CS 224S / LINGUIST 285
Spoken Language Processing

Andrew Maas

Stanford University

Spring 2017

Lecture 16: Parametric TTS, Intoxication, Depression, Trauma, Personality

Original slides by Dan Jurafsky

Evaluation of TTS

- Intelligibility Tests
  - Diagnostic Rhyme Test (DRT)
    - Humans do listening identification choice between two words differing by a single phonetic feature
      - Voicing, nasality, sustenation, sibilation
    - 96 rhyming pairs
    - Veal/feel, meat/beat, vee/bee, zee/thee, etc
      - Subject hears "veal", chooses either "veal or "feel"
      - Subject also hears "feel", chooses either "veal" or "feel"
    - % of right answers is intelligibility score.
- Overall Quality Tests
  - Have listeners rate space on a scale from 1 (bad) to 5 (excellent) (Mean Opinion Score)
- AB Tests (prefer A, prefer B) (preference tests)

Huang, Acero, Hon

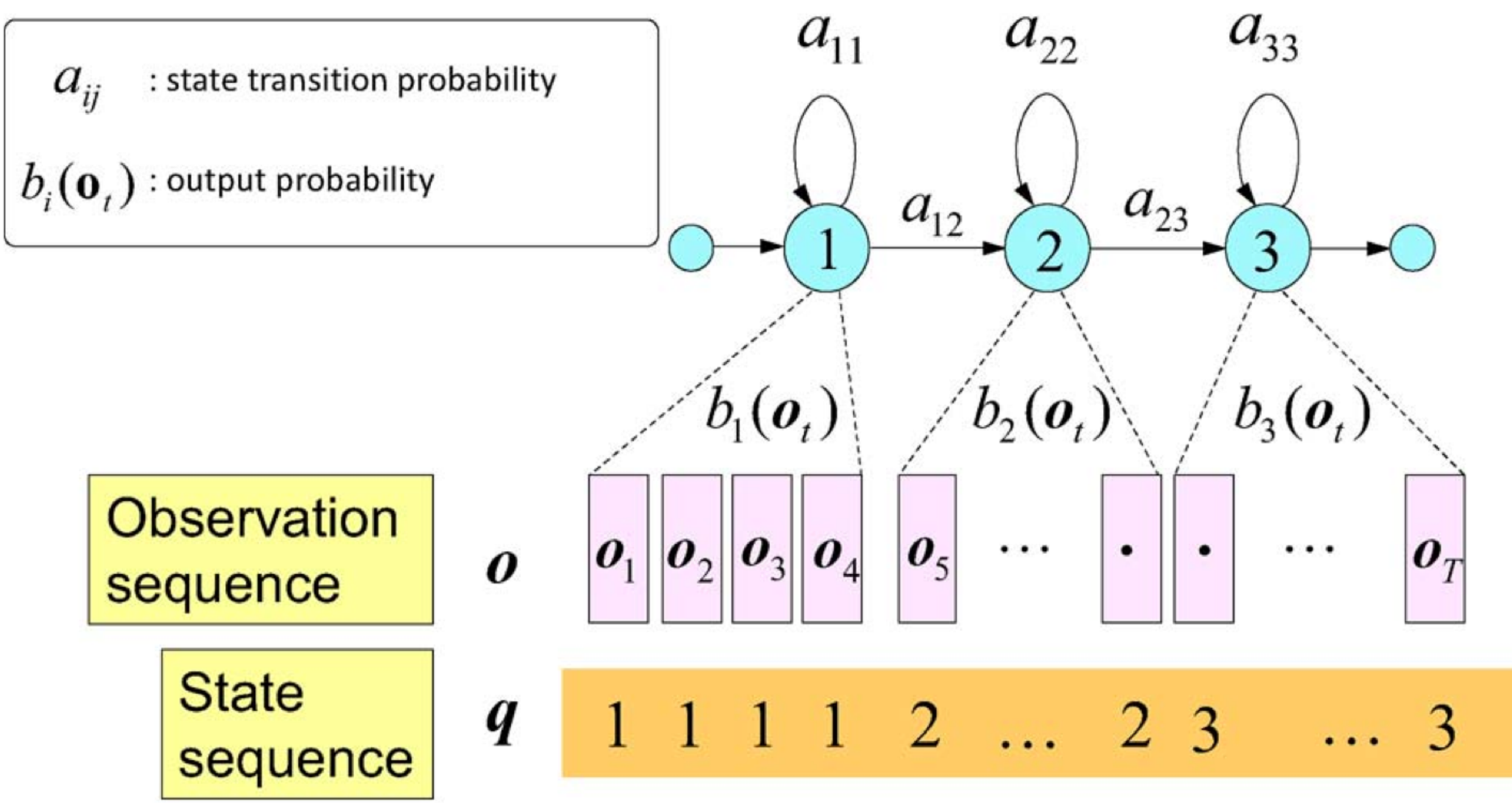
Parametric Synthesis

- Developed by Tokuda and Zen
- Proposed in mid-'90s, popular since 2007ish
- Big idea: Use classifiers/regressors to predict all of F0, duration, *spectral envelope*. Synthesize everything
- Initial work uses the same HMM we used for ASR, but in reverse

Parametric Synthesis

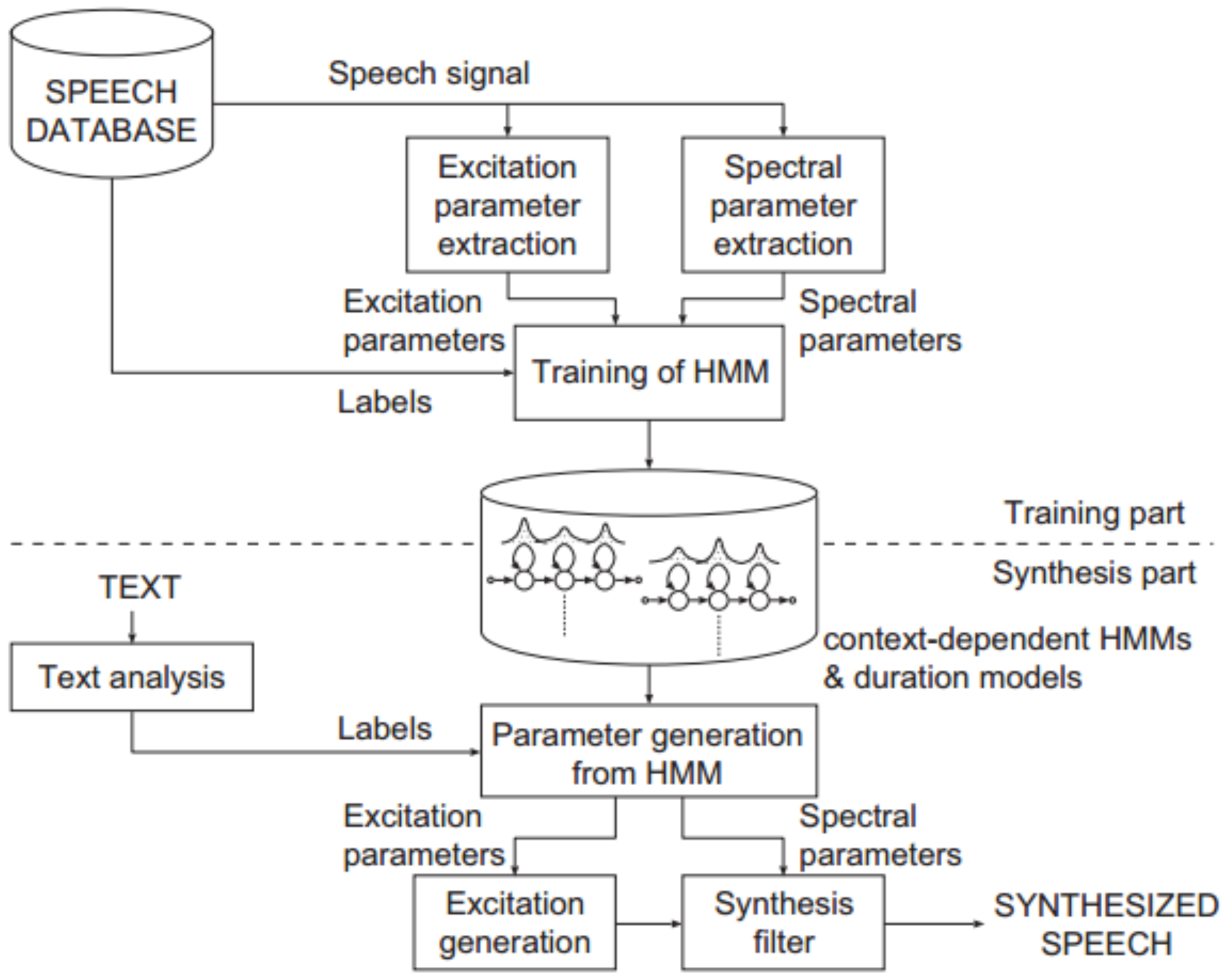
+ Small footprint
+ Don’t need huge amount of data to train
+ Flexible: easier to modify pitch for emotional change, or use MLLR adaptation to change voice characteristics
+ Smooth: no discontinuities in spectrum and prosody due to join artifacts
- Too smooth: flat, monotone, spectral smearing in time
- Vocoding effects: buzzy unnatural sound

HMM synthesis



Tokuda et al 2009

HTS system overview



(Tokuda, Zen, & Black. 2009)

What does the HMM produce?

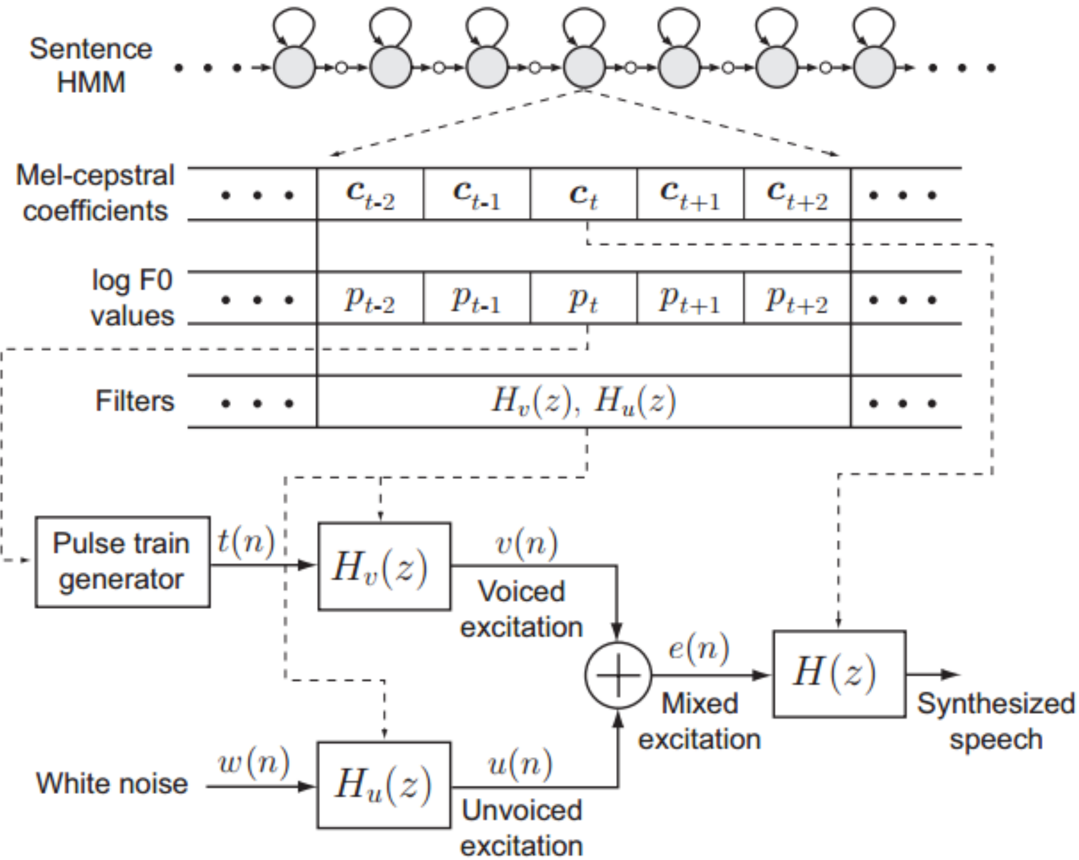
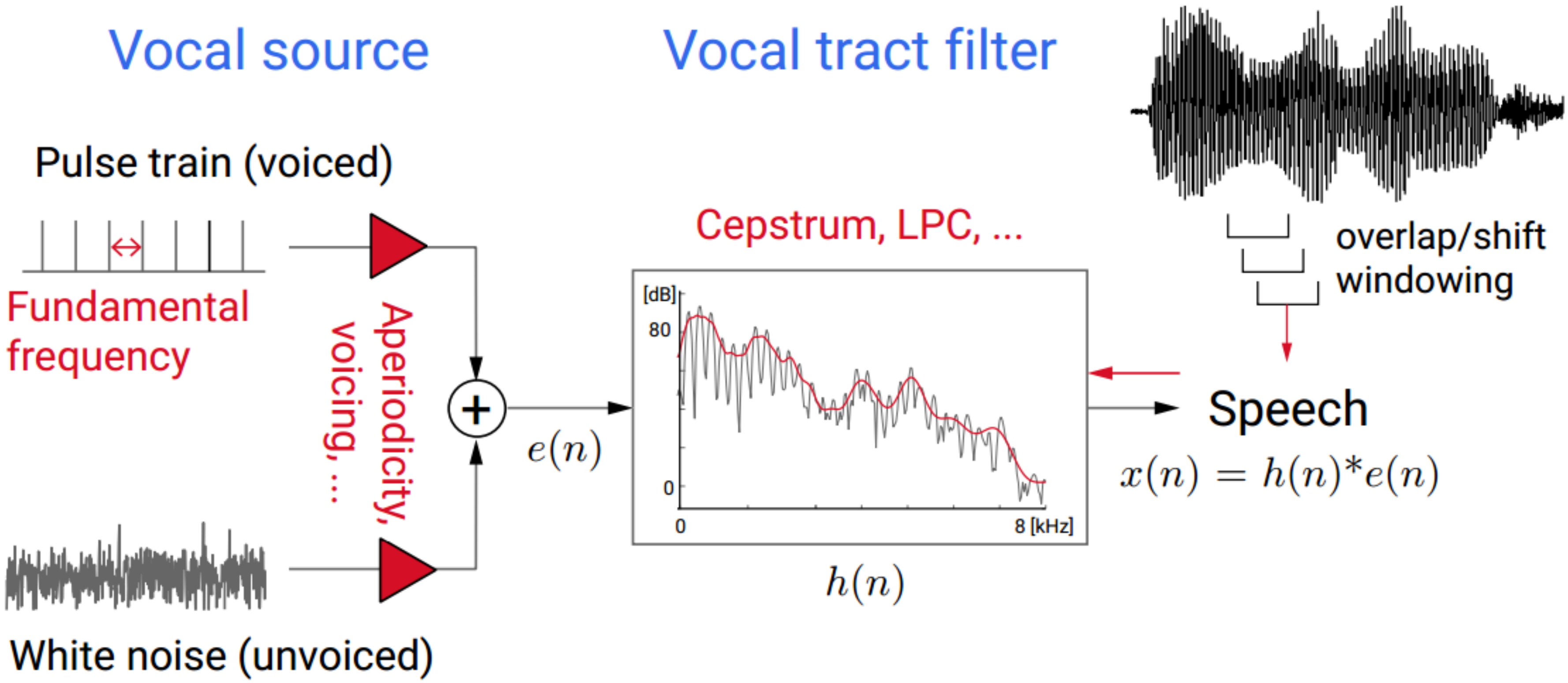


Figure 10: ML-based excitation scheme proposed by Maia et al. for HMM-based speech synthesis: filters $H_v(z)$ and $H_u(z)$ are associated with each state.

