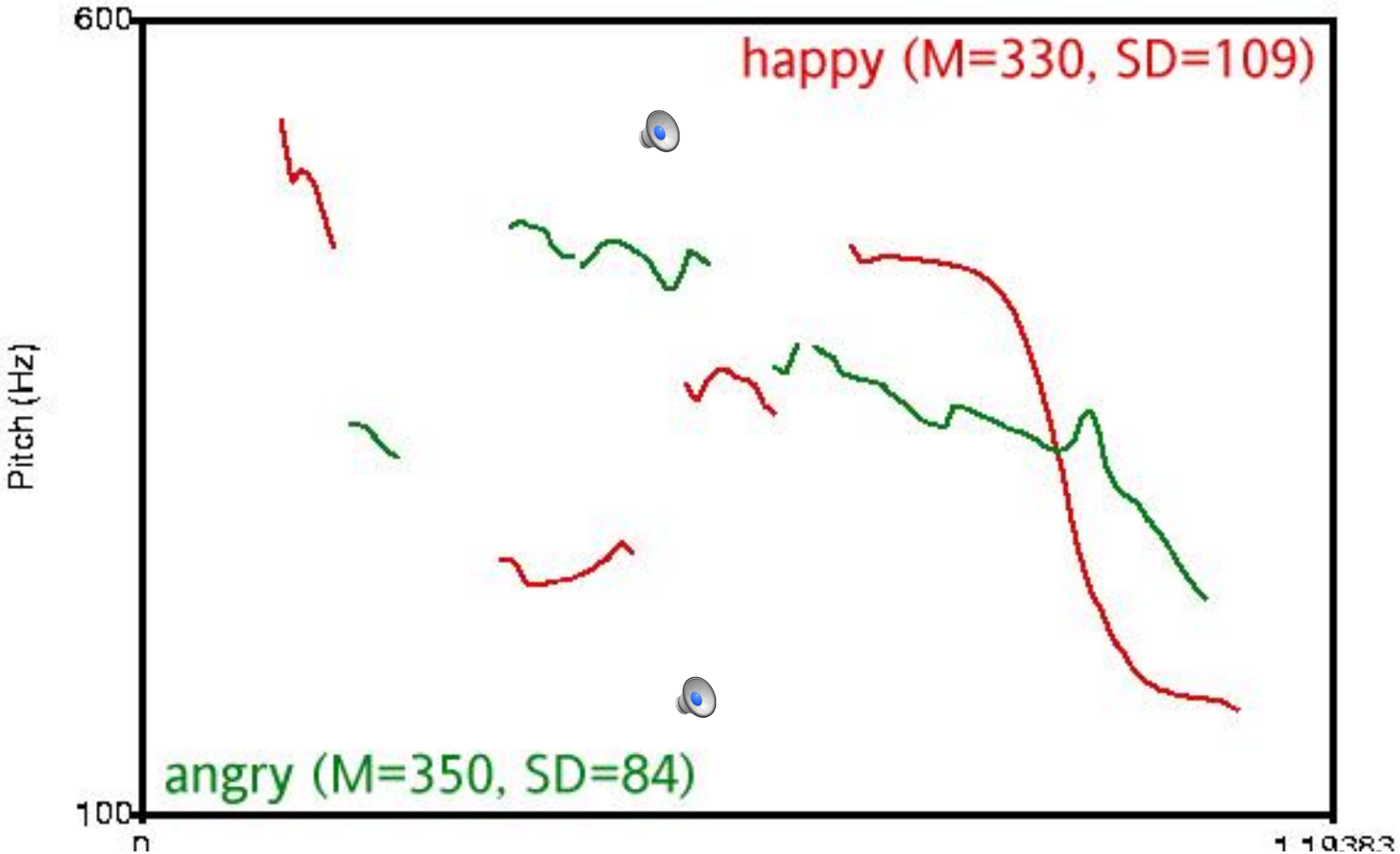
- Spectral Tilt: H2 – H1
  - computed over 30 ms window centered on middle of vowel
  1. Vowel with hand-labeled *nuclear* stress
     - main stressed vowel of the intonation phrase
  2. Loudest vowel
     - vowel with highest RMS
- Syllable Length
- ToBI
  - Nuclear accent type (L*, H*, L+H*)
  - Boundary tone type (H-H%, L-L%, etc)

Global Pitch Statistics



Slide from Jackson Liscombe

Global Pitch Statistics



Slide from Jackson Liscombe

Correlation between emotion and acoustics

Feature	sad	angry	bored	frust	anxs	friend	conf	happy	inter	encour
F0_MIN	-0.36		-0.36	-0.11		0.32	0.20	0.39	0.35	0.30
F0_MAX	-0.38	0.08	-0.51			0.31	0.24	0.39	0.42	0.29
F0_MEAN	-0.35		-0.53		0.10	0.32	0.23	0.39	0.43	0.29
F0_RANGE	-0.35	0.09	-0.47			0.28	0.23	0.34	0.38	0.25
F0_STDV					0.09					
F0_ABOVE		0.12	-0.09	0.12	0.14					
RMS_MIN	-0.16				-0.08		0.13	0.10		
RMS_MAX	-0.27	0.14	-0.37	0.10	0.08	0.11	0.22	0.21	0.26	0.14
RMS_MEAN	-0.28	0.12	-0.36		0.12	0.13	0.23	0.22	0.28	0.16
RMS_RANGE	-0.27	0.14	-0.37	0.10	0.08	0.11	0.22	0.20	0.27	0.14
RMS_STDV	-0.27	0.15	-0.35	0.10	0.08	0.10	0.23	0.20	0.26	0.13
VCD	-0.19		-0.10	-0.14	-0.17	0.16	0.23	0.23	0.14	0.20
SYL_LNGTH	0.23		0.23			-0.15	-0.09	-0.19	-0.19	-0.17
TILT_STRESS	-0.12	0.17		0.10	-0.11		0.18			
TILT_RMS		0.25	0.09	0.22		-0.17		-0.11		-0.13

positive-activation emotions (angry, frustrated, happy, confident):

high F0, RMS, speed

positive versus negative valence: TILT (positive valence = negative tilt)

Human labels for each sentence

	not at all	a little	somewhat	quite	extremely
How **frustrated** does this person sound?					
How **confident** does this person sound?					
How **interested** does this person sound?					
How **sad** does this person sound?					
How **happy** does this person sound?					
How **friendly** does this person sound?					
How **angry** does this person sound?					
How **anxious** does this person sound?					
How **bored** does this person sound?					
How **encouraging** does this person sound?					

Liscombe et al. Experiments

- Binary Classification for Each Emotion,
  - 'not at all' versus other
  - Ripper, 90/10 split
  - 75% accuracy compared to 62% most-frequent-class baseline

Emotion	Feature	Accuracy
angry	F0_*, RMS_*, TILT_*, VCD	77.27%
confident	F0_RANGE, F0_MEAN	76.14%
happy	F0_MIN	81.25%
interested	F0_STDV	75.57%
encouraging	VCD	73.86%
sad	F0_MAX	81.25%
anxious	TILT_RMS	78.41%
bored	TILT_RMS	80.11%
friendly	TILT_STRESS	75.00%
frustrated	F0_MAX	75.00%

Example 2 - Ang 2002

Ang, Shriberg, Stolcke. 2002. “Prosody-based automatic detection of annoyance and frustration in human-computer dialog”

DARPA Communicator “ Travel Planning” 837 dialogs, 21,819 utts

- How reliably can humans and machines label annoyance and frustration?
- What prosodic or other features are useful?