(Tokuda, Zen, & Black. 2009)

Synthesis with source-filter model

Piece-wise stationary, source-filter generative model $p(x \mid o)$



Slide from Heiga Zen

Key Questions in Parametric Synthesis

+ *What parameters do we predict?* Usually MFCCs for spectrum, log F0, voicing/excitation
+ *How do we combine them?* Exact parameterization and combining them well reduces robotic buzzy effects
+ *How do we make predictions?* Choice of HMM, machine learning approaches. Less important than the vocoding/combination issues

HTS Example

- Listen to the “low level” buzzy quality characteristic of most parametric systems
- Listen to clarity/impact of plosives compared to concatenative example

Comparing vocoder/excitation models

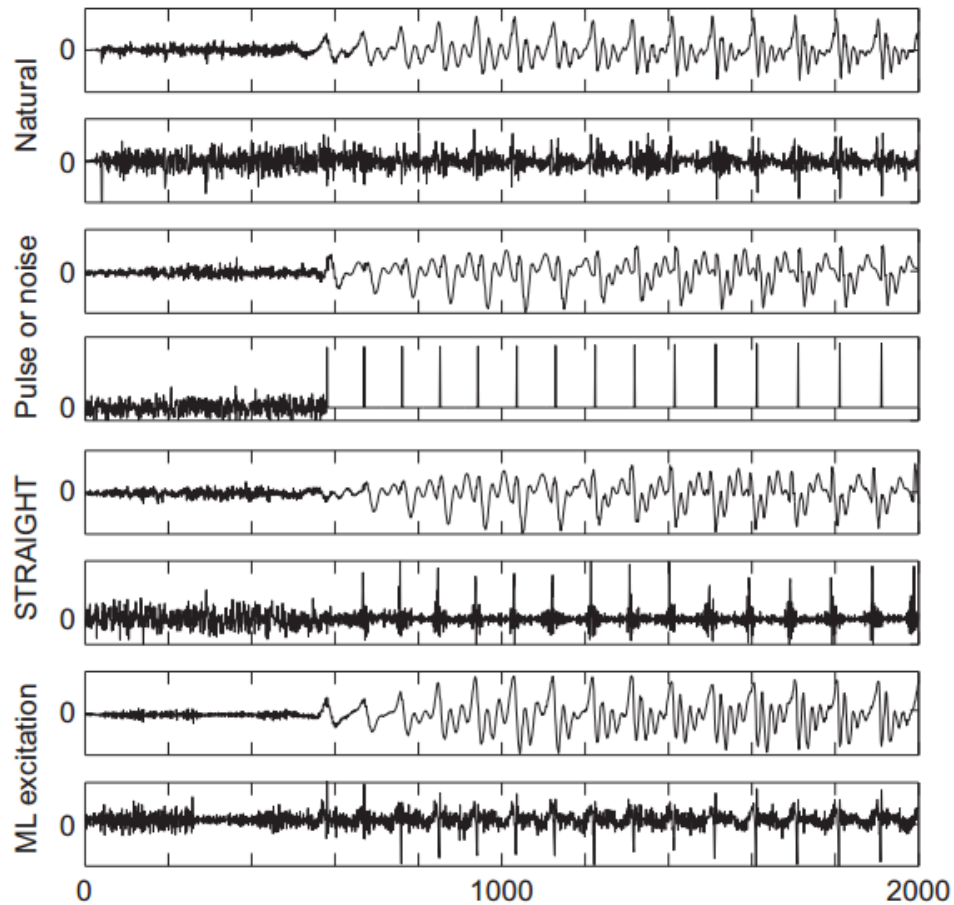


Figure 11: Waveforms from top to bottom: natural speech and its residual, speech and excitation synthesized with simple periodic pulse-train or white-noise excitation, speech and excitation synthesized with STRAIGHT vocoding method, and speech and excitation synthesized with ML excitation method.

(Tokuda, Zen, & Black. 2009)

End to end neural net synthesis

- TTS as a language model of individual samples

$$p(\mathbf{x}) = \prod_{t=1}^{T} p(x_t \mid x_1, \ldots, x_{t-1})$$

- Condition on phoneme sequence / prosodic features **h**

$$p(\mathbf{x} \mid \mathbf{h}) = \prod_{t=1}^{T} p(x_t \mid x_1, \ldots, x_{t-1}, \mathbf{h})$$

(Van den Oord *et al*. 2016)

Wavenet end to end synthesis

- TTS as a language model of individual samples

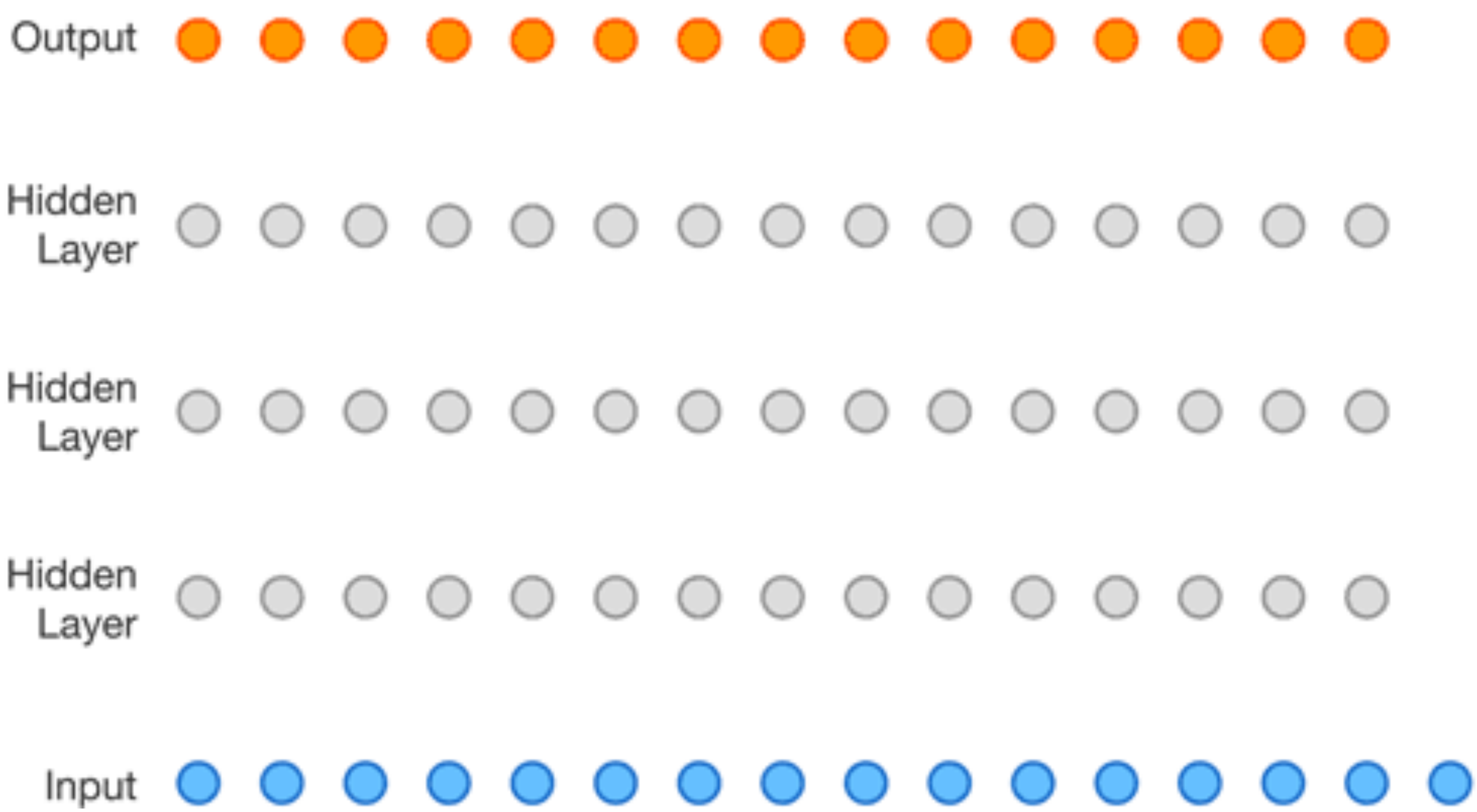
$$p(\mathbf{x}) = \prod_{t=1}^{T} p(x_t \mid x_1, \ldots, x_{t-1})$$

- Condition on phoneme sequence / prosodic features **h**

$$p(\mathbf{x} \mid \mathbf{h}) = \prod_{t=1}^{T} p(x_t \mid x_1, \ldots, x_{t-1}, \mathbf{h})$$

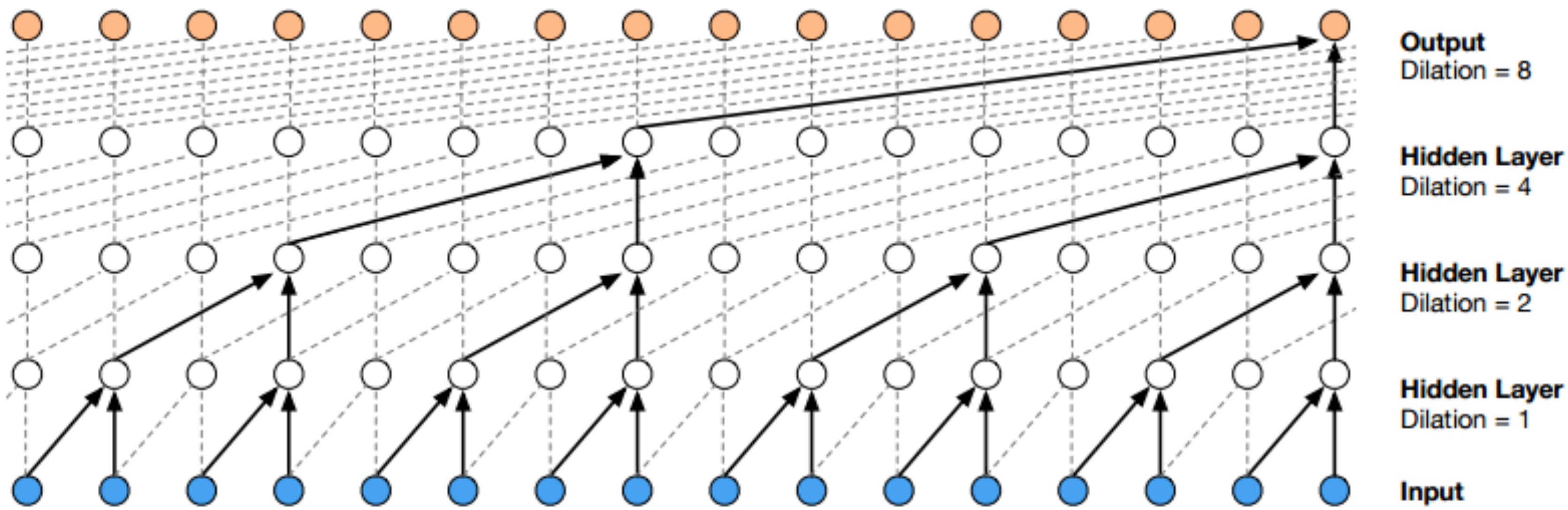
(Van den Oord *et al*. 2016)

Causal convolution architecture



(Van den Oord *et al*. 2016)

Dilated causal convolutions



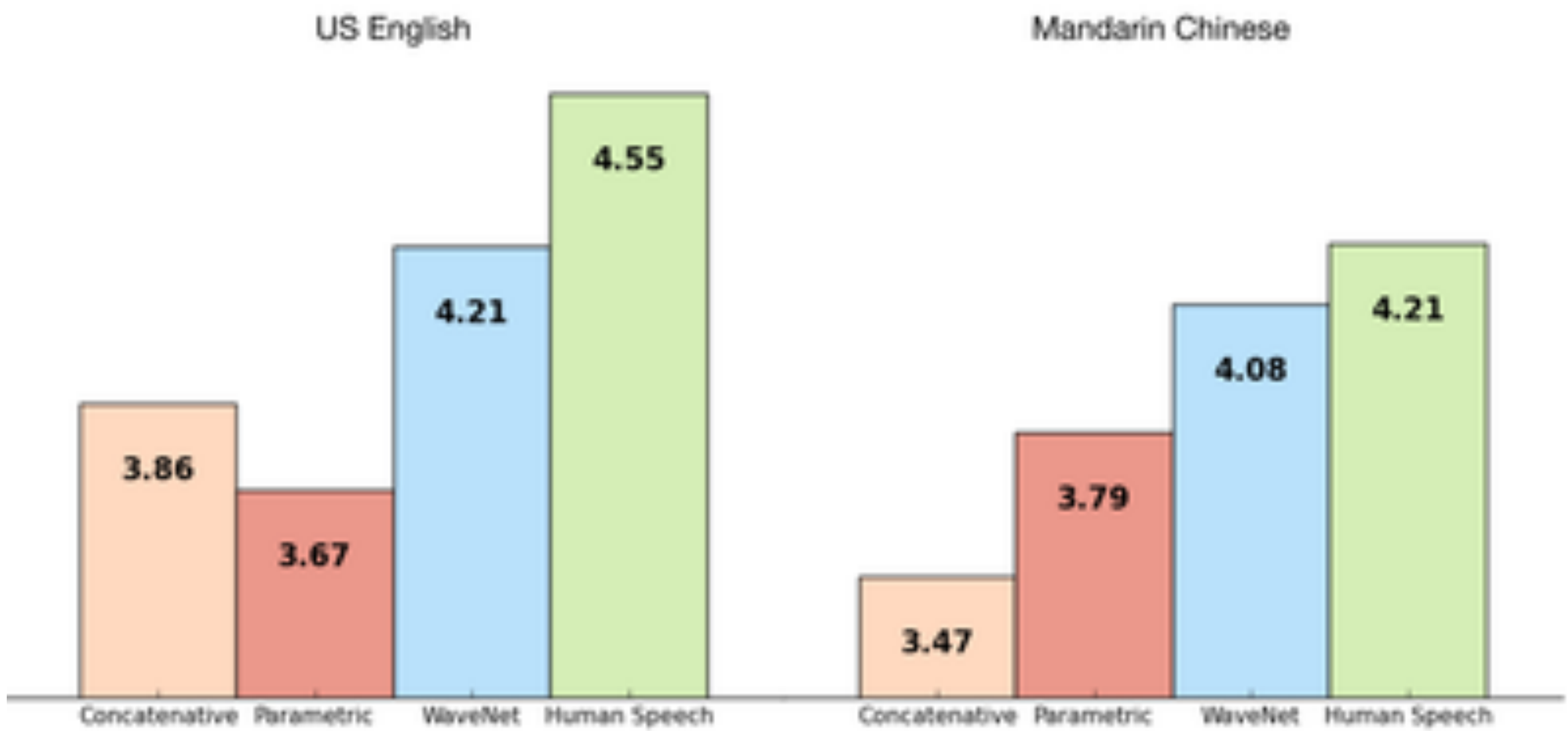
(Van den Oord *et al*. 2016)

Output encoding

- Real values for samples don't work well
- Many quantizations already exist for speech (from telephony mostly)
- Output is a *softmax classifier* over 256 quantized values (mu law)

(Van den Oord *et al*. 2016)

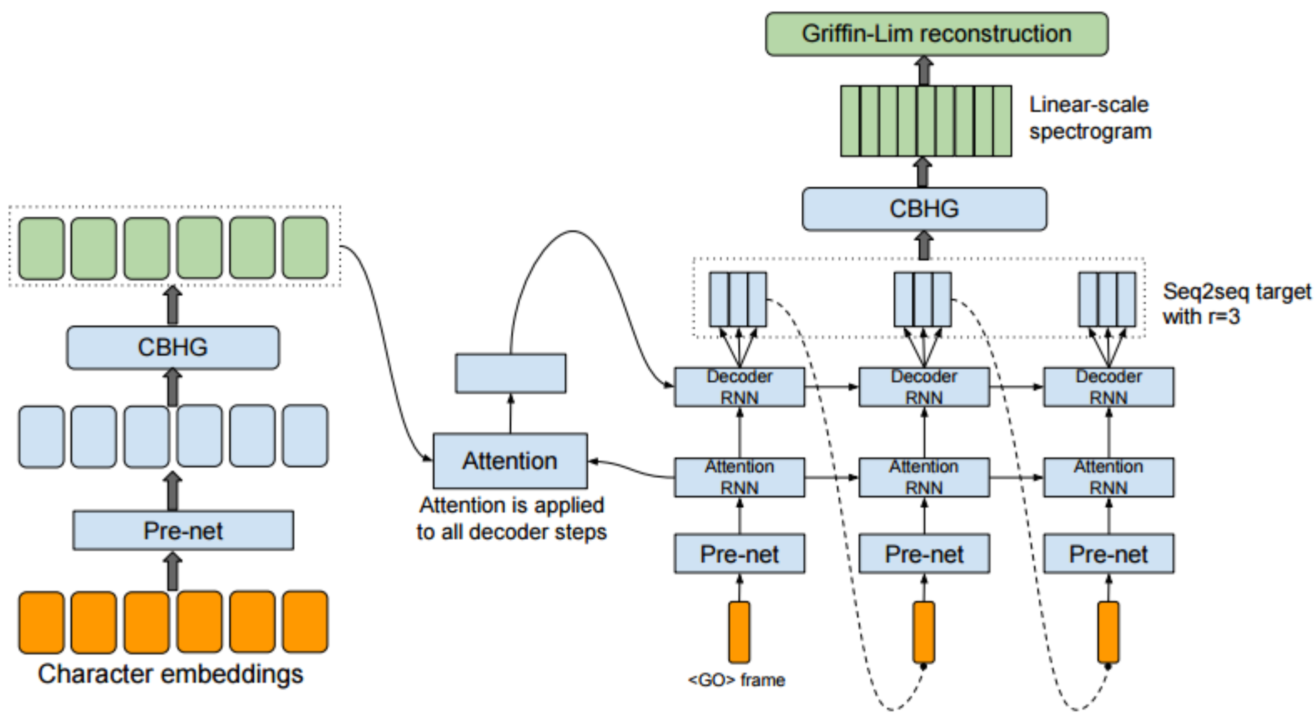
Mean opinion score results



(Van den Oord *et al*. 2016)