Data Annotation

- 5 undergrads with different backgrounds
- Each dialog labeled by 2+ people independently
  - 2nd “Consensus” pass for all disagreements, by two of the same labelers

Slide from Shriberg, Ang, Stolcke

Data Labeling

Emotion: neutral, annoyed, frustrated, tired/disappointed, amused/surprised, no-speech/NA

Speaking style: hyperarticulation, perceived pausing between words or syllables, raised voice

Repeats and corrections: repeat/rephrase, repeat/rephrase with correction, correction only

Miscellaneous useful events: self-talk, noise, non-native speaker, speaker switches, etc.

Slide from Shriberg, Ang, Stolcke

Emotion Samples

- Neutral
  - July 30 1
  - Yes 2
- Disappointed/tired
  - No 6
- Amused/surprised
  - No 7
- Annoyed
  - Yes 3
  - Late morning (HYP) 8
- Frustrated
  - Yes 4
  - No 5
  - 9
  - No, I am … (HYP) 10
  - There is no Manila…

Slide from Shriberg, Ang, Stolcke

Emotion Class Distribution

	Count	%
Neutral	17994	83.1
Annoyed	1794	8.3
No-speech	1437	6.6
Frustrated	176	0.8
Amused	127	0.6
Tired	125	0.6
TOTAL	21653	

To get enough data, grouped annoyed and frustrated, versus else (with speech)

Slide from Shriberg, Ang, Stolcke

Prosodic Model

- Classifier: CART-style decision trees
- Downsampled to equal class priors
- Automatically extracted prosodic features based on recognizer word alignments
- Used 3/4 for train, 1/4th for test, no call overlap

Slide from Shriberg, Ang, Stolcke

Prosodic Features

- Duration and speaking rate features
  - duration of phones, vowels, syllables
  - normalized by phone/vowel means in training data
    - true or recognized phones
  - normalized by speaker (all utterances, first 5 only)
  - speaking rate (vowels/time)
- Pause features
  - duration and count of utterance-internal pauses at various threshold durations
  - ratio of speech frames to total utt-internal frames

Slide from Shriberg, Ang, Stolcke

Features (cont.)

- Spectral tilt features
  - average of 1st cepstral coefficient
  - average slope of linear fit to magnitude spectrum
  - difference in log energies btw high and low bands
  - extracted from longest normalized vowel region

Slide from Shriberg, Ang, Stolcke

Pitch Features

- minimum and maximum utterance pitch
  - raw and speaker-normalized
- maximum pitch inside longest normalized vowel
- slopes at various locations
- normalized by speakers F0 range

Slide from Shriberg, Ang, Stolcke

Language Model Features

- Train two 3-gram class-based LMs
  - one on frustration, one on other.
- Given a test utterance, chose class that has highest LM likelihood (assumes equal priors)
- In prosodic decision tree, use sign of the likelihood difference as input feature

Slide from Shriberg, Ang, Stolcke

Results (cont.)

- H-H labels agree 72%
- H labels agree 84% with “consensus” (biased)
- Tree model agrees 76% with consensus-- better than original labelers with each other
- Language model features alone (64%) are not good predictors

Slide from Shriberg, Ang, Stolcke

Prosodic Predictors of Annoyed/Frustrated

- Pitch:
  - high maximum fitted F0 in longest normalized vowel
  - high speaker-norm. (1st 5 utts) ratio of F0 rises/falls
  - maximum F0 close to speaker's estimated F0 "topline"
  - minimum fitted F0 late in utterance (no "?" intonation)
- Duration and speaking rate:
  - long maximum phone-normalized phone duration
  - long max phone- & speaker- norm.(1st 5 utts) vowel
  - low syllable-rate (slower speech)

Slide from Shriberg, Ang, Stolcke

Ang et al '02 Conclusions

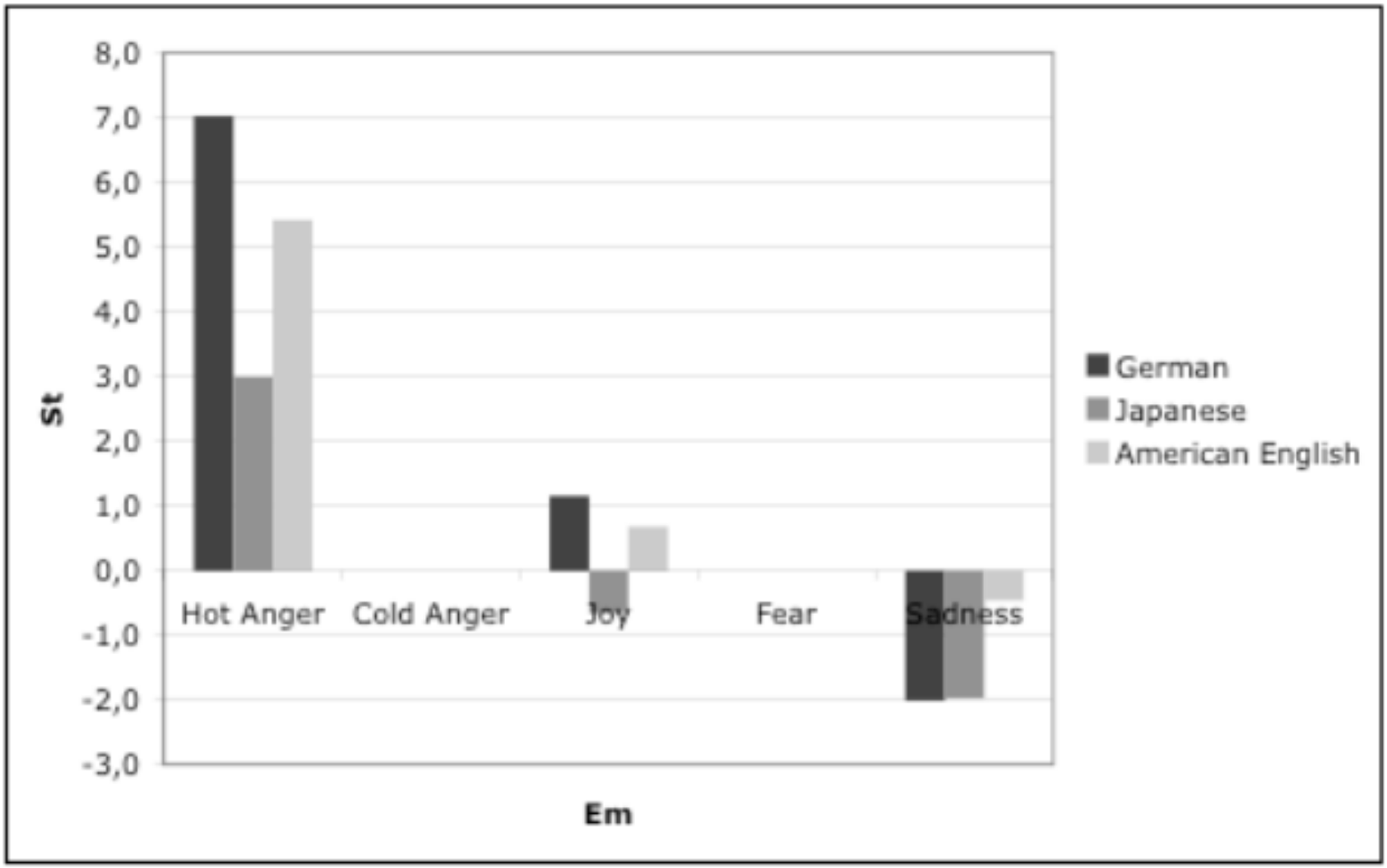
- Emotion labeling is a complex task
- Prosodic features:
  - duration and stylized pitch
  - speaker normalizations help