Sequence to sequence with attention

- Predict frames and use a more standard vocoder
- Input is character sequences rather than phonetic features



(Wang *et al*. 2017)

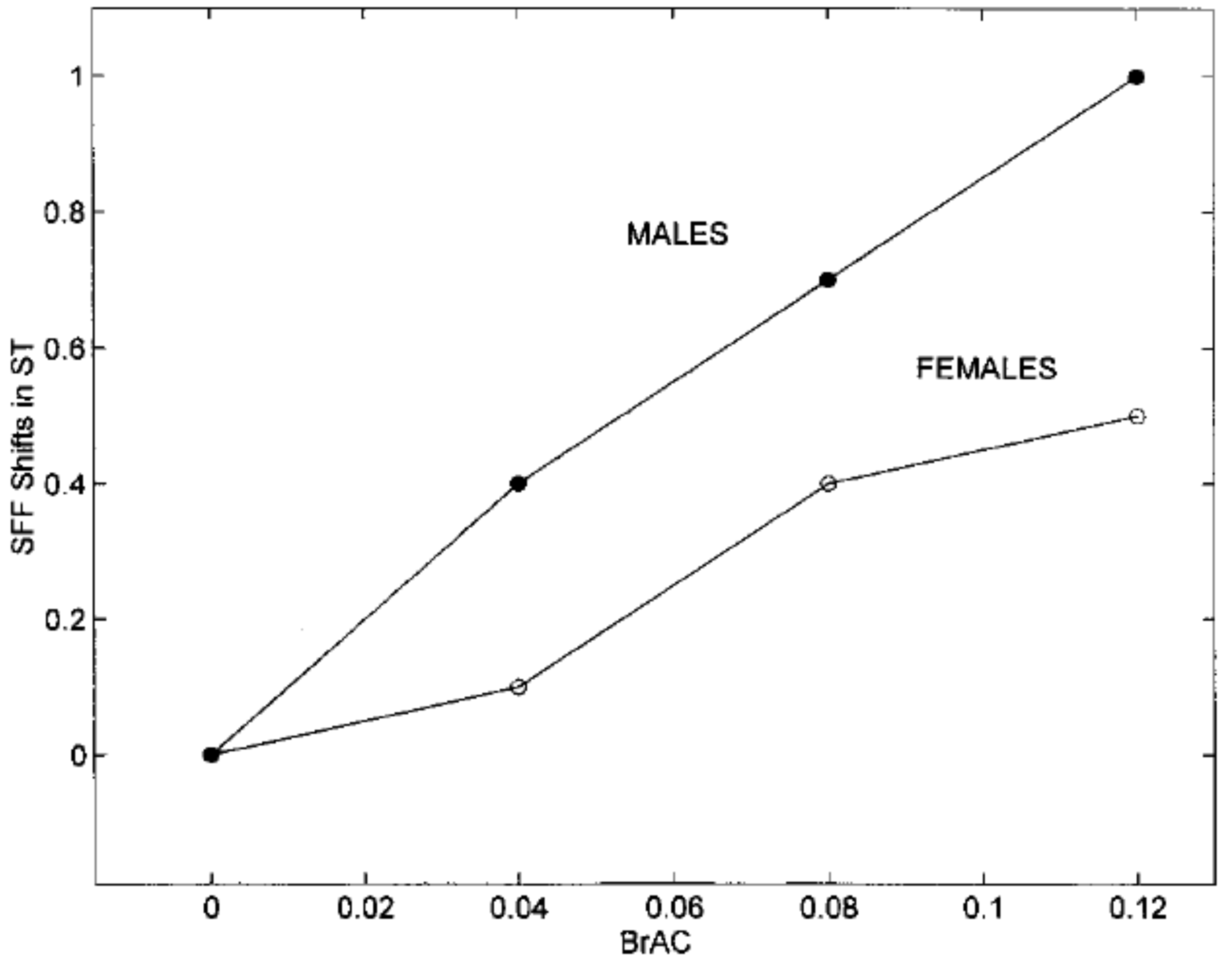
Intoxication

Hollien et al 2001

- Methods:
  - 35 young adults, 19 males, 16 females
  - given series of doses of alcohol
  - speech collected at 4 BAC stages

    - Rainbow passage
    - difficult words (buttercup, shapupie)
    - extemp speech ("Tell us about your favorite TV program)
    - head-mounted mikes
  - Investigated:
    - F0 mean and variance
    - duration/rate of speech
    - intensity
    - disfluencies

Hollien et al 2001 Results: F0



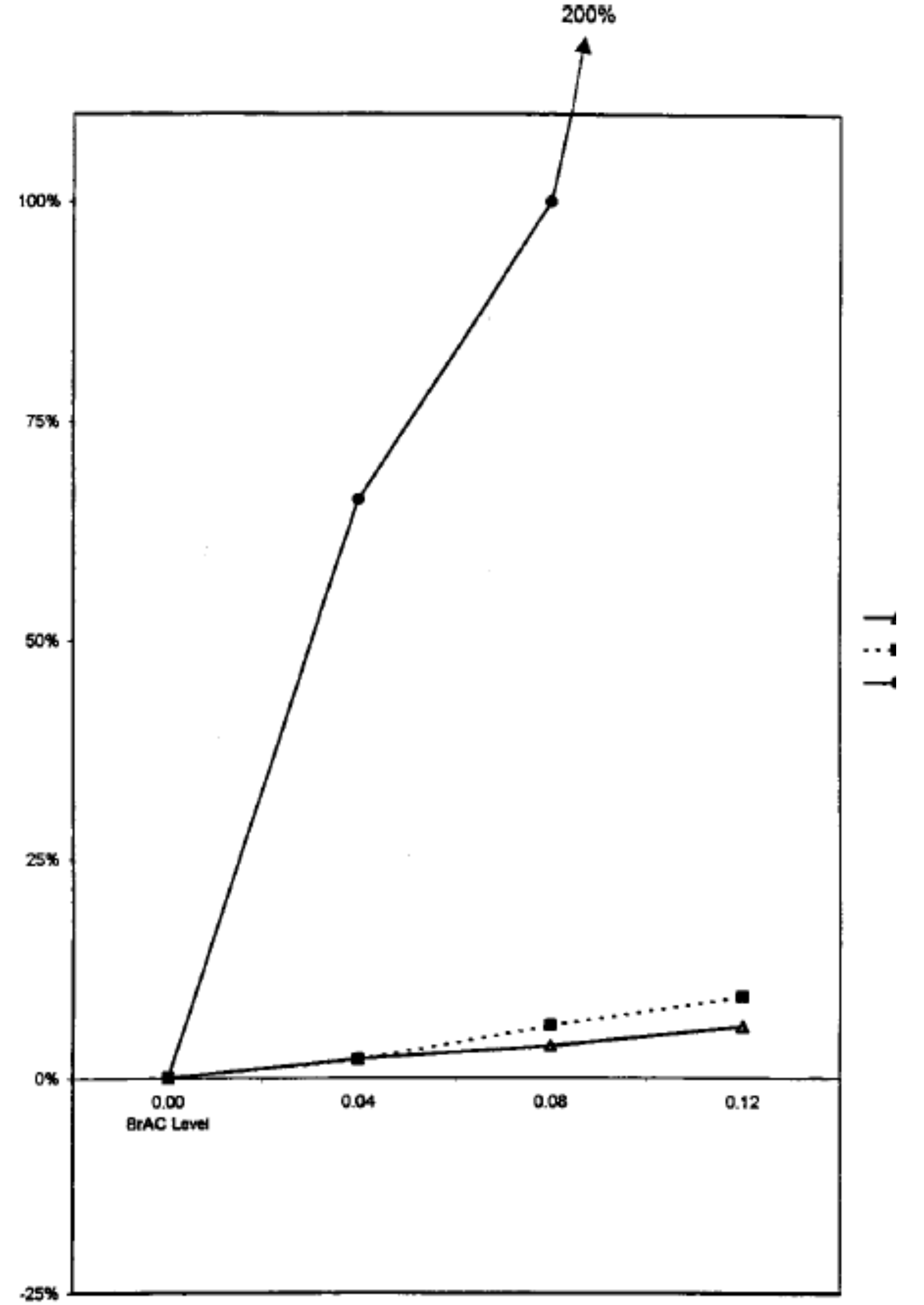
Hollien et al 2001 Results: Duration

Group	Level of intoxication (BrAC)				Shift
	0.00	0.04	0.08	0.12	(0.00–0.12)
Men					
Mean (s)	25.3	25.8	26.8	27.6	+2.3
S.D. (s)	2.9	2.5	2.1	2.5	
Women					
Mean (s)	25.1	25.5	25.7	27.5	+2.4
S.D. (s)	2.2	2.2	2.4	2.7	

Hollien et al 2001 Results: Disfluencies

		Experimental condition (BrAC)			
Subjects	*N*	0.00	0.04	0.08	0.12
Males	19				
Mean		3.2	4.7	6.5	8.6
SD		2.0	2.6	3.1	3.4
Females	16				
Mean		2.2	3.5	4.7	6.1
SD		1.7	2.2	2.7	3.0
Mean	35	2.7	4.1	5.6	7.4

Hollien et al 2001 Results: Magnitudes



Hollien et al 2001 Results: Speaker Specific Effects

- 20% of speakers did not follow these trends

A famous case study

- Johnson, K., Pisoni, D. & Bernacki, R. (1990) Do voice recordings reveal whether a person is intoxicated?: A case study. Phonetica. 47: 215-237.

Exxon Valdez

Exxon Valdez oil spill

From Wikipedia, the free encyclopedia Coordinates: 60.83333°N 146.86667°W

The ***Exxon Valdez* oil spill** occurred in Prince William Sound, Alaska, on March 24, 1989, when the *Exxon Valdez*, an oil tanker bound for Long Beach, California, struck Prince William Sound's Bligh Reef and spilled 260,000 to 750,000 barrels (41,000 to 119,000 m^3) of crude oil.[1][2] It is considered to be one of the most devastating human-caused environmental disasters.[3] As

***Exxon Valdez* oil spill**

3 days after *Exxon Valdez* ran aground

Location	Prince William Sound, Alaska
Coordinates	60.83333°N 146.86667°W
Date	24 March 1989
Cause	

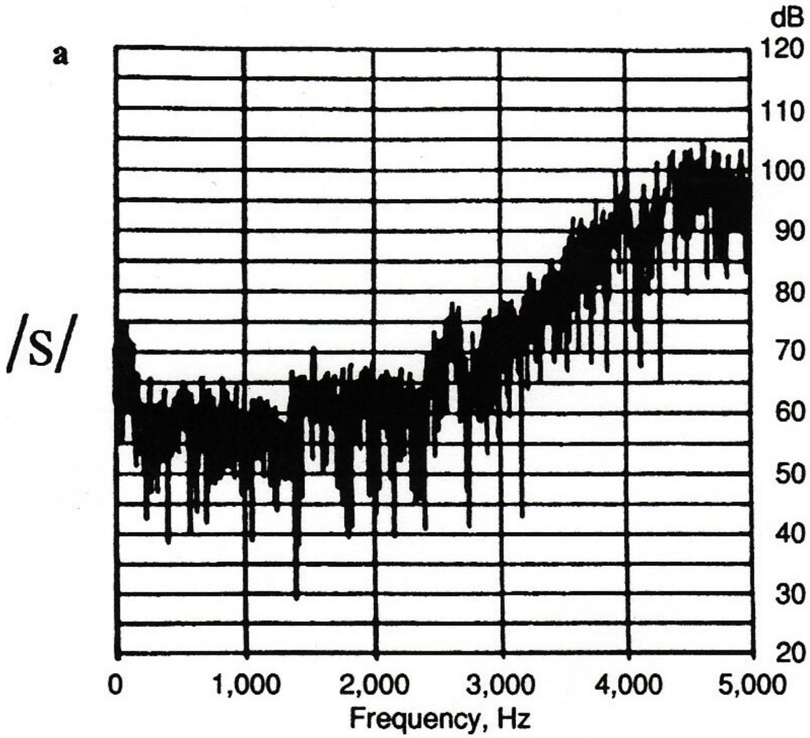
Was Captain Hazelwood drunk?

- Not clear if this is relevant, since seems like other questionable corporate things were going on:
  - he was asleep below deck
  - The third mate was in charge of the wheelhouse
  - the ship's radar was broken
- But is a well-studied case

Johnson et al examined 3 kinds of cues

- Segmental Effects (phoneme, syllable, word level)
- Disfluencies
- Suprasegmental Effects (stress, intonation, etc.)

Keith Johnsons /s/ and/ʃ/



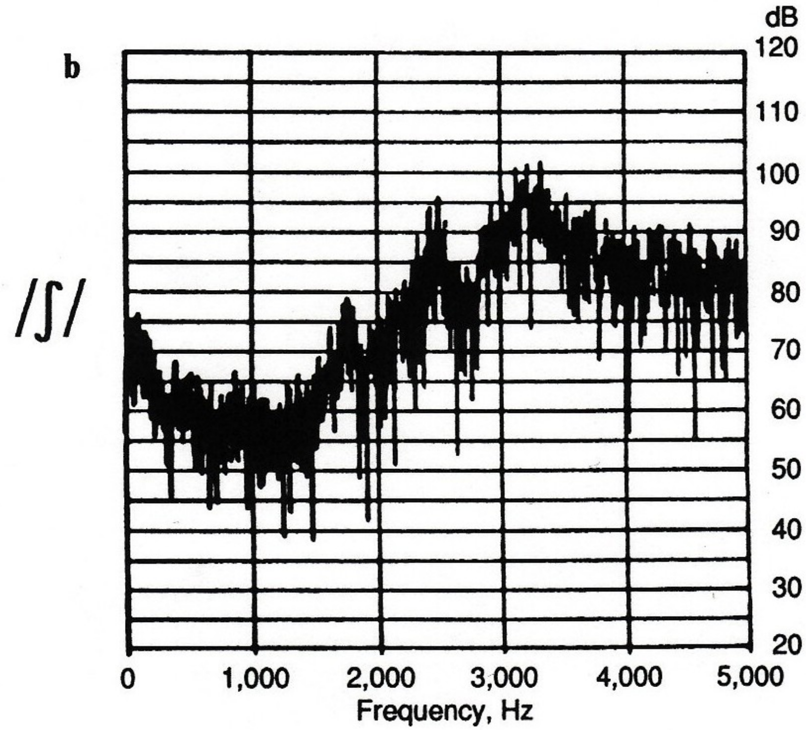


Fig. 1. Power spectra of /s/ (**a**) and /ʃ/ (**b**) produced by K. J. in a quiet recording booth with recording equipment responsive up to 5,000 Hz.

e.g. “sun” vs “shun”

/ʃ/: Captain Hazelwood

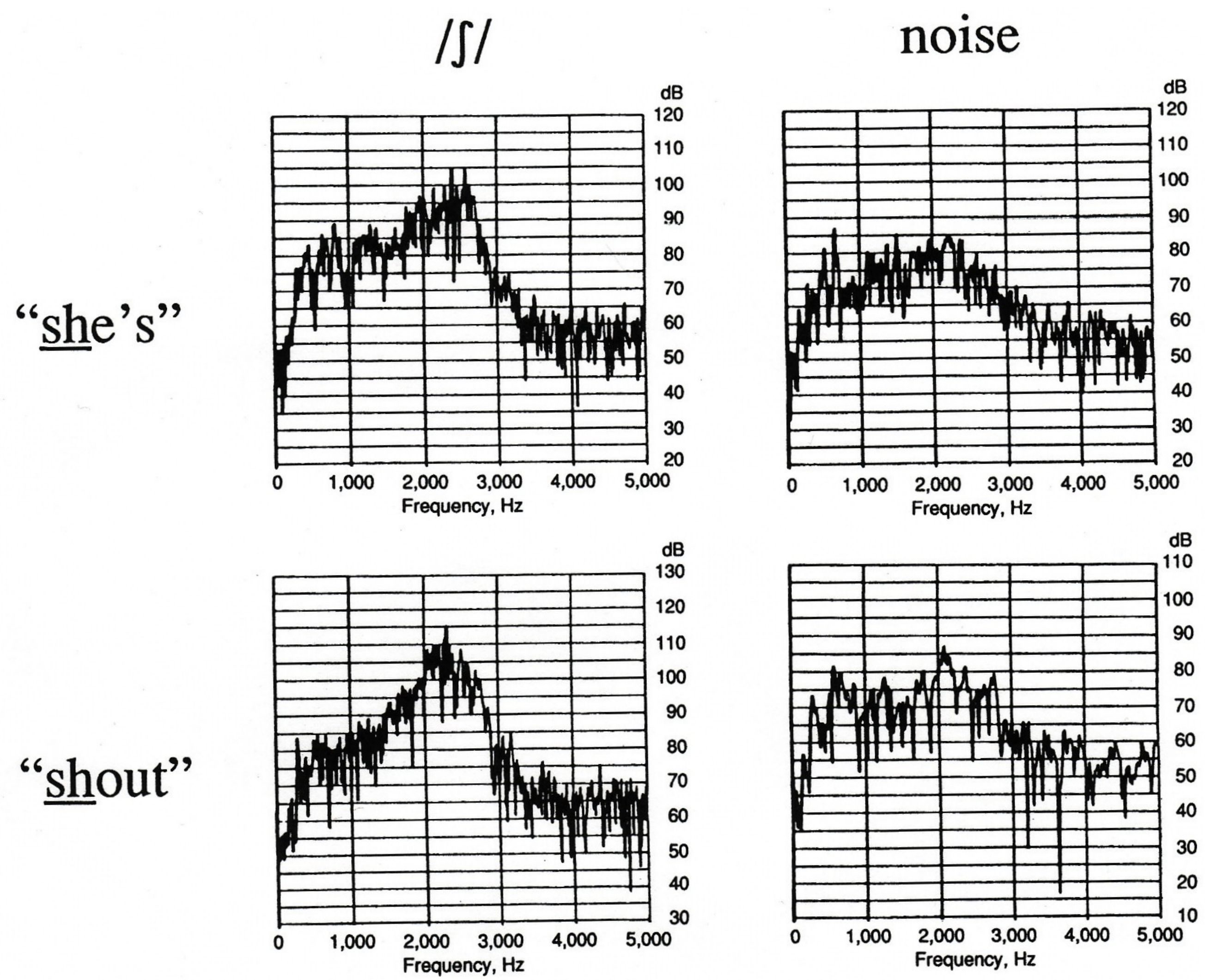
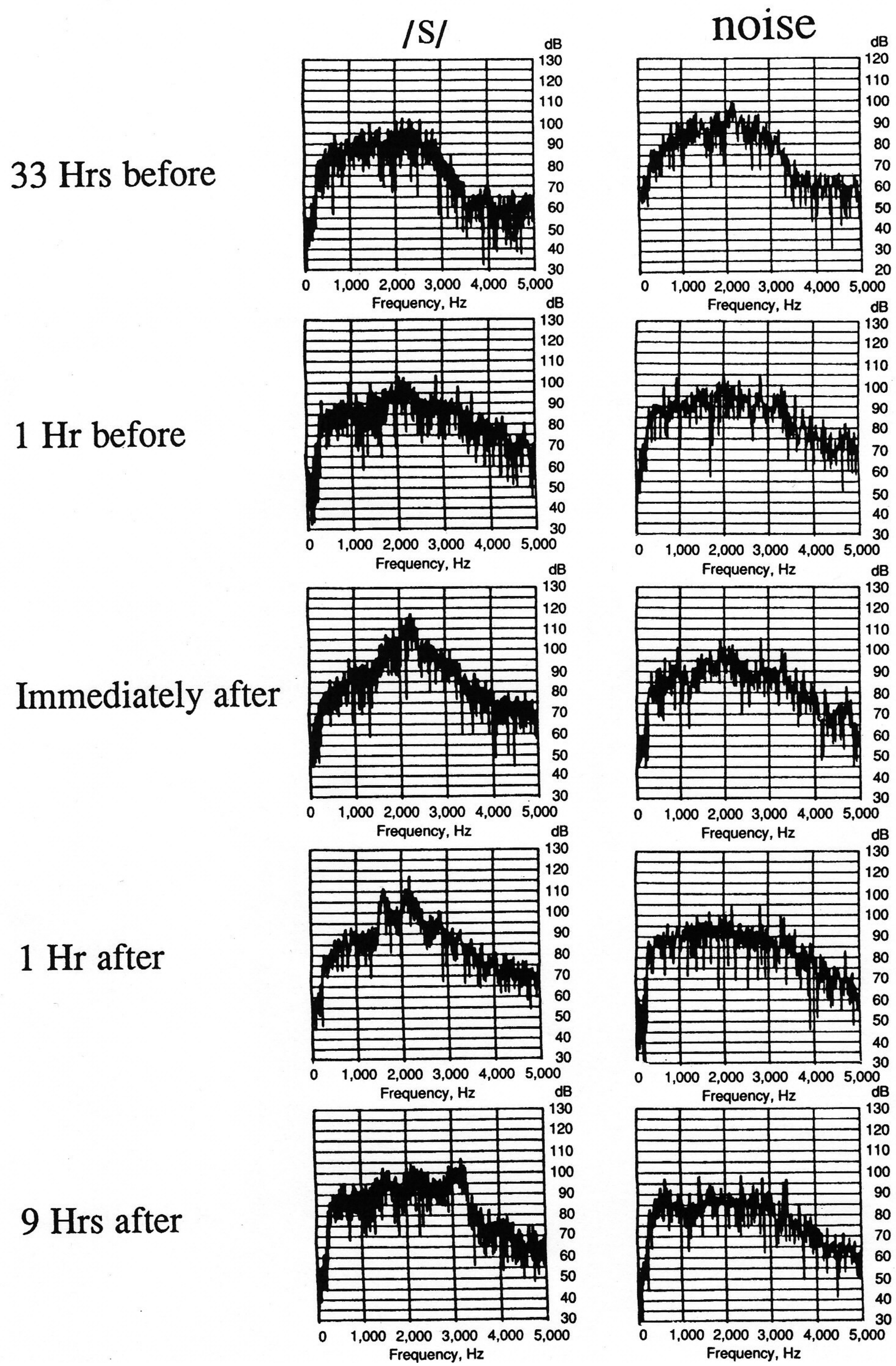


Fig. 2. Power spectra of /ʃ/ produced by Captain Hazelwood in the words *she's* and *shout* recorded 33 h before the accident. Each spectrum is paired with a spectrum of the background noise from a nearby open-mike pause.


/s/
noise
33 Hrs before
1 Hr before
Immediately after
1 Hr after
9 Hrs after
dB
Frequency, Hz
0
1,000
2,000
3,000
4,000
5,000

/ʃ/ noise

“she’s”

“shout”

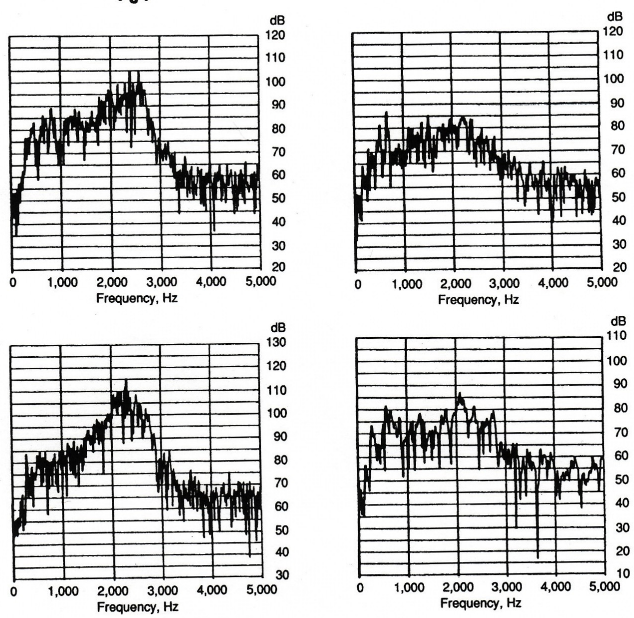
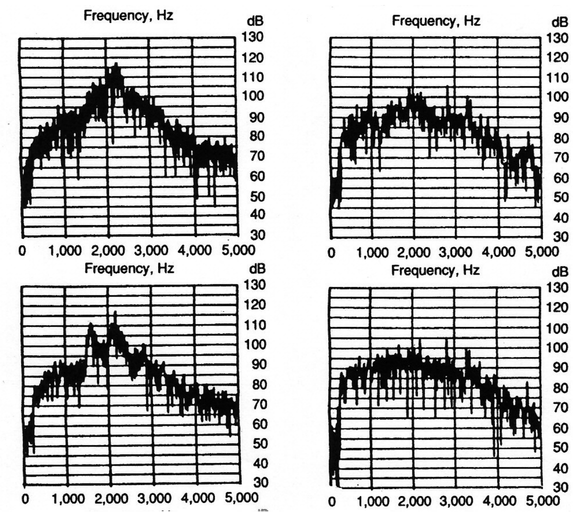


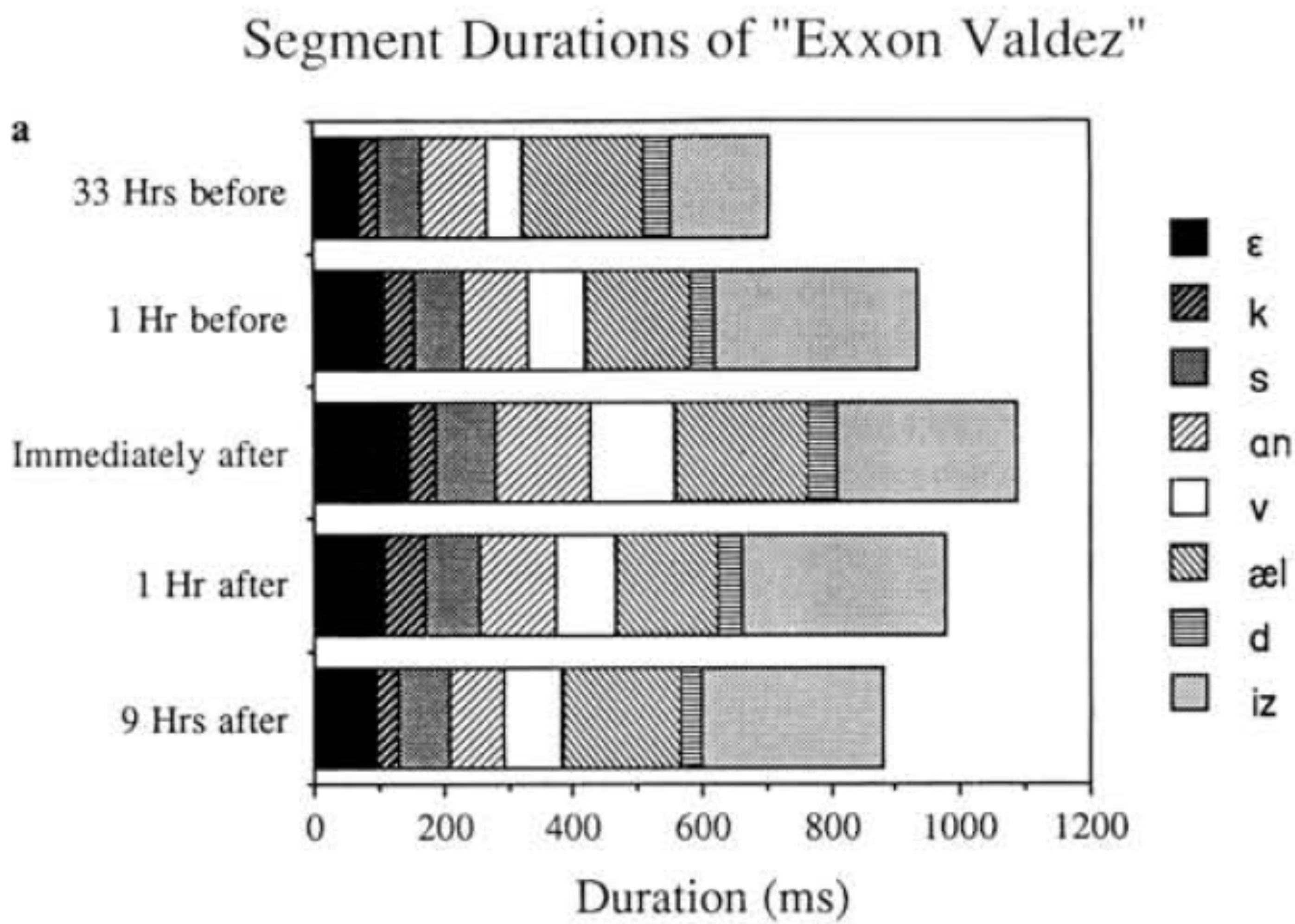
Fig. 2. Power spectra of /ʃ/ produced by Captain Hazelwood in the words *she’s* and *shout* recorded 33 h before the accident. Each spectrum is paired with a spectrum of the background noise from a nearby open-mike pause.

Immediately after

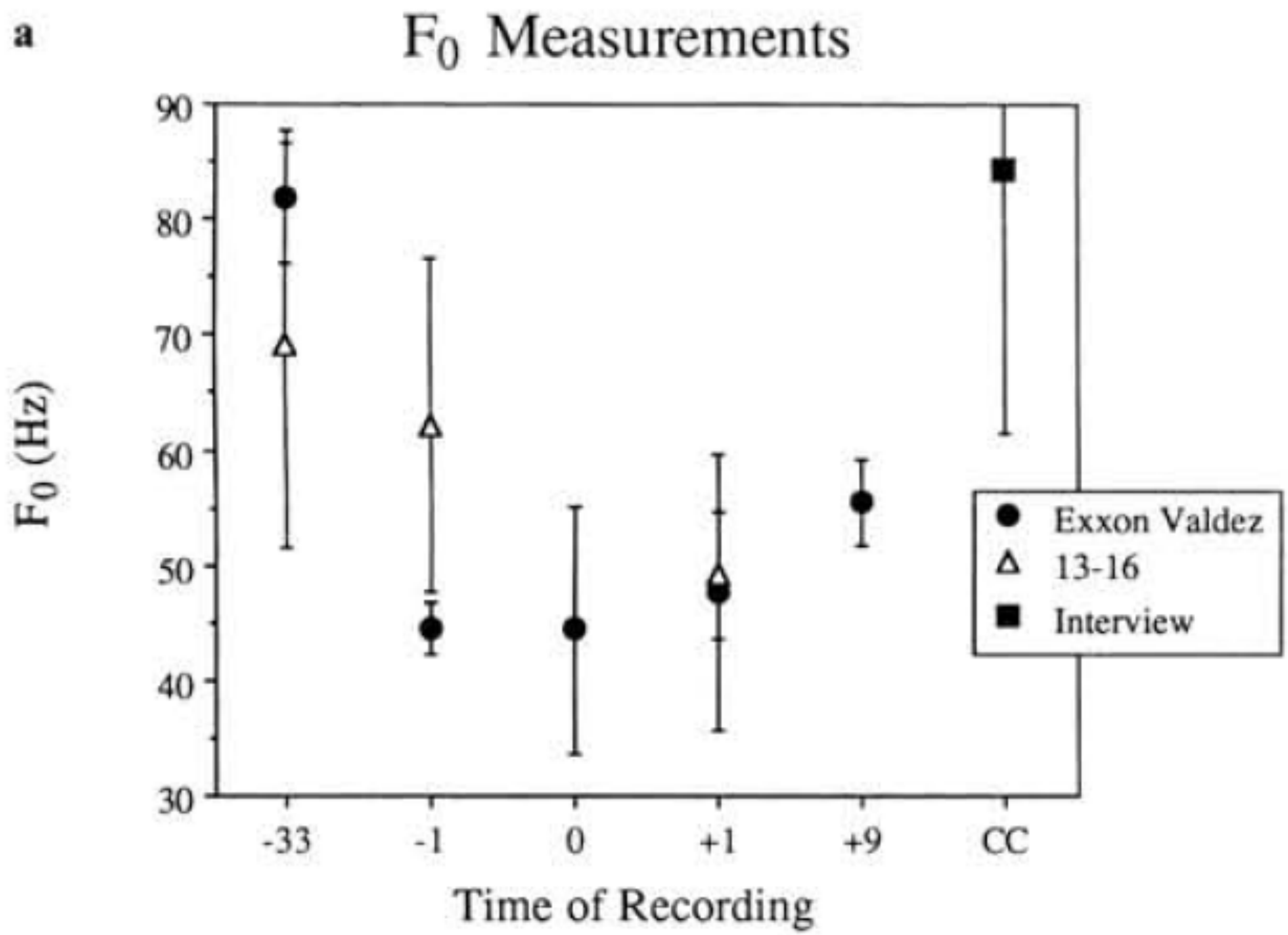
1 Hr after



Duration


Segment Durations of "Exxon Valdez"
a
33 Hrs before
1 Hr before
Immediately after
1 Hr after
9 Hrs after
0
200
400
600
800
1000
1200
Duration (ms)
ɛ
k
s
ɑn
v
æl
d
iz