- "N-gram probability ratio" is a bad feature

Example 3: Basic Emotions across languages

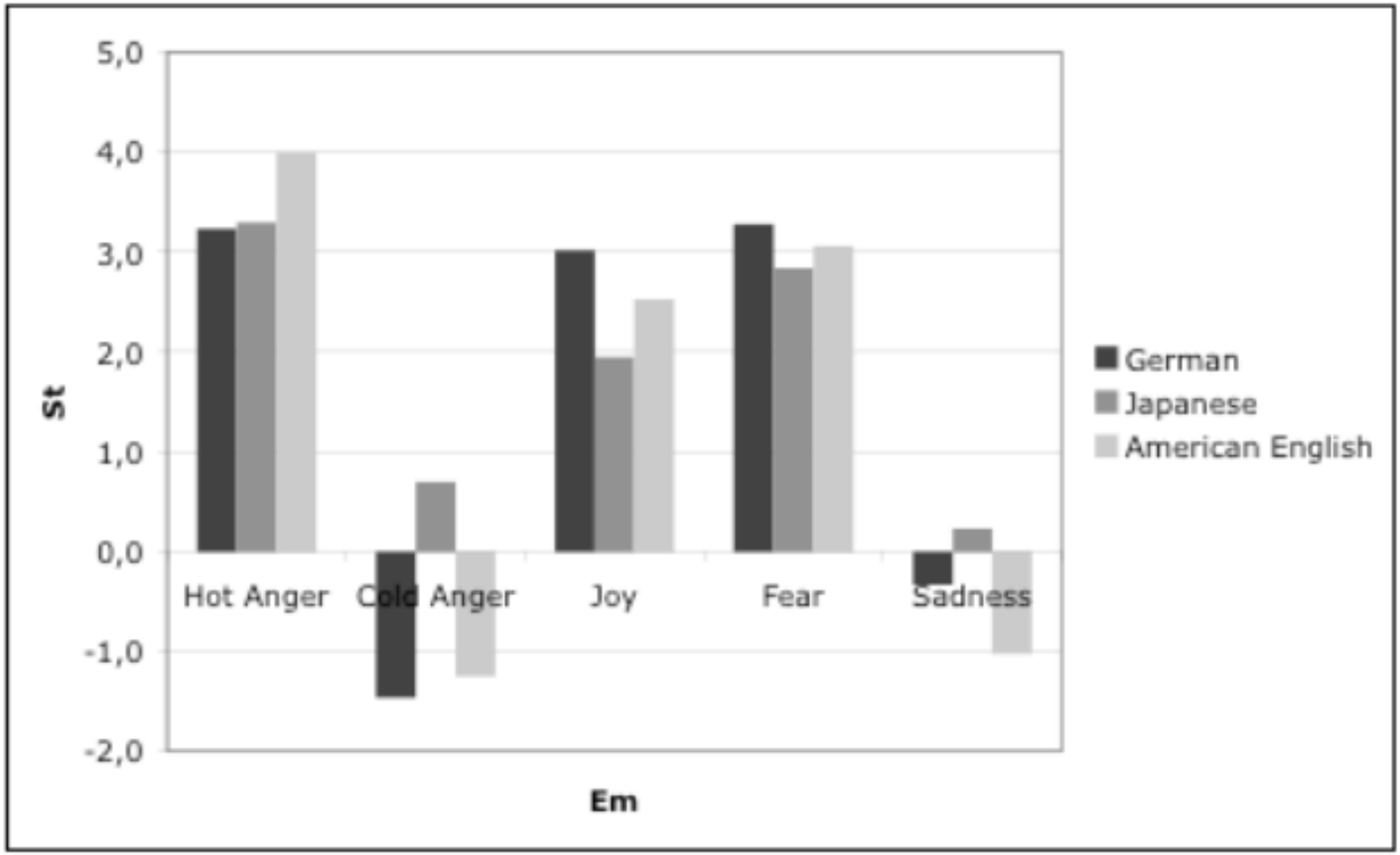
- Braun and Katerbow
- F0 and the basic emotions
- Using “comparable corpora”
  - English, German and Japanese
- Dubbing of the TV show Ally McBeal into German and Japanese

Results: Male speakers



Difference between emotional and neutral speech

Results: Female speakers



Difference between emotional and neutral speech

Perception

- A Japanese male joyful speaker:
- Confusion matrix: % of misrecognitions

- Japanese perceiver:

Fear	17
Joy	**40**
Neutral	28
Sadness	16
Anger	0

American perceiver:

Fear	19
Joy	**19**
Neutral	27
Sadness	33
Anger	1

Example 4: Intelligent Tutoring Spoken Dialogue System

ITSpoke

- Diane Litman, Katherine Forbes-Riley, Scott Silliman, Mihai Rotaru

- Jackson Liscombe, Julia Hirschberg, Jennifer J. Venditti. 2005. Detecting Certainness in Spoken Tutorial Dialogues

Tutorial corpus

- 151 dialogues from 17 subjects
- student first writes an essay, then discusses with tutor

- both are recorded with microphones
- manually transcribed and segmented into turns
- 6778 student utterances (average 2.3 seconds)
- each utterance hand-labeled for certainty

PROBLEM (TYPED): If a car is able to accelerate at 2 m/s^2, what acceleration can it attain if it is towing another car of equal mass?

ESSAY (TYPED): The maximum acceleration a car can reach when towing a car behind it of equal mass will be halved. Therefore, the maximum acceleration will be $1m/s^2$.

DIALOGUE (SPOKEN): *. . . 9.1 min. into session . . .*

TUTOR$_1$: Uh let us talk of one car first.

STUDENT$_1$: ok. *(EMOTION = NEUTRAL)*

TUTOR$_2$: If there is a car, what is it that exerts force on the car such that it accelerates forward?

STUDENT$_2$: The engine *(EMOTION = POSITIVE)*

TUTOR$_3$: Uh well engine is part of the car, so how can it exert force on itself?

STUDENT$_3$: um... *(EMOTION = NEGATIVE)*

Corpus Statistics

```
64.2%  neutral
18.4%  certain
13.6%  uncertain
 3.8%  mixed
```

Uncertainty in ITSpoke

um <sigh> I don't even think I have an idea here ..... now .. mass isn't weight ..... mass is ..... the ..... space that an object takes up ..... is that mass?