F0


a
F_0 Measurements
90
80
70
60
50
40
30
F_0 (Hz)
-33
-1
0
+1
+9
CC
Time of Recording
Exxon Valdez
13-16
Interview

Summary

Table 3. Summary of phenomena found in the analysis of the NTSB tape (numbers in parentheses indicate the time of recording)

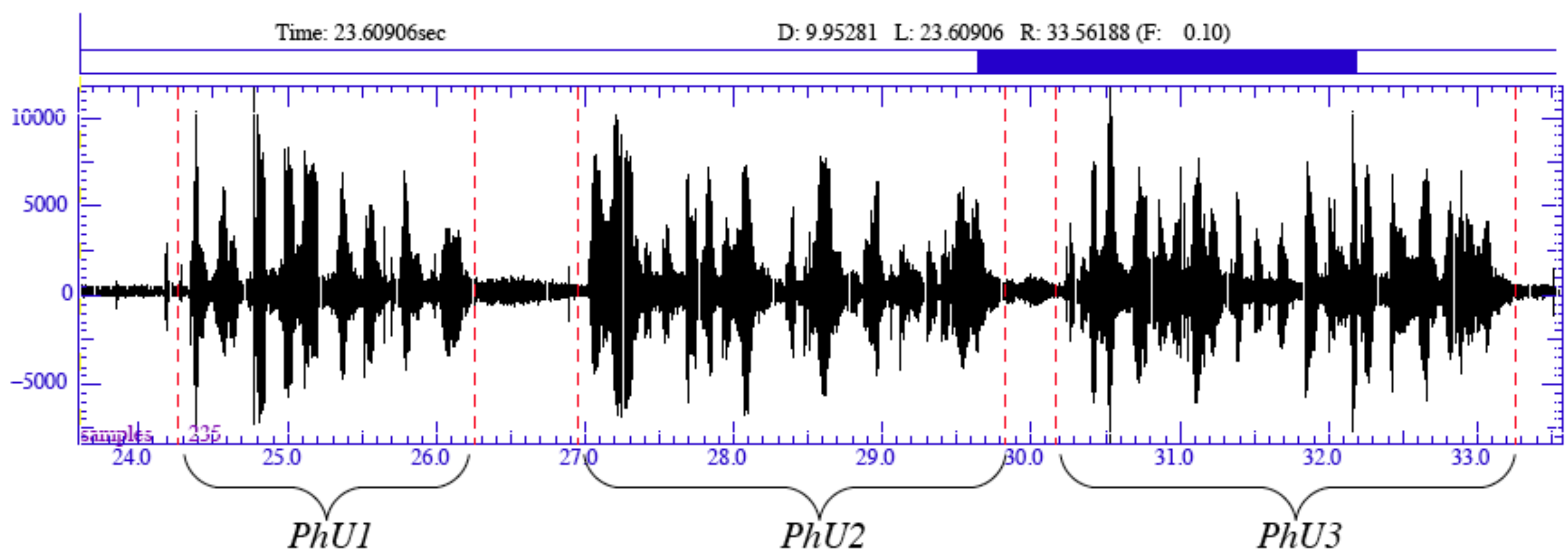
Gross effects	revisions (–1) Exxon Ba, uh Exxon Valdez (–1) departed disembarked (–1) I, we'll (–1) columbia gla, columbia bay
Segmental effects	misarticulation of /r/ and /l/ (0) northerly, little, drizzle, visibility (/s/ becomes /ʃ/ (fig. 3) final devoicing (e.g. /z/ → /s/) (–1,0,+1) Valdez → Valdes
Suprasegmental effects	reduced speaking rate (fig. 4, 5) mean change in pitch range (talker-dependent, fig. 6) increased F_0 jitter (fig. 6)

Problems

- If intoxicated speech, why wasn't s pronounced as sh 1 hour before?
- Other kinds of speaker state could cause drop in F0, slower speech, and disfluencies?
  - Stress, just having woken up, trauma….

Automatic Classification

- Use of prosodic speech characteristics for automated detection of alcohol intoxication
  Michael Levit, Richard Huber, Anton Batliner, Elmar Noeth
- Break utterance into phrases automatically, based on
  - fundamental frequency (where possible);
  - zero-crossing rate



Then use 4 classes of features

- Prosodic
  - F0 max, F0 min, energy max, energy min, pause length
- Duration of voiced regions, unvoiced regions, etc.
- Jitter and shimmer
  - jitter is variation in pitch
  - shimmer is variation in energy
- Average cepstrum and cepstral slope

Methods

- Alcoholized speech samples collected at the Police Academy of Hessen, Germany
  - 120 readings (87 minutes) of a fable
  - 33 male speakers
  - BAC between 0 and .24/mille

Alcohol Blood Level	0.0	< 0.4	< 0.8	< 1.2	< 1.6	< 2.0	< 2.4
Recordings	32	20	20	18	20	7	3

  - Binary task: above or below 0.8/mille
  - leave-one-out cross-validation
  - neural net classifier

Results of Levit et al.

- Used dev set to find best classifier
- This suggested two feature classes:
  - Prosodic features
  - Jitter/shimmer
- Results with this classifier
  - 62% phrase-accuracy
  - 69% for the whole speech sample
    - voting of the phrases

New Corpus!

- Alcohol Language Corpus
  - Florian Schiel et al 2009, 2010
- http://www.bas.uni-muenchen.de/forschung/Bas/BasALCeng.html
- 162 speakers (77 female, 85 male)
  - recorded in a car (sometimes with engine running)
  - command and control speech ("turn off the radio")
  - spontaneous dialogue, monologue, question answering
  - read speech
  - counts of disfluencies, etc
- sample, drunk: 
- sample, sober: 

Automatic detection in ALC : Paralinguistic Challenge 2011

- Human: 66-72% (Schiel 2011, Ultes, Schmitt, Minker 2011)
- Machine: roughly 65%-70%
- Example features from winning system:

Bone, Daniel, Matthew Black, Ming Li, Angeliki Metallinou, Sungbok Lee, and Shrikanth S. Narayanan. 2011. Intoxicated Speech Detection by Fusion of Speaker Normalized Hierarchical Features and GMM Supervectors. In *INTERSPEECH*, pp. 3217-3220.

- Prosody (f0, duration, energy, jitter, shimmer)
- Spectral (MFCC, MFB log-energy, formants)
- Computed over whole utterance and small windows
- normalized phoneme duration
- iterative speaker normalizaton

Depression

Stirman and Pennebaker

- Suicidal poets
- 300 poems from early, middle, late periods of
  - 9 suicidal poets
  - 9 non-suicidal poets

Stirman and Pennebaker: 2 models

- Durkheim disengagement model:
  - suicidal individual has failed to integrate into society sufficiently, is detached from social life
  - detach from the source of their pain, withdraw from social relationships, become more self-oriented
  - prediction:
    - more self-reference, less group references
- Hopelessness model:
  - Suicide takes place during extended periods of sadness and desperation, pervasive feelings of helplessness, thoughts of death
  - prediction:
    - more negative emotion, fewer positive, more refs to death

Methods

- 156 poems from 9 poets who
  - committed suicide
  - published, well-known
  - in English
  - have written within 1 year of commmiting suicide
- Control poets matched for nationality, education, sex, era.

The poets

TABLE 2. Suicidal Poets and Their Controls

Suicidal Poet	Age at Death	Control Poet	Cutoff Age	Nationality	Other Similarities
Randall Jarrell (1914–1965)	51	Robert Lowell (1917–1977)	50 (1967)	American	PhD
John Berryman (1914–1972)	58	Lawrence Ferlinghetti (1919–)	59 (1978)	American	PhD
Sylvia Plath (1932–1963)	31	Denise Levertov (1923–1997)	37 (1960)	American	Lived England and US
Anne Sexton (1928–1974)	46	Adrienne Rich (1929–)	46 (1975)	American	Lived in New England, college education
Adam L. Gordon (1833–1870)	37	Matthew Arnold (1822–1888)	45 (1867)	British	
Sarah Teasdale (1884–1933)	49	Edna St. V. Millay (1892–1950)	49 (1941)	American	
Hart Crane (1899–1932)	33	Joyce Kilmer (1886–1918)	32 (1918)	American	
Sergei Esenin (1895–1925)	30	Boris Pasternak (1890–1960)	35 (1930)	Russian	
Vladimir Mayakovsky (1893–1930)	37	Osip Mandelstam (1891–1938)	37 (1928)	Russian	

Stirman and Pennebaker: Results

TABLE 1. Means for LIWC Categories

	Suicide Group			Control Group			Effects
	Early	Middle	Later	Early	Middle	Late	
Disengagement theory							
I (me, my)	4.0	3.4	4.0	2.5	1.6	2.5	S
We (us, our)	.73	1.3	.85	.69	.40	1.1	S,P**
Communication (talk, share)	1.2	1.1	1.0	.89	1.1	1.3	—
Hopelessness theory							
Negative emotion (hate, worthless)	2.2	1.8	1.7	2.3	2.1	1.7	—
Positive emotion (happy, love)	3.3	3.1	3.9	2.9	2.9	2.5	—
Death (dead, grave)	.52	.47	.69	.34	.43	.41	S**
Other findings							
Sexual words (lust, breast)	.60	.84	.47	.36	.36	.31	S

Note: Means reflect percentage of total words used in each poem within the relevant category. Effects refer to: S = suicide vs. nonsuicide main effect, P = phase of career main effect. All effects are significant $p \leq .05$, except ** $p \leq .08$.

Significant factors

- Disengagement theory
  - I, me, mine
  - we, our, ours
- Hopelessness theory
  - death, grave
- Other
  - sexual words (lust, breast)

Rude et al: Language use of depressed and depression-vulnerable college students

- Beck (1967) cognitive theory of depression
  - depression-prone individuals see the world and themselves in pervasively negative terms
- Pyszynski and Greenberg (1987)
  - think about themselves
  - after the loss of a central source of self-worth, unable to exit a self-regulatory cycle concerned with efforts to regain what was lost.
  - results in self-focus, self-blame
- Durkheim social integration/disengagement
  - perception of self as not integrated into society is key to suicidality and possibly depression

Methods

- College freshmen
  - 31 currently-depressed (standard inventories)
  - 26 formerly-depressed
  - 67 never-depressed
- Session 1: take depression inventory
- Session 2: write essay
  - please describe your deepest thoughts and feelings about being in college… write continuously off the top of your head. Don't worry about grammar or spelling. Just write continuously.

Results

- depressed used more “I,me” than never-depressed
  - turned out to be only “I”
- and used more negative emotional words
- not enough “we” to check Durkheim model
- formerly depressed participants used more “I” in the last third of the essay

Ramirez-Esparza et al: Depression in English and Spanish

- Study 1: Use LIWC counts on posts from 320 English and Spanish forums
  - 80 posts each from depression forums in English and Spanish
  - 80 control posts each from breast cancer forums
- Run the following LIWC categories
  - I
  - we
  - negative emotion
  - positive emotion

Results of Study 1

	English		Spanish			
Categories	Dep.	Breast Cancer	Dep.	Breast Cancer		
	N=80 Mean (SD)	N=80 Mean (SD)	N=80 Mean (SD)	N=80 Mean (SD)	t-value	t-value
First person singular	12.24 (2.97)	4.03 (3.01)	9.30 (2.34)	5.03 (2.76)	-17.39*	-10.54*
First person plural	.18 (.33)	.72 (1.06)	.22 (.39)	1.02 (1.28)	4.36*	5.32*
Positive Emotions	1.72 (1.14)	2.54 (1.72)	2.99 (1.36)	3.53 (1.93)	3.56*	2.04^{+}

Case Study: Online Forum Posts

- From depression forums:
  - 404 English posts
  - 404 Spanish posts
- Create a term by document matrix of content words
  - 200 most frequent content words
- Do a factor analysis
  - dimensionality reduction in term-document matrix
  - Used 5 factors

English Factors

FACTOR 1: Treatment		FACTOR 1: Disclosure		FACTOR 3: Family	
Medication	.62	Tell	.43	Mom	.49
Effect	.46	People	.41	Daughter	.48
Depression	.43	Know	.39	Child	.48
Side	.35	Happy	.35	Family	.48
Week	.34	Talk	.35	Brother	.43
Therapy	.34	Feel	.34	Sister	.43
Suffer	.34	Want	.34	Dad	.41
Disorder	.33	Suppose	.33	Son	.33
Doctor	.33	Read	.32	Love	.33
Antidepressant	.32	Hurt	.32	Girl	.33
Experience	.32	Wrong	.32	Young	.32
Major	.32	Emotional	.31	Parent	.32
Mental	.31	Mind	.31	House	.31
Psychiatrist	.31	Sad	.31	Husband	.30
		Make	.31	Crazy	.30

FACTOR 4: Symptoms		FACTOR 5 School	
Sleep	.51	Constant	.46
Hour	.48	Relationshi	.45
Food	.44	School	.40
Wake	.44	High	.40
Morning	.44	Lack	.39
Night	.41	University	.38
Bed	.39	Social	.38
Stay	.38	College	.36
Weight	.37	Move	.32
Eat	.36	Friend	.32
Place	.32	Girlfriend	.32
		Class	.31

Spanish Factors

FACTOR 1: Family		FACTOR 2: Relationship History	
MADRE/mother	.48	RELACION/relationship	.51
HERMANO/brother	.46	ENAMORADO/love	.48
ABUELA/grandmother	.44	CONOCI/met	.45
PADRES/parents	.44	HABLAR/talk	.45
PAPA/father	.43	CHICO/guy	.44
HORRIBLE/horrible	.39	AMIGOS/friends	.44
CASA/house	.39	NOVIO/boyfriend	.43
SUICIDIO/suicide	.36	JUNTOS/together	.37
DINERO/money	.34	ESPECIAL/special	.36
ESTUDIOS/studies	.33	EMPECE/start	.33
CLASE/class	.32	TIEMPO/time	.33
FEA/ugly	.32	DEJAR/leave	.33
ASCO/disgust	.32	FINAL/end	.33
COMER/eat	.31	HISTORIA/history	.31
PEQUENIOO/small	.31	MESES/months	.31
FAMILIA/family	.31	LLEGAR/arrive	.31

FACTOR 3: Hopelessness		FACTOR 4: School	
NOCHE/night	.45	TIMIDA/timid	.41
SEGUNDO/second	.44	CARRERA/career	.41
MORIR/die	.42	COLEGIO/school	.40
PAZ/peace	.39	CONFIANZA/trust	.37
OJOS/eyes	.39	ESTUDIOS/studies	.37
ESPERO/hope	.38	INCAPAZ/incapable	.33
TERRIBLE/terrible	.36	UNIVERSI./university	.33
FUERTE/strong	.34	TONTERIA/foolishness	.32
CORAZON/heart	.31		
SUENIOS/dreams	.31		

FACTOR 5: Treatment
PSICOLOGO/psychologist
ANSIEDAD/anxiety
EMPRESA/company
ANTIDEP/antidepressants
SINTOMAS/symptoms
MEDICAMENTO/medicines

Speech features for Depression

Sanchez, Michelle Hewlett, Dimitra Vergyri, Luciana Ferrer, Colleen Richey, Pablo Garcia, Bruce Knoth, and William Jarrold. "Using Prosodic and Spectral Features in Detecting Depression in Elderly Males." In *INTERSPEECH*, pp. 3001-3004. 2011.