[71-67-1:92-113]

Slide from Jackson Liscombe

Acoustic-Prosodic Features

- 4 normalized fundamental frequency (f0) features: maximum, minimum, mean, standard deviation
- 4 normalized energy (RMS) features: maximum, minimum, mean, standard deviation
- 4 normalized temporal features: total turn duration, duration of pause prior to turn, speaking rate, amount of silence in turn

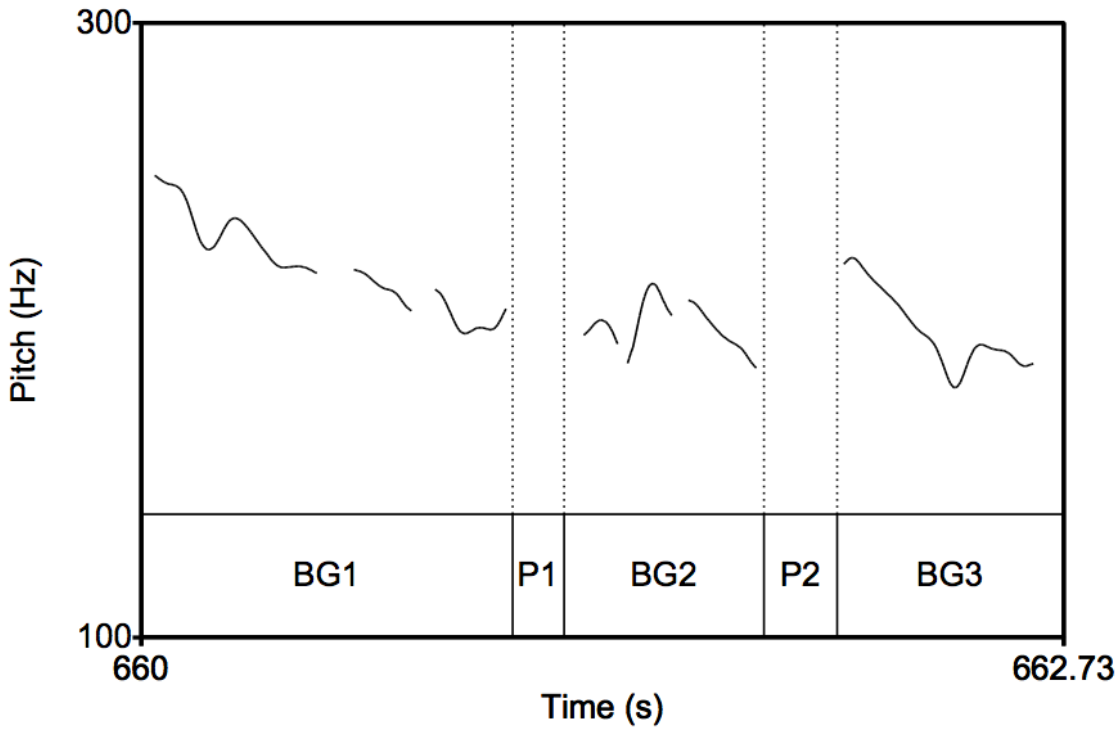
Non-Acoustic-Prosodic Features

- lexical items in turn
- 6 automatic features: turn begin time, turn end time, isTemporalBarge-in, isTemporalOverlap, #words in turn, #syllables in turn
- 6 manual features: #false starts in turn, isPriorTutorQuestion, isQuestion, isSemanticBarge-in, #canonical expressions in turn, isGrounding

Identifier Features: subject, subject gender, problem

Turns and Breath groups

- Turns were very long
- All features also extracted over breath groups
- Data was labeled only for turns, so use BG as feature
  - 45 features from first, last and longest BG



Liscombe et al: ITSpoke Experiment

- Human-Human Corpus
- AdaBoost(C4.5) 90/10 split in WEKA
- Classes: Uncertain vs Certain vs Neutral
- Results:

Features	Accuracy
Baseline	66%
Acoustic-prosodic	75%

Slide from Jackson Liscombe

A tutorial system that adapts to uncertainty

Forbes-Riley, Kate, and Diane Litman. "Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor." Speech Communication 53, no. 9 (2011): 1115-1136.

---

tutor1: What will the velocity of the object be a second after that (where the initial velocity is 9.8m/s and the acceleration is 9.8m/s^2)?
student1: Is it 19.6 m/s? *[Correct+Uncertain]*

---

nonadaptive-tutor2: Good. Now back to the man and his keys. We've shown that they both have an acceleration of 9.8m/s^2. So, how do their velocities compare with each other, at every second during the fall?

---

adaptive-tutor2: Good. A second later, its velocity will be 9.8m/s + 9.8m/s = 19.6m/s. This is because its acceleration tells us that every second, its velocity increases by 9.8m/s. So, what'll its velocity be, a second after that?
adaptive-student2: Um. I'm thinking. 29.4 m/s. *[Correct+Certain]*
adaptive-tutor3: Right. Now back to the man and his keys. We've shown that they both have an acceleration of 9.8m/s^2. So, how do their velocities compare with each other, at every second during the fall?

Features for uncertainty

Acoustic-Prosodic Features

4 fundamental frequency (f0) features: maximum, minimum, mean, standard deviation

4 energy (RMS) features: maximum, minimum, mean, standard deviation

3 temporal features: turn duration, prior pause duration, internal silence

Lexical and Dialogue Features

ITSPOKE-recognized lexical items in turn

tutor goal name

problem name

turn number

per-dialogue running totals and averages for 11 acoustic-prosodic features

Identifier Feature:

subject gender

Conclusions

- Uncertainty is very hard to detect
- Certainty is easier
- Even so, the system improved learner outcomes

Disengagement in ITSpoke 2

Kate Forbes-Riley, Diane Litman, Heather Friedberg, Joanna Drummond. 2012. Intrinsic and Extrinsic Evaluation of an Automatic User Disengagement Detector for an Uncertainty-Adaptive Spoken Dialogue System. NAACL 2012.

T_1: What is the definition of Newton's Second Law?

U_1: I have no idea <*sigh*>. **(DISE,** ***incorrect*****, UNC)**

. . .

T_2: What's the numerical value of the man's acceleration? Please specify the units too.

U_2: The speed of the elevator. Meters per second. **(DISE,** ***incorrect*****, UNC)**

. . .

T_3: What are the forces acting on the keys after the man releases them?

U_3: graaa-vi-tyyyyy <*sings the answer*> **(DISE,** ***correct*****, CER)**

Figure 1: Corpus Example Illustrating the User Turn Labels ((Dis)Engagement, (In)Correctness, (Un)Certainty)

Disengagement Features

- **Acoustic-Prosodic Features**

  temporal features: turn duration, prior pause duration, turn-internal silence

  fundamental frequency (f0) and energy (RMS) features: maximum, minimum, mean, std. deviation

  running totals and averages for all features

- **Lexical and Dialogue Features**

  dialogue name and turn number

  question name and question depth

  ITSPOKE-recognized lexical items in turn

  ITSPOKE-labeled turn (in)correctness

  incorrect runs

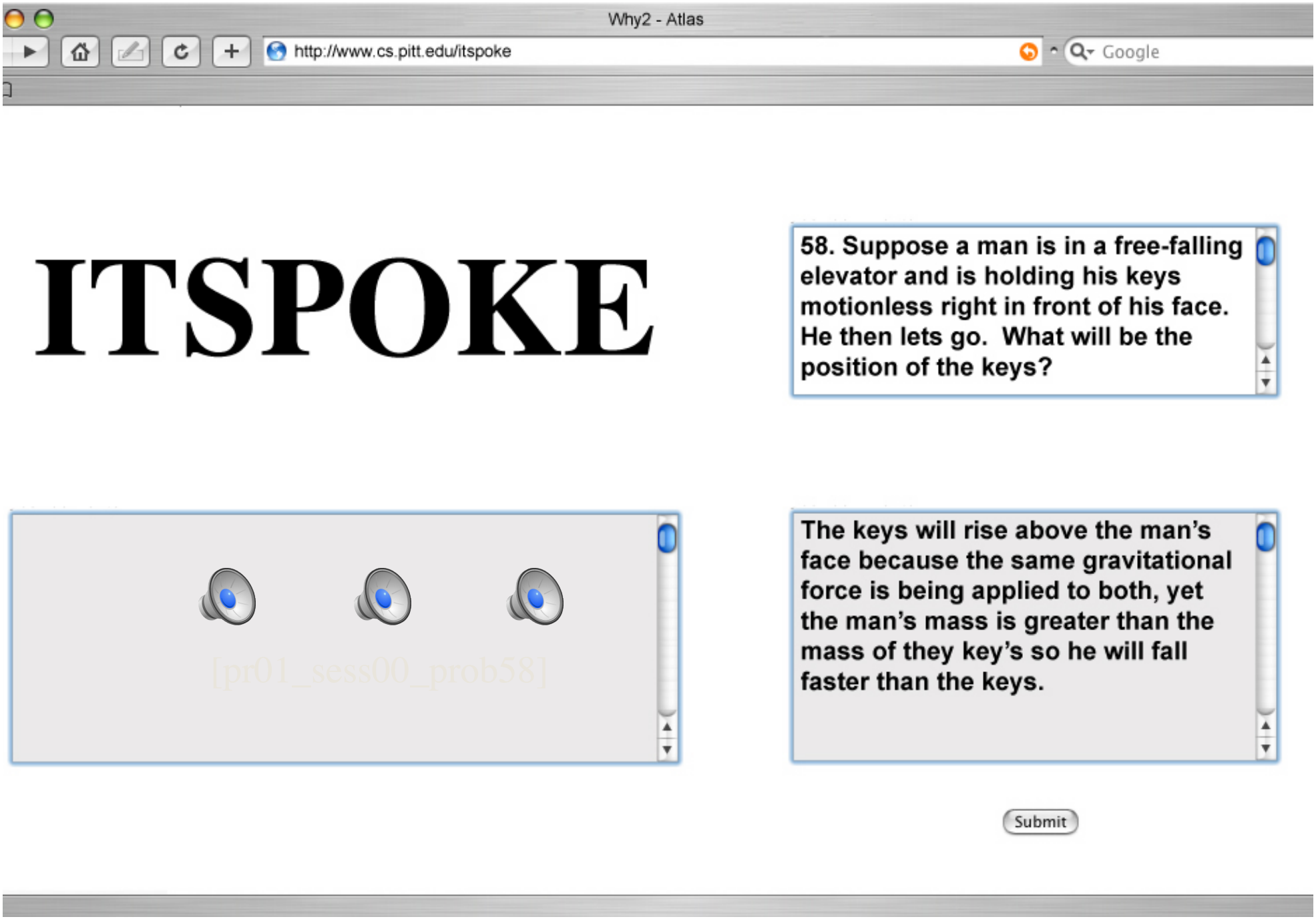
- **User Identifier Features**:

  gender and pretest score

Upper level of tree consists entirely of prosody, question name/depth

Most important feature: Pause prior to start of turn

<250ms means disengagement!!!!



Slide from Jackson Liscombe

Scherer summary: Prosodic features for emotion

	Stress	Anger/rage	Fear/panic	Sadness	Joy/elation	Boredom
Intensity	↗	↗	↗	↘	↗	
F0 floor/mean	↗	↗	↗	↘	↗	
F0 variability		↗		↘	↗	↘
F0 range		↗	↗(↘)	↘	↗	↘
Sentence contours		↘		↘		
High frequency energy		↗	↗	↘	(↗)	
Speech and articulation rate		↗	↗	↘	(↗)	↘

Interpersonal stance

Scherer's typology of affective states

Emotion: relatively brief episode of synchronized response of all or most organismic subsystems in response to the evaluation of an external or internal event as being of major significance

angry, sad, joyful, fearful, ashamed, proud, desperate

Mood: diffuse affect state …change in subjective feeling, of low intensity but relatively long duration, often without apparent cause

cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stance: affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange

distant, cold, warm, supportive, contemptuous

Attitudes: relatively enduring, affectively colored beliefs, preferences predispositions towards objects or persons

liking, loving, hating, valuing, desiring

Personality traits: emotionally laden, stable personality dispositions and behavior tendencies, typical for a person

nervous, anxious, reckless, morose, hostile, envious, jealous

Interpersonal Stance: Our Goals

- Friendliness
- Assertiveness
- Flirtation
- Awkwardness

Methodology

- Speed-dating
- Participants rate each other

speed dating *noun*

Menu

speed dating [uncountable]

an event at which you meet and talk to a lot of different people for only a few minutes at at a time. People do this in order to try to meet someone and have a romantic relationship.

E.J. Finkel, P.W. Eastwick. 2008. Speed-dating. *Current Directions in Psychological Science*, 17 (3) (2008), p. 193

Place, S. S., Todd, P. M., Penke, L., & Asendorpf, J. B. (2009). The ability to judge the romantic interest of others. *Psychological Science*, 20(1), 22-26.

M.E. Ireland, R.B. Slatcher, P.W. Eastwick, L.E. Scissors, E.J. Finkel, J.W. Pennebaker. 2011. Language style matching predicts relationship initiation and stability. *Psychological Science*, 22 (1) (2011), p. 39

What do you do for fun? Dance?

Uh, dance, uh, I like to go, like camping. Uh, snowboarding, but I'm not good, but I like to go anyway.

You like boarding.

Yeah. I like to do anything. Like I, I'm up for anything.

Really?

Yeah.

Are you open-minded about most everything?

Not everything, but a lot of stuff-

What is not everything [laugh]

I don't know. Think of something, and I'll say if I do it or not. [laugh]

Okay. [unintelligible].

Skydiving. I wouldn't do skydiving I don't think.

Yeah I'm afraid of heights.

F: Yeah, yeah, me too.

M: [laugh] Are you afraid of heights?

F: [laugh] Yeah [laugh]

Background: Previous work on Pickiness in Dating

- Finkel and Eastwick 2009, Psych Science
- Men are less selective than women in speed dating
- Novel explanation: act of physically approaching a partner increases attraction to that partner
  - traditional events, always men rotates
- Ran 15 speed dating events
  - in 8, men rotated: men more selective
  - in 7, women rotated: men equally selective to women
- Conclusion?

Background: Friendliness

- English (Liscombe et al.; 2003)
  - Friendly speech: higher f0 min, mean, max
    - but all other positive valence similar
  - Higher spectral tilt (H2-H1) of stressed vowel
- Swedish (House; 2005)
  - Higher F0 in questions (especially late in syllable) friendlier than low F0 or a peak early in the syllable.
- Chinese (Chen et al. 2004, Li and Wang 2004)
  - statements and questions produced by actors,
  - Friendly speech had higher mean F0, faster

Background: Flirtation/Attractiveness

- Attractiveness
  - Raised F0 in women's voices
    - Preferred by men (Feinberg et al 2008, Jones et al 2010)
    - Rated more attractive by men (Collins and Missing, 2003; Puts et al., 2011).
  - Lowered F0 or close harmonics in men's voices
    - labeled by women as more attractive or masculine (Collins and Missing, 2003; Puts et al., 2011).
- Flirtatiousness:
  - Higher F0 or dispersed formants in women's voices
    - perceive as more flirtatious by other women (Puts et al., 2011).