Alghowinem, Sharifa, Roland Goecke, Michael Wagner, Julien Epps, Michael Breakspear, and Gordon Parker. "From Joyous to Clinically Depressed: Mood Detection Using Spontaneous Speech." In *FLAIRS Conference*. 2012.

Moore, Elliot, Mark A. Clements, John W. Peifer, and Lydia Weisser. "Critical analysis of the impact of glottal features in the classification of clinical depression in speech." *Biomedical Engineering, IEEE Transactions on* 55, no. 1 (2008): 96-107

D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, and D. M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 829–837, Jul. 2000.

E. Moore, II, M. Clements, J. Peifer, and L. Weisser, "Comparing objective feature statistics of speech for classifying clinical depression," in *Proc. 26th Ann. Conf. Eng. Med. Biol.*, 2004, vol. 1, pp. 17–20.

Commonly used features:

- F0 variance (monopitch)
- loudness variance (monoloudness)
- rate of speech (slower)
  - response delay, pauses
- spectral features

Topic 3: Trauma

Social stage model of collective coping (Pennebaker & Harber, 1993). After a traumatic experience:

- Stage 1: people cope by sharing their thoughts about the upsetting experience
- Stage 2, a few weeks later: decrease in talking, but still thinking about event
- Stage 3: 6-8 weeks later: reduction in both talking and thinking

What are the linguistic characteristics of stage 1?

Cohn, Mehl, Pennebaker: Linguistic Markers of Psychological Change Surrounding September 11, 2001

- 1084 LiveJournal users
- all blog entries for 2 months before and after 9/11
- Lumped prior two months into one “baseline” corpus.
- Investigated changes after 9/11 compared to that baseline
- Using LIWC categories

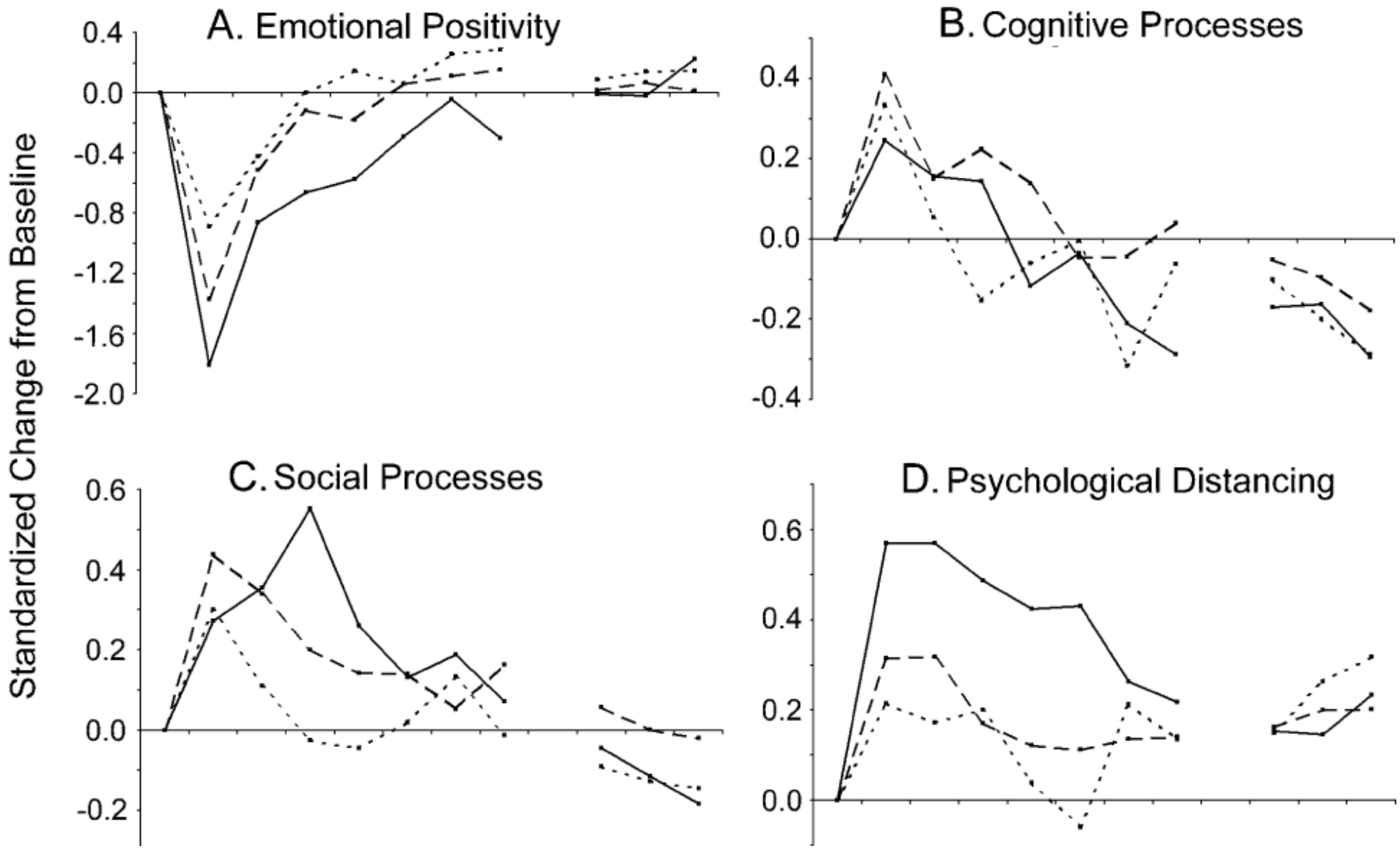
Variables examined

- Emotional positivity
  - difference between LIWC scores for positive emotion words (happy, good, nice) and negative emotion words (kill, ugly, guilty).
- cognitive processing
  - think, question, because: concerned with organizing and intellectually understanding issues
- social orientation
  - talk, share, friends and personal pronouns besides I/me. (essentially counts # of references to other people)

Last factor: Psychological Distancing

- psychological distancing
  - factor-analytic:
    - + articles,
    - + words > 6 letters long
    - - I/me/mine
    - - would/should/could
    - - present tense verbs
  - low score = personal, experiential lg, focus on here and now
  - high score: abstract, impersonal, rational tone

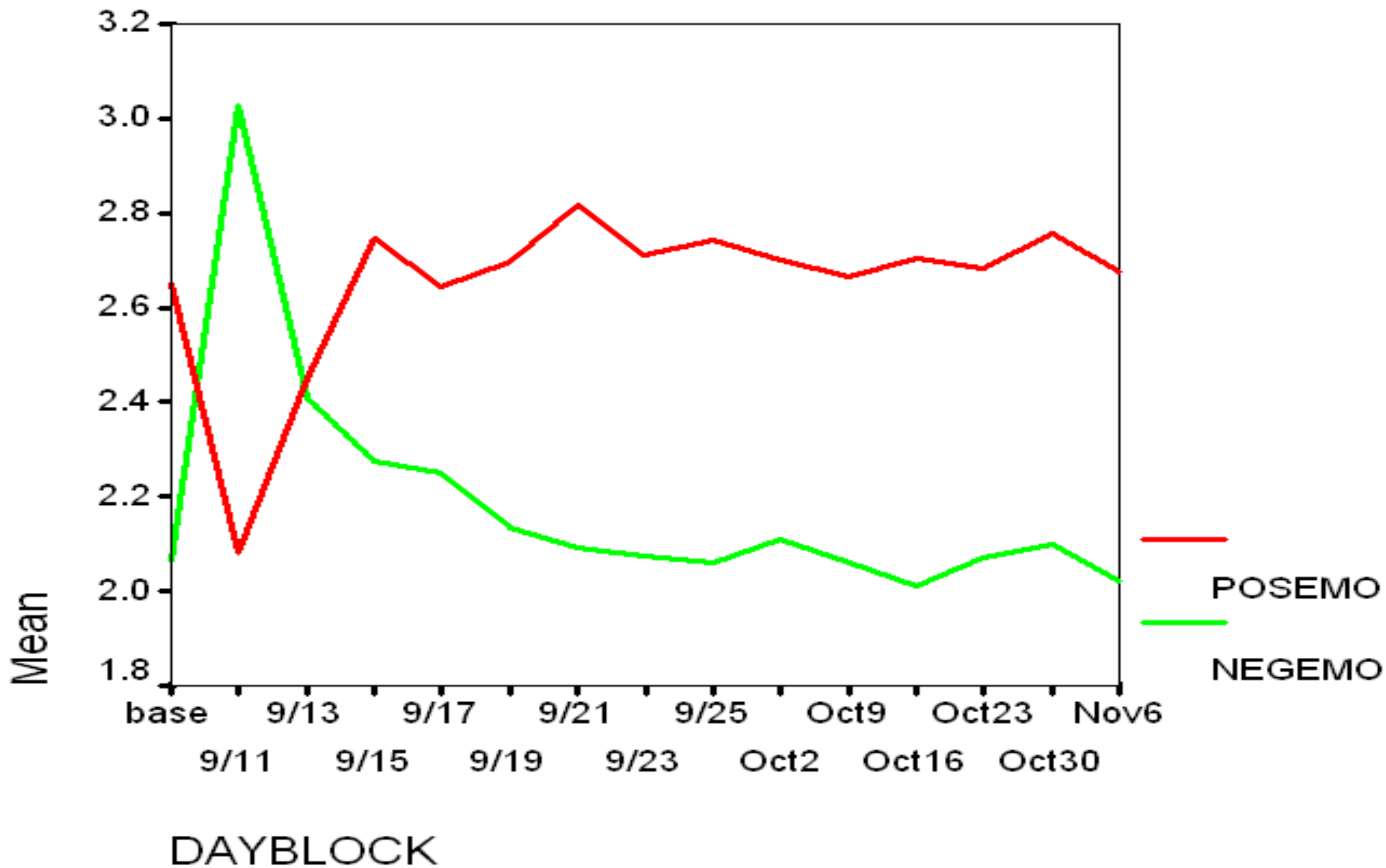
Results



A. Emotional Positivity
B. Cognitive Processes
C. Social Processes
D. Psychological Distancing
Standardized Change from Baseline
0.4
0.0
-0.4
-0.8
-1.2
-1.6
-2.0
0.2
-0.2
0.6

LiveJournal.com September 11, 2001 study: Positive and negative emotion words

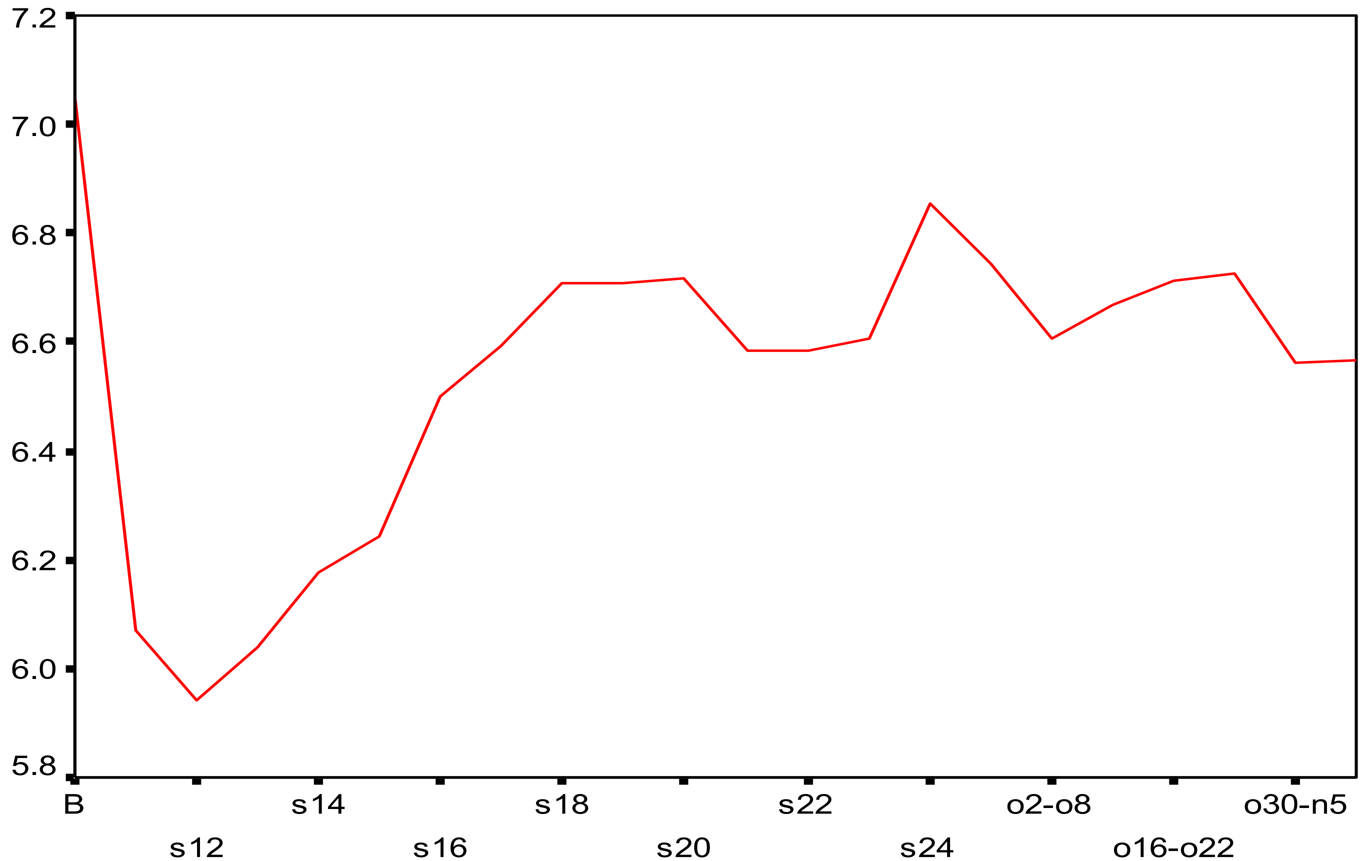
Cohn, Mehl, Pennebaker. 2004. Linguistic markers of psychological change surrounding September 11, 2001. Psychological Science 15, 10: 687-693.



Graph from Pennebaker slides

Livejournal.com:
I, me, my on or after Sep 11, 2001

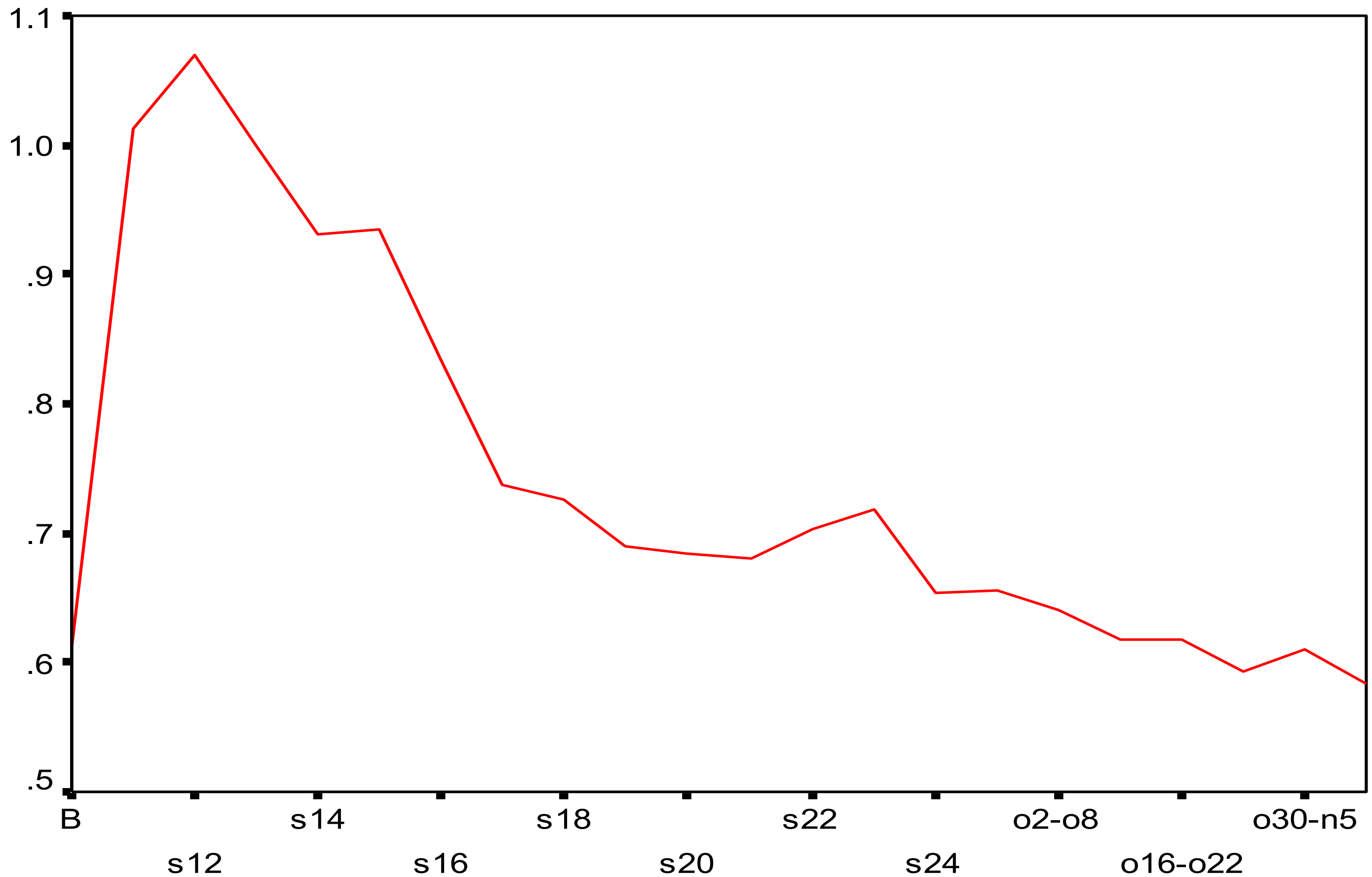
Cohn, Mehl, Pennebaker. 2004. Linguistic markers of psychological change surrounding September 11, 2001. Psychological Science 15, 10: 687-693.