Extracting social meaning

- Stance
  - Friendly, flirt, awkward, assertive
- Social Bond
  - Clicking or Connection
  - Romantic Interest
- **946 4-minute dates**
  - ~800K words, hand-transcribed
  - ~60 hours, from shoulder sash recorders
  - 3 events, 20x20=400 dates x 3
  - Date perceptions, demographics, preferences

Data annotation

- Each speaker wore a microphone
- So each date had two recordings
- The wavefile from each speaker was manually segmented into 4-minute dates
- Professional transcription service produced:
  - words, laughter, disfluencies
  - timestamps for turn beginning and end (1 second)
    - for 10% of the dates, timestamp at 0.1 second granularity
  - using both recordings

Study 1:
What we attempted to predict

- Conversational style:
  - How often did they behave in the following ways on this date?
    - On a scale of 1-10 (1=never, 10=constantly)

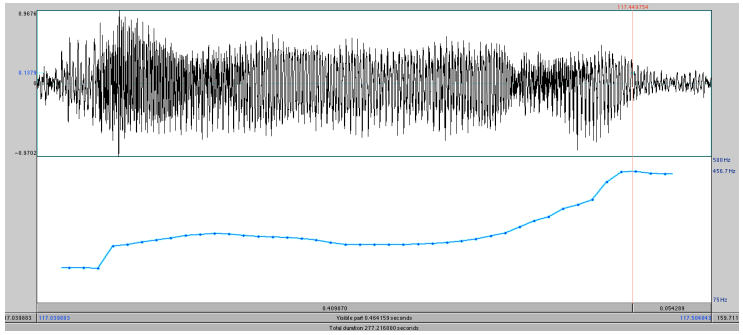
awkward

friendly

flirtatious

assertive

Features



- Prosodic
  - pitch (min, mean, max, std)
  - intensity (min, max, mean, std)
  - duration of turn
  - rate of speech (words per second)
- Lexical
  - negation words (don’t, didn’t, won’t, can’t, not, never)
  - hedges (kind of, sort of, probably, I don’t know)
  - personal pronouns (I, you, we, us)
- Dialog
  - questions
  - backchannels (“uh-huh”, “yeah”)
  - appreciations (“Wow!”, “That’s great!”)
  - sympathy (“That’s awful!” “Oh, that sucks!”)

LIWC

Linguistic Inquiry and Word Count

Pennebaker, Francis, & Booth, 2001

dictionary of 2300 words grouped into > 70 classes, modified:

I: I'd, I'll, I'm, I've, me, mine, my, myself (not counting I mean)

YOU: you, you'd, you'll, your, you're, yours, you've (not counting you know)

SEX: sex, sexy, sexual, stripper, lover, kissed, kissing

LOVE: love, loved, loving, passion, passions, passionate

HATE: hate, hates, hated

SWEAR: suck*, hell*, crap*, shit*, screw*, damn*, heck, f.ck*, ass*, ...

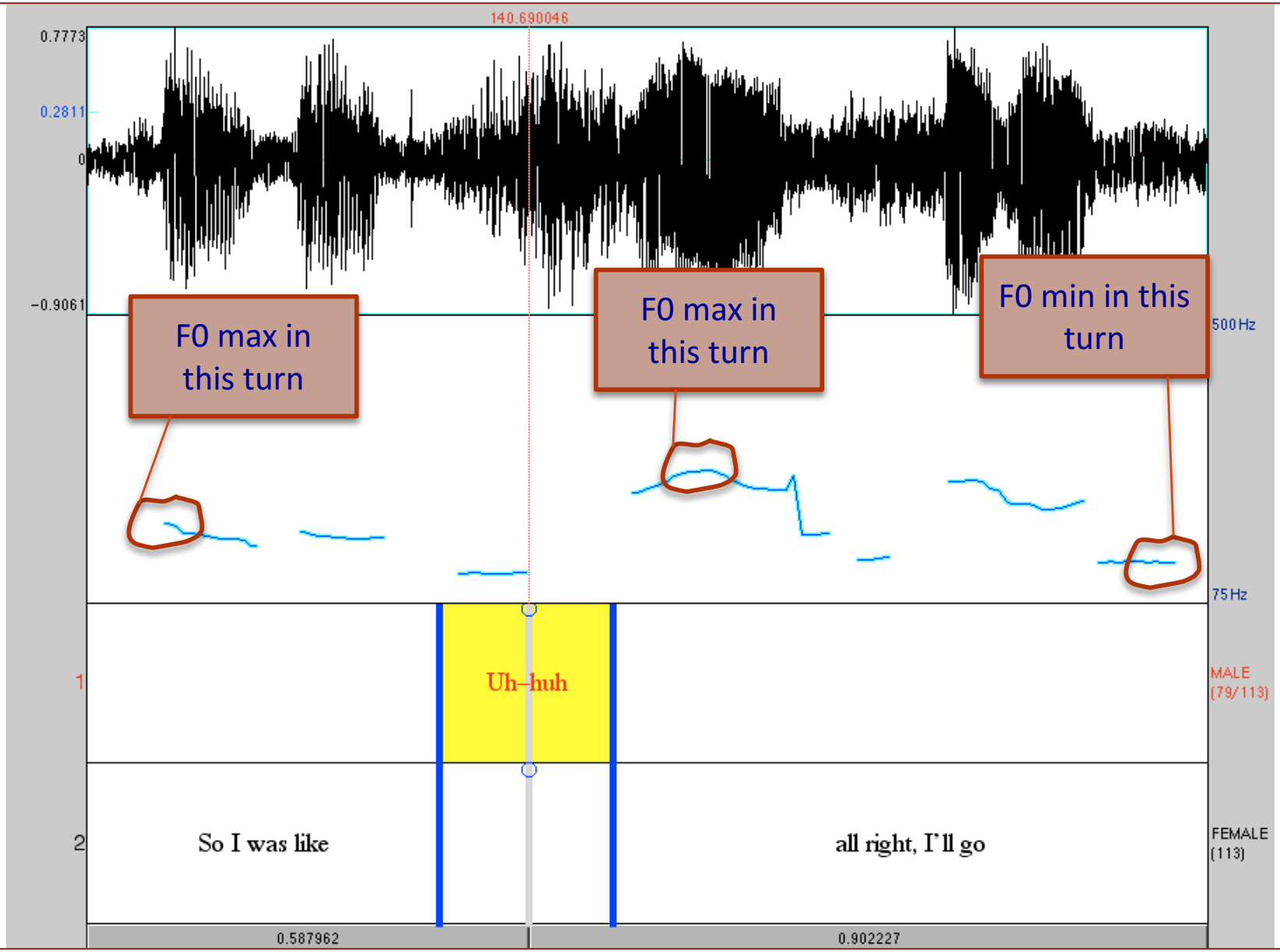
NEGEMOTION: bad, weird, hate, crazy, problem*, difficult, tough, awkward, boring

NEGATE: don't, not, no, didn't, never, can't, doesn't, wasn't, nothing, isn't, …

Additional lexical features

- Hedges
  - kind of, sort of, a little, I don't know, I guess
- Work terms
  - research, advisor, lab, work, finish, PhD, department
- Metadiscussion of dating
  - speed date, flirt, event, dating, rating
- UH or UM:
  - M: Um, eventually, yeah, but right now I want to get some more experience, uh, in research.
- Like, you know, I mean:

Speed date features extracted within turns: used for whole side



Features: Pitch

- F0 min, max, mean
  - Thus to compute, e.g., F0 min for a conversation side
    - Take F0 min of each turn (not counting zero values)
    - Average over all turns in the side
    - “F0 min, F0 max, F0 mean”
  - We also compute measures of variation
    - Standard deviation, pitch range
    - F0 min sd, F0 max sd, F0 mean sd
    - pitch range = (f0 max – f0 min)

Prosodic features

F0 MIN	minimum (non-zero) F0 per turn, averaged over turns
F0 MIN SD	standard deviation from F0 min
F0 MAX	maximum F0 per turn, averaged over turns
F0 MAX SD	standard deviation from F0 max
F0 MEAN	mean F0 per turn, averaged over turns
F0 MEAN SD	standard deviation (across turns) from F0 mean
F0 SD	standard deviation (within a turn) from F0 mean, averaged over turns
F0 SD SD	standard deviation from the f0 sd
PITCH RANGE	f0 max - f0 min per turn, averaged over turns
PITCH RANGE SD	standard deviation from mean pitch range
RMS MIN	minimum amplitude per turn, averaged over turns
RMS MIN SD	standard deviation from RMS min
RMS MAX	maximum amplitude per turn, averaged over turns
RMS MAX SD	standard deviation from RMS max
RMS MEAN	mean amplitude per turn, averaged over turns
RMS MEAN SD	standard deviation from RMS mean
TURN DUR	duration of turn in seconds, averaged over turns
TURN DUR SD	standard deviation of turn duration

Replace 18 factors with 6

- Factor analysis, 6 factors explain 85% of variance

	Factor1 Max F0	Factor2 Loudness	Factor3 Min F0	Factor4 Var Loudness	Factor5 Long Turn	Factor6 Var. F0
avtndur	20	5	-18	-1	91	-21
sdtndur	6	1	-5	2	95	-7
avpmin	-34	12	84	-3	-23	8
sdpmin	-16	0	91	2	-2	18
avpmax	92	10	1	3	22	-7
sdpmax	-75	-5	-17	6	-2	53
avpmean	59	19	67	-1	-6	-3
sdpmean	12	-4	30	10	-14	73
avpsd	91	-8	-21	1	-3	20
sdpsd	-34	-10	-1	5	-20	76
avimin	-1	51	13	-66	-10	20
sdimin	-5	22	8	66	7	16
avimax	9	91	2	-1	7	-14
sdimax	1	-59	0	70	-6	10
avimean	4	94	12	-23	-1	-1
sdimean	4	-29	-2	89	-9	5
avprange	90	5	-25	4	26	-8
sdprange	-74	-5	15	8	7	51

Prosodic Features: 6 factors

- Higher Pitch Ceiling
- Louder (Min, mean and max)
- Higher Pitch Floor
- Variable Loudness
- Longer Turns
- Variable Pitch

- plus Rate of Speech

Positive and negative assessments

(Goodwin, 1996; Goodwin and Goodwin, 1987; Jurafsky et al., 1998

Sympathy

(that's|that is|that seems|it is|that sounds)

(very|really|a little|sort of)?

(terrible|awful|weird|sucks|a problem|tough|too bad)

Appreciations ("Positive feedback")

(Oh)? (Awesome|Great|All right|Man|No kidding|wow|my god)

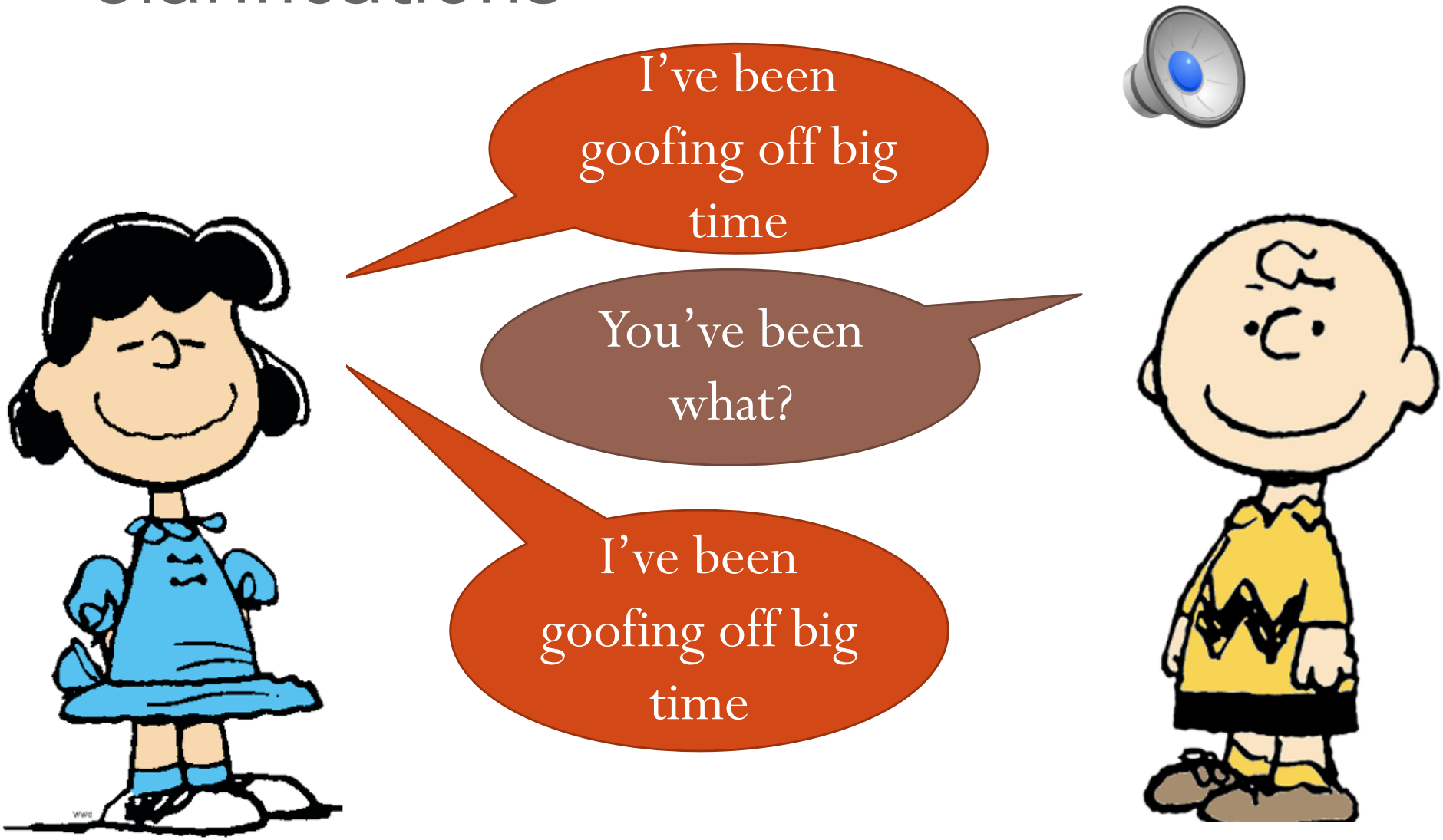
That

('s|is|sounds|would be)

(so|really)?

(great|funny|good|interesting|neat|amazing|nice|not bad|fun)

Clarifications



Regular Expression Patterns for Clarifications

- What?
- Sorry
- Excuse me
- Huh?
- Who?
- Pardon?
- Say again?
- Say it again?
- What's that
- What is that

Turn-initial Laughter

Laughing at your date's joke:

MALE: .. "speed filling out forms" is what this should be called.