Graph from Pennebaker slides

September 11 LiveJournal.com study: *We, us, our*

Cohn, Mehl, Pennebaker. 2004. Linguistic markers of psychological change surrounding September 11, 2001. Psychological Science 15, 10: 687-693.



Graph from Pennebaker slides

Trauma after Princess Diana's death

- Princess Diana died August 30, 1997
- Over the next 4 weeks, scraped all conversations from "The UK Experience" chat room on AOL.
- 121 chat sessions among 3,139 participants.
- Compared to baseline rates:
  - Increase in we
  - Decrease in I
  - Increase in negative emotional words

Texas A&M Bonfire tragedy

- Gortner and Pennebaker
- Examined student newspaper in the weeks after the tragedy:

- Increase in we
- Increase in I
- Increase in negative emotion

Another domain of trauma? Restaurant Reviews

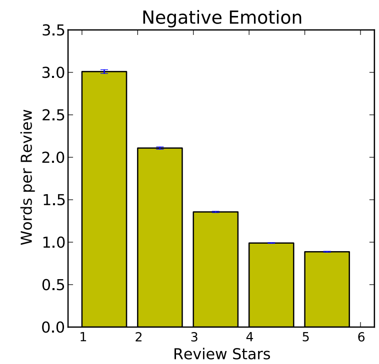
Jurafsky, Chahuneau, Routledge, Smith 2014.

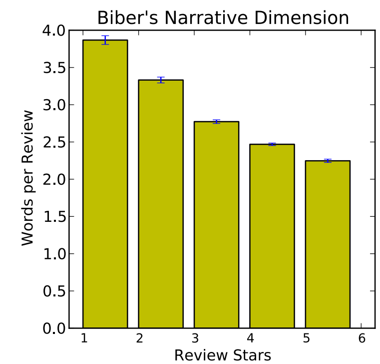
- 6562 restaurants
  - 900K reviews www.yelp.com
- Negative (★):
  - The bartender... absolutely horrible... we waited 10 min before we even got her attention... and then we had to wait 45 - FORTY FIVE! - minutes for our entrees... stalk the waitress to get the cheque... she didn't make eye contact or even break her stride to wait for a response...

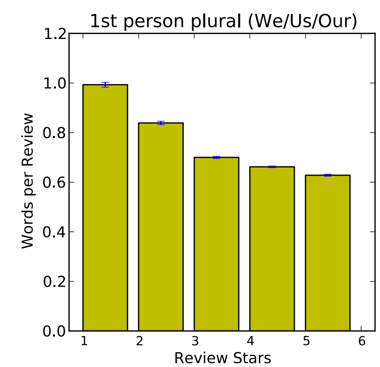
What makes a bad review bad?

- Negative sentiment language
  - horrible awful terrible worst bad disgusting
- narrative
  - past tense
    - *waited*, *didn't* make eye contact, *was* disappointing.
  - 3rd person pronouns
    - he she his her
  - other people
    - manager, customer, minutes, money, waitress, waiter, bill, attitude, management, business, apology, mistake, table, charge, order, hostess,
- mentions of we and us

we waited 10 min before **we** even got her attention... and then **we** had to wait 45 - FORTY FIVE! - minutes for **our** entrees... ...







We just saw texts with these characteristics!

- Negative sentiment, past tense narratives about others
- Enormous increase in “we” and “us”: solace in community
- Chat group discussions after Princess Diana’s death
  - Stone, L.D. & Pennebaker, J.W. (2002). Trauma in real time: Talking and avoiding online conversations about the death of Princess Diana. Basic and Applied Social Psychology, 24, 172-182
- Blog posts after September 11, 2001
  - Cohn, M.A., Mehl, M.R., & Pennebaker, J.W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. Psychological Science, 15, 687-693
- Student newspaper reports after a campus tragedy
  - Gortner, E.-M., & Pennebaker, J.W. (2003). The archival anatomy of a disaster: Media coverage and community-wide health effects of the Texas A&M Bonfire Tragedy. Journal of Social and Clinical Psychology, 22, 580-603

- Conclusion: **Awful reviews are trauma narratives**

Personality

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

Personality and Cultural Values

- Personality refers to the structures and propensities inside a person that explain his or her characteristic patterns of thought, emotion, and behavior.
  - Personality captures what people are like.
  - Traits are defined as recurring regularities or trends in people's responses to their environment.
    - Cultural values, defined as shared beliefs about desirable end states or modes of conduct in a given culture, influence the expression of a person's traits.

The Big Five Dimensions of Personality

- Extraversion vs. Introversion
  (sociable, assertive, playful vs. aloof, reserved, shy)
- Emotional stability vs. Neuroticism
  (calm, unemotional vs. insecure, anxious)
- Agreeableness vs. Disagreeable
  (friendly, cooperative vs. antagonistic, faultfinding)
- Conscientiousness vs. Unconscientious
  (self-disciplined, organised vs. inefficient, careless)
- Openness to experience
  (intellectual, insightful vs. shallow, unimaginative)

Aside: Do Animals Have Personalities?

- Gosling (1998) studied spotted hyenas. He:
  - had human observers use personality scales to rate the different hyenas in the group
  - did a factor analysis on these findings
  - found five dimensions

    three closely resembled the Big Five traits of neuroticism, openness to experience, and agreeableness

Slide from **Randall E. Osborne**

The Big Five Personality Traits

- Conscientiousness - dependable, organized, reliable, ambitious, hardworking, and persevering.

The Big Five Personality Traits, Cont'd

- Agreeableness - warm, kind, cooperative, sympathetic, helpful, and courteous.
  - Strong desire to obtain acceptance in personal relationships as a means of expressing personality.
  - Agreeable people focus on "getting along," not necessarily "getting ahead."

The Big Five Personality Traits, Cont’d

- Extraversion - talkative, sociable, passionate, assertive, bold, and dominant.
  - Easiest to judge in zero acquaintance situations — situations in which two people have only just met.
  - Prioritize desire to obtain power and influence within a social structure as a means of expressing personality.
  - High in positive affectivity — a tendency to experience pleasant, engaging moods such as enthusiasm, excitement, and elation.

The Big Five Personality Traits:

Neuroticism - nervous, moody, emotional, insecure, jealous.

- experience unpleasant moods such as hostility, nervousness, and annoyance.
- more likely to appraise day-to-day situations as stressful.
- less likely to believe they can cope with the stressors that they experience.
- related to locus of control (attribute causes of events to themselves or to the external environment)
  - neurotics hold an external locus of control: believe that the events that occur around them are driven by luck, chance, or fate.
  - less neurotic people hold internal locus of control: believe that their own behavior dictates events.

External and Internal Locus of Control

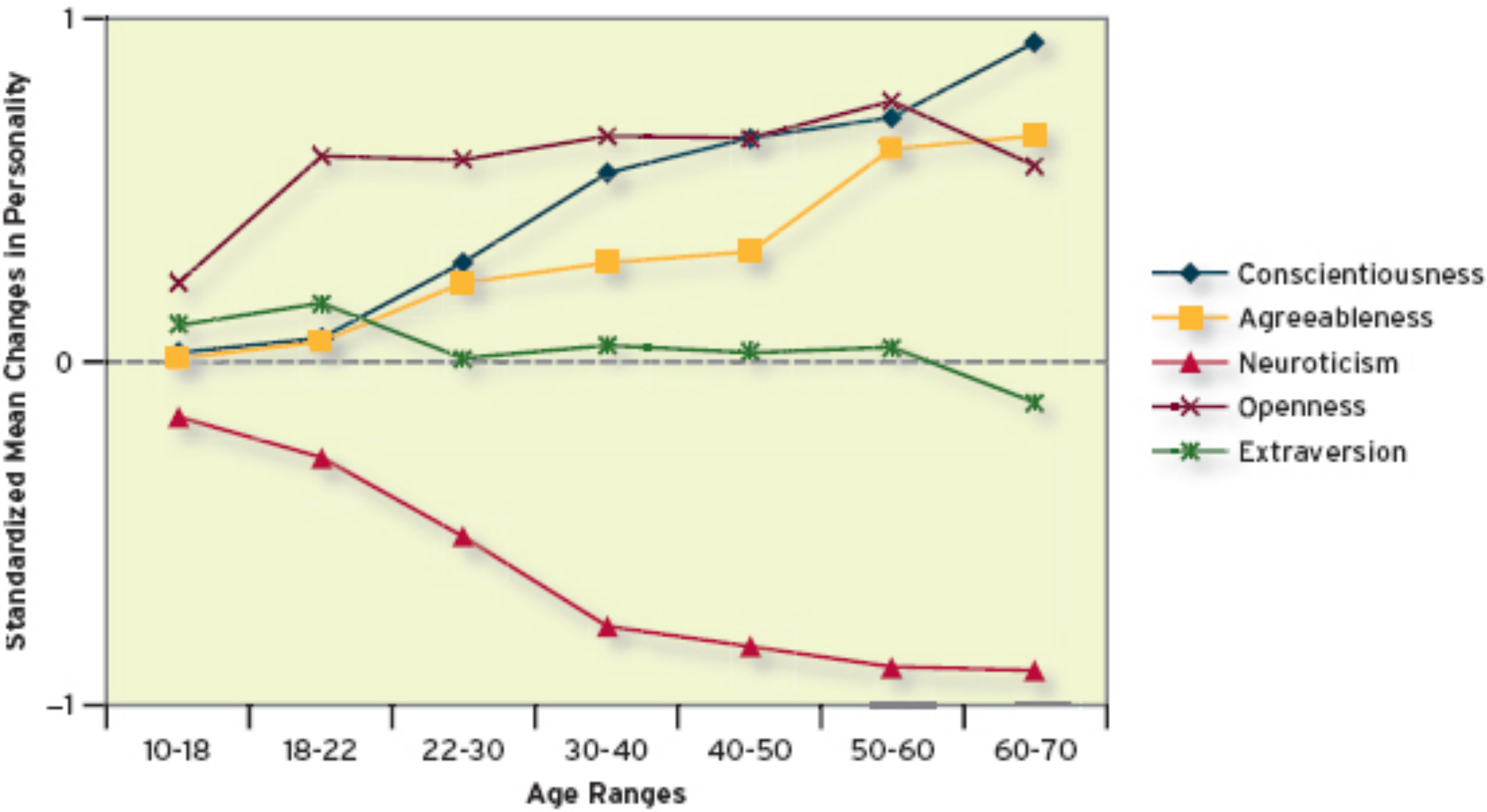
PEOPLE WITH AN EXTERNAL LOCUS OF CONTROL TEND TO BELIEVE:	PEOPLE WITH AN INTERNAL LOCUS OF CONTROL TEND TO BELIEVE:
Many of the unhappy things in people's lives are partly due to bad luck.	People's misfortunes result from the mis-takes they make.
Getting a good job depends mainly on being in the right place at the right time.	Becoming a success is a matter of hard work; luck has little or nothing to do with it.
Many times exam questions tend to be so unrelated to course work that studying is really useless.	In the case of the well-prepared student, there is rarely if ever such a thing as an unfair test.
This world is run by the few people in power, and there is not much the little guy can do about it.	The average citizen can have an influence in government decisions.
There's not much use in trying too hard to please people; if they like you, they like you.	People are lonely because they don't try to be friendly.

The Big Five Personality Traits, Cont'd

Openness to experience - curious, imaginative, creative, complex, refined, and sophisticated.

- Also called "Inquisitiveness" or "Intellectualness" or even "Culture."
- high levels of creativity, the capacity to generate novel and useful ideas and solutions.
- Highly open individuals are more likely to migrate into artistic and scientific fields.

Changes in Big Five Dimensions Over the Life Span



Take the Big Five Inventory

http://www.outofservice.com/bigfive/

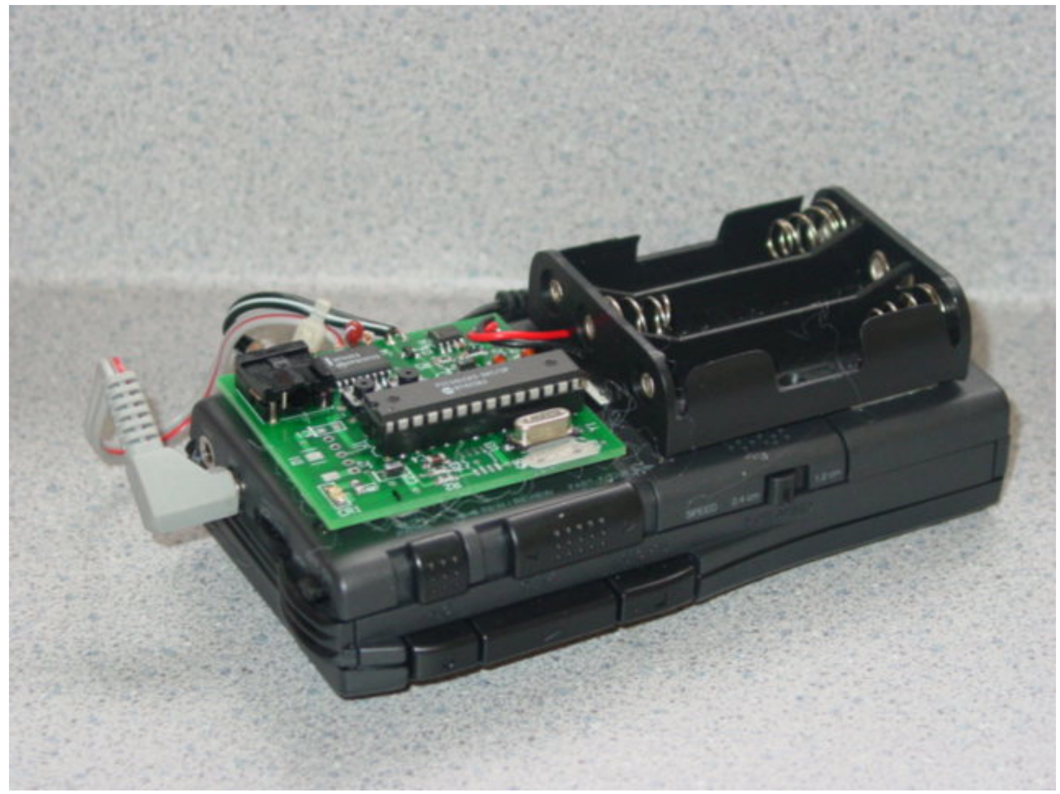
Corpora for studying personality: Natural speech

Electronically Activated Recorder (EAR)

Mehl, M. R., Pennebaker, J. W., Crow, M. D., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, and Computers, 33,* 517-523.