FEMALE: [laughter] Yeah.

or teasing:

MALE: You're on the rebound.

FEMALE: huh--uh.

MALE: [laughter] Defensive.

Turn medial/final Laughter

Laughing at yourself

- FEMALE: Why are you single?
- MALE: Why I'm single? That's a good question. [laughter]

- MALE: And I stopped by the--the beer party the other day.
- FEMALE: Oh goodness. And you saw me in- [laughter]
-

Other features

- DISFLUENT RESTARTS : # of disfluent restarts in side
  - Uh, I–there's a group of us that came in–
- INTERRUPT: # of turns in side where speakers overlapped
  - M: But-and also obviously–
    - F: It sounds bigger.
  - M: –people in the CS school are not quite as social in general as other–

Accommodation features

- Speakers change their behavior to match (or not match) their interlocutor

  Natale 1975, Giles, Mulac, Bradac, & Johnson 1987, Bilous & Krauss 1988, Giles, Coupland, and Coupland, 1991, Giles and Coupland 1992, Niederhoffer and Pennebaker 2002, Pardo 2006, Nenkova and Hirschberg 2008, inter alia.

  - Matching rate of speech
  - Matching F0
  - Matching intensity (loudness)
  - Matching vocabulary and grammar
  - Matching dialect
- Our question:
  - Is accommodation characteristic of certain interpersonal styles?

Simple measures of accommodation

- Words that I used in turn i
  - that you used in turn i-1
    - function words (I, you, the, if, and, it, to…)
    - content words (department, party, lunch….)
- correlation between our rates of speech
  - if I get faster do you get faster?
- laugh accommodation
  - if I laugh do you laugh in the next turn?

Various sets of studies

- Social science
  - mixed-effects logistic regressions with many control factors
    - BMI, height, age difference
    - foreign versus non-foreign student
- Engineering
  - various classifiers, without the control factors

Controls for Social Science Studies

- **Actor and Partner Traits**
  - Male gender – male (1,0)
  - Height – inches (standardized by gender)
  - BMI – Body mass index = weight (lb) / [height (in)]2 x 703 BM standardized by gender)
  - Foreign – foreign born (1,0)
  - Dating experience – respondent's dating expertise

    (7=several times a week, 6=twice a week, 5=once a week, 4=twice a month, 3=once a month, 2=several times a year, 1=almost never)
  - Looking for relationship – whether respondent is seeking relationship or not (goal is to meet new people, get a date, or a serious relationship, = 1; if it seemed like fun, to say they did it, or other, = 0)

- **Dyad Traits**
  - Order – date's order in evening (1=first, 20=last).
  - Met before – dummy variable for knowing one another prior.
  - Age difference – actor's age in days – partner's age in days.

Feature Normalization

- word features are normalized by speakers total #words
- log rate of speech
- All the features standardized (mean=0, variance=1) globally across the training set before training.

Engineering studies
16 binary classifiers

- Female ±Awkward, Male ±Awkward,
- Female ±Friendly, Male ±Friendly,
- Female ±Flirtatious, Male ±Flirtatious,
- Female ±Assertive, Male ±Assertive

- Each study run twice, on:
  - self-assessed
  - alter-assessed
- Multiple classifier experiments
  - L1-regularized logistic regression
  - SVM w/RBF kernel

Test set

- For each of the 16 experiments
  - Sort all 946 dates
  - Choose top 10% as positive class
  - Choose bottom 10% as negative class
  - ignore 80% of dates in the middle!
- 5-fold cross-validation within this small training and test set

- Goal: distinguishing social interactants who are reported to exhibit (or not exhibit) clear social intentions or styles

Results using SVM Classifier

Using my speech to predict what my date says about me

	Male speaker	Female speaker
Flirting	65%	78%
Friendly	71	64
Awkward	67	67
Assertive	65	69

Results using SVM Classifier

- Using my speech to predict what I say about myself

	Male speaker	Female speaker
Flirting	66%	74%
Friendly	76	71
Awkward	63	67
Assertive	73	64

What do flirters do?

- Women when flirting:
  - raise pitch ceiling
  - talk faster
  - say “I” and “like”, use more hedges
  - laugh at themselves
- Men when flirting:
  - raise their pitch floor
  - laugh at their date (teasing?)
  - say “you”
  - don’t use words related to academics
  - say “um”, “I mean”, “you know”

Unlikely words for male flirting

academia

interview

teacher

phd

advisor

lab

research

management

finish

What makes someone seem friendly? "Collaborative conversational style"

Related to the "collaborative floor" of Edelsky (1981), Coates (1996)

- **Friendly people:**
  - laugh at themselves
  - don't use negative emotions
- **Friendly men**
  - are sympathetic and agree more often
  - don't interrupt
  - don't use hedges
- **Friendly women:**
  - higher max pitch
  - laugh at their date

What makes an awkward conversationalist?

- Awkward people:
  - use more hedges
  - ask more questions
- Awkward men
  - don’t talk about academics
  - do swear or use negative emotion
- Awkward women:
  - do talk about academics
  - talk more, and talk faster
  - don’t laugh at their date
  - don’t use “I”

Assertive

- Assertive men
  - talk more
  - use more negative emotion
  - lower their pitch floor
  - use more agreements and appreciations
  - use more “um”, “you”
  - use less negation
- Assertive women:
  - use more negation (“no”, “didn’t”, “don’t”)
  - talk about academics
  - are less sympathetic
  - accommodate more (content words)
  - use more “I” and “I mean”
  - use less negative emotion

How useful are linguistic features?

How useful are linguistic features?

	Linguistic features	Height, weight, etc features	Both
Male flirt	66	64	72
Female flirt	74	55	76
Male assert	73	55	72

Nonlinguistic features

- Nonlinguistic features help mainly in detecting flirting men
- Men are more likely to (say they) flirt:
  - If alter has low BMI
  - if self has high BMI
  - later in the evening

- (Men more sensitive to physical features?)

Study on Bond formation

Click:

- How well did I click with this person? (1-10)

Willing:

- Do I want to go on a date with this person?

How does clicking happen?

- Sociology literature:
  - bonding or “sense of connection” is caused by
    - **homophily**: select mate who shares your attributes and attitudes
    - **motives and skills**
    - **mutual coordination and excitement**
      - (Durkheim: religious rituals, unison singing, military)
- What is the role of language?
  - Background: speed dating has power asymmetry
    - **women are pickier**
    - Lot of other asymmetric role relationships (teacher-student, doctor-patient, boss-employee, etc.)

Our hypothesis:
targeting of the empowered party

- The conversational target is the woman
  - both parties should talk about her more
- The woman's face is important
  - the man should align to the woman and show understanding
- The woman's engagement is key
  - in a successful bonding, she should be engaged

Results: Clicking associated with:

Hierarchical regression dyad model, net of actor, partner, dyad features

- both parties talk about the woman
  - women use *I,*
  - men use *you*
- man supports woman's face
  - men use *appreciations* and *sympathy*,
  - men *accommodate* women's laughter
  - *men* interrupt with *collaborative completions*
- woman is engaged
  - women raise their pitch, vary loudness and pitch
  - women avoid hedges

Conclusions

- How to date:
  - Don’t talk about your advisor
  - Focus on the empowered party
  - Flirting women raise pitch ceiling – flirting men raise pitch floor
- How to be friendly
  - be sympathetic, ask clarification questions, agree, accommodate