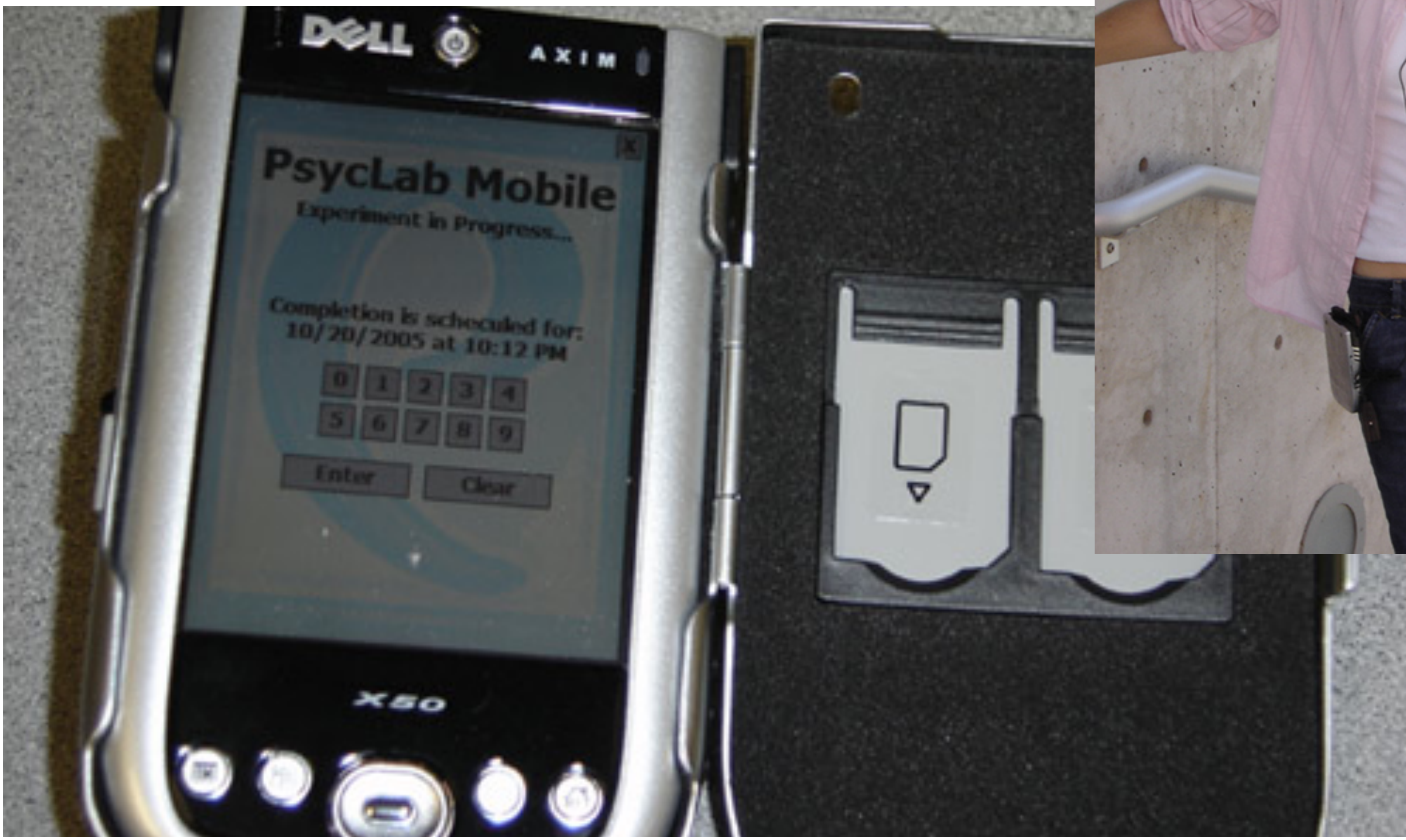
- a modified digital voice recorder that periodically records brief snippets of ambient sounds
- Attaches to the belt or in a purse-like bag while participants go about their daily lives.

Analog EAR-1: 90 minute tape 1997-2000



Digital EAR-2: digital voice recorder, flash drive 2001-2004

PDA Ear-3 2005-



Mairesse et al. Two Corpora

- Pennebaker and King (1999)
  - 2,479 essays from psychology students (1.9 million words), “write whatever comes into your mind” for 20 minutes
- Mehl et al. (2006)
  - Speech from Electronically Activated Recorder (EAR)
  - Random snippets of conversation recorded, transcribed
  - 96 participants, total of 97,468 words and 15,269 utterances).

Mehl *et al.* (2006) data

Mehl, Matthias R., Samuel D. Gosling, and James W. Pennebaker. 2006. "Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life." *Journal of personality and social psychology*

- 96 psych freshman at UT Austin took the 44-item Big Five Inventory
- Agreed to wear EAR two weekdays continuously (when awake)
  - External mike clipped to collar
- 30-s on, 12.5-min off cycle = 4.8 recordings/hour
  - They were told they could erase anything they didn’t want researchers to hear
  - afterwards they reported wearing about 75% of their waking time
- Each sound file
  - transcribed
  - coded for environmental situation (location, activity)
  - 23 LIWC variables coded
  - 18 trained students listened to the files and assigned Big Five Inventory scores

Ears (speech) corpus

Introvert	Extravert
- Yeah you would do kilograms. Yeah I see what you're saying. - On Tuesday I have class. I don't know. - I don't know. A16. Yeah, that is kind of cool. - I don't know. I just can't wait to be with you and not have to do this every night, you know? - Yeah. You don't know. Is there a bed in there? Well ok just...	- That's my first yogurt experience here. Really watery. Why? - Damn. New game. - Oh. - That's so rude. That. - Yeah, but he, they like each other. He likes her. - They are going to end up breaking up and he's going to be like.
Unconscientious	**Conscientious**
- With the Chinese. Get it together. - I tried to yell at you through the window. Oh. xxxx's fucking a dumb ass. Look at him. Look at him, dude. Look at him. I wish we had a camera. He's fucking brushing his t-shirt with a tooth brush. Get a kick of it. Don't steal nothing.	- I don't, I don't know for a fact but I would imagine that historically women who have entered prostitution have done so, not everyone, but for the majority out of extreme desperation and I think. I don't know, i think people understand that desperation and they don't don't see [...]

Essays corpus

Introvert	**Extravert**
I've been waking up on time so far. What has it been, 5 days? Dear me, I'll never keep it up, being such not a morning person and all. But maybe I'll adjust, or not. I want internet access in my room, I don't have it yet, but I will on Wed??? I think. But that ain't soon enough, cause I got calculus homework [...]	I have some really random thoughts. I want the best things out of life. But I fear that I want too much! What if I fall flat on my face and don't amount to anything. But I feel like I was born to do BIG things on this earth. But who knows... There is this Persian party today.
Neurotic	**Emotionally stable**
One of my friends just barged in, and I jumped in my seat. This is crazy. I should tell him not to do that again. I'm not that fastidious actually. But certain things annoy me. The things that would annoy me would actually annoy any normal human being, so I know I'm not a freak.	I should excel in this sport because I know how to push my body harder than anyone I know, no matter what the test I always push my body harder than everyone else. I want to be the best no matter what the sport or event. I should also be good at this because I love to ride my bike.

Sample Features

Feature	**Type**	**Example**
Anger words	LIWC	hate, kill, pissed
Metaphysical issues	LIWC	God, heaven, coffin
Physical state/function	LIWC	ache, breast, sleep
Inclusive words	LIWC	with, and, include
Social processes	LIWC	talk, us, friend
Family members	LIWC	mom, brother, cousin
Past tense verbs	LIWC	walked, were, had
References to friends	LIWC	pal, buddy, coworker
Imagery of words	MRC	Low: future, peace - High: table, car
Syllables per word	MRC	Low: a - High: uncompromisingly
Concreteness	MRC	Low: patience, candor - High: ship
Frequency of use	MRC	Low: duly, nudity - High: he, the

LIWC FEATURES (Pennebaker et al., 2001):

- **Standard counts:**
  - Word count (WC), words per sentence (WPS), type/token ratio (Unique), words captured (Dic), words longer than 6 letters (Sixltr), negations (Negate), assents (Assent), articles (Article), prepositions (Preps), numbers (Number)
  - Pronouns (Pronoun): 1st person singular (I), 1st person plural (We), total 1st person (Self), total 2nd person (You), total 3rd person (Other)
- **Psychological processes:**
  - Affective or emotional processes (Affect): positive emotions (Posemo), positive feelings (Posfeel), optimism and energy (Optim), negative emotions (Negemo), anxiety or fear (Anx), anger (Anger), sadness (Sad)
  - Cognitive Processes (Cogmech): causation (Cause), insight (Insight), discrepancy (Discrep), inhibition (Inhib), tentative (Tentat), certainty (Certain)
  - Sensory and perceptual processes (Senses): seeing (See), hearing (Hear), feeling (Feel)
  - Social processes (Social): communication (Comm), other references to people (Othref), friends (Friends), family (Family), humans (Humans)
- **Relativity:**
  - Time (Time), past tense verb (Past), present tense verb (Present), future tense verb (Future)
  - Space (Space): up (Up), down (Down), inclusive (Incl), exclusive (Excl)
  - Motion (Motion)
- **Personal concerns:**
  - Occupation (Occup): school (School), work and job (Job), achievement (Achieve)
  - Leisure activity (Leisure): home (Home), sports (Sports), television and movies (TV), music (Music)
  - Money and financial issues (Money)
  - Metaphysical issues (Metaph): religion (Relig), death (Death), physical states and functions (Physcal), body states and symptoms (Body), sexuality (Sexual), eating and drinking (Eating), sleeping (Sleep), Grooming (Groom)
- **Other dimensions:**
  - Punctuation (Allpct): period (Period), comma (Comma), colon (Colon), semi-colon (Semic), question (Qmark), exclamation (Exclam), dash (Dash), quote (Quote), apostrophe (Apostro), parenthesis (Parenth), other (Otherp)
  - Swear words (Swear), nonfluencies (Nonfl), fillers (Fillers)

Utterance type

Labeled by parsing each utterance and then using heuristic rules based on parse tree:

Commands: imperatives, “can you”, etc.

Backchannels: yeah, ok, uh-huh, huh

Questions

Assertions (anything else)

Prosodic features

Computed via Praat

pitch (mean, min, max, sd):

intensity (mean, min, max, sd)

voiced time

rate of speech (words/second)

Classifiers from Weka

- Classification (binary)
  C4.5 Decision Tree (J48)
  Nearest neighbor
  Naïve Bayes
  Ripper
  Adaboost
  SVM with linear kernels
- Regression (predict Likert values)
  linear regression
  M5' regression tree
  SVMOreg
- Ranking (training set T of ordered pairs
  T = {(x,y)|x,y, are language samples from two individuals, x has a higher score than y for that personality trait}
  Rankboost

Ears (speech) corpus

Data	Trait	Base	J48	NN	NB	JRIP	ADA	SMO
Obs	Extra	47.78	66.78	59.33	73.00•	60.44	73.00 •	65.78
Obs	Emot	51.11	62.56	58.22	73.89•	56.22	48.78	60.33
Obs	Agree	47.78	48.78	51.89	61.33•	51.89	52.89	56.33
Obs	Consc	47.78	57.67	61.56	67.67•	61.56	60.22 •	57.11
Obs	Open	47.78	52.22	46.78	57.00	49.67	50.56	55.89
Self	Extra	47.78	48.78	49.67	57.33	50.56	54.44	49.89
Self	Emot	51.11	45.56	46.78	50.44	46.78	41.89	44.33
Self	Agree	52.22	47.89	50.89	58.33	56.89	55.22	52.33
Self	Consc	51.11	33.44	45.56	39.33	43.11	46.11	53.22
Self	Open	51.11	52.00	42.22	61.44	45.00	56.00	47.78

Ears (speech) corpus, from observer, Naïve Bayes classifier

Feature set	None	Type	LIWC	MRC	Prosody	All
Set size	0	4	88	14	11	
Extraversion	47.78	45.67	68.89●	68.78●	67.56●	73
Emotional stability	51.11	60.22	69.89●	60.78	61.78	73.89
Agreeableness	47.78	57.56	54.00	58.67	50.44	61.33
Conscientiousness	47.78	59.67	60.22	66.78●	52.11	67.67
Openness to experience	47.78	53.11	61.11	54.00	**64.56●**	57

Summary

- Much easier to classifier observer-labeled than self-labeled
- Simpler classifiers like NB did well
  - not much data: 96 people, 97K words

Feature analysis: Observed Extraversion

more words

higher pitch

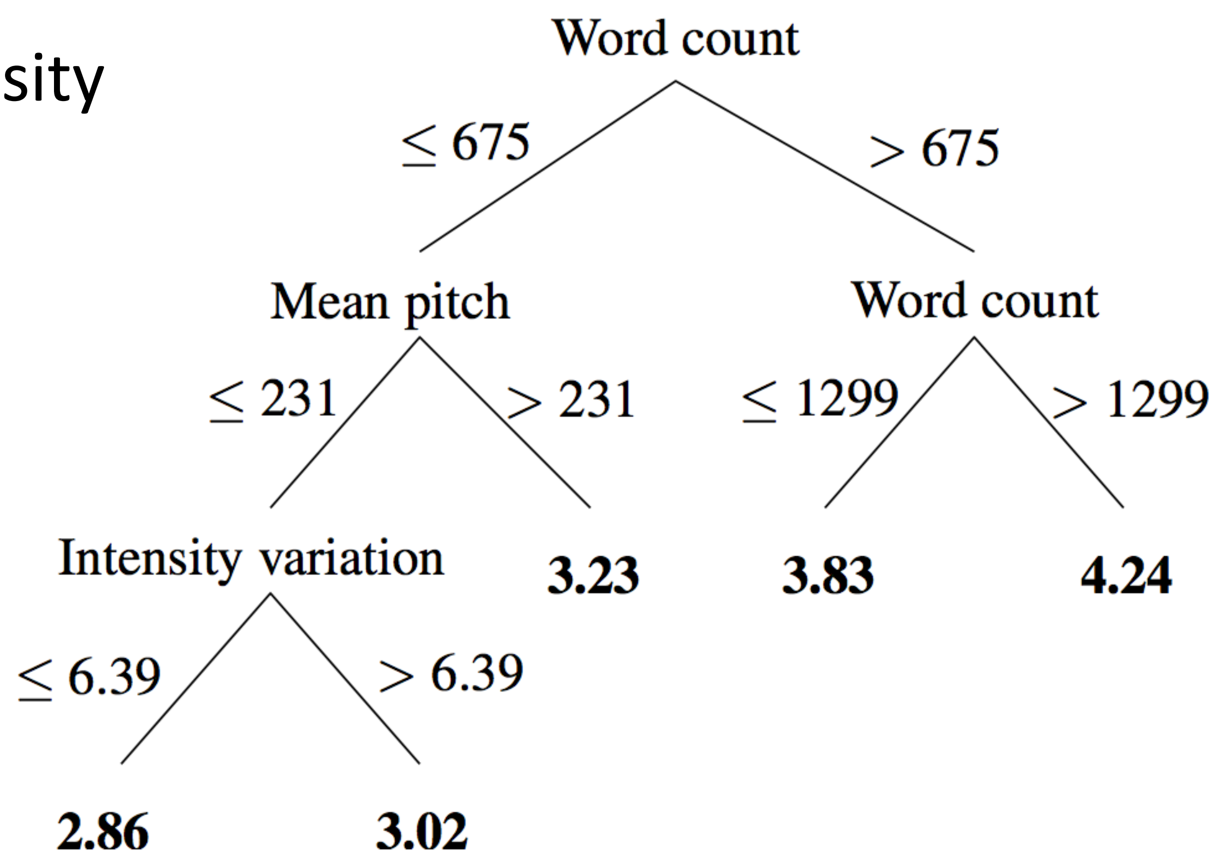
more concrete, imageable words

greater variation in intensity

greater mean intensity

more word repetitions

M5’ Regression Tree



Agreeableness

-swear	**Self-assessed**:	**Other-assessed**:
-anger	pitch variation	long words, short sents
+backchannel	max intensity	

other-assessed:

Agreeableness model with all features					
#	**Positive rules**	α	#	**Negative rules**	α
1	Nphon $\geq$ 2.66	0.56	11	Fam $\geq$ 601.61	-0.16
2	Tentat $\geq$ 2.83	0.50	12	Swear $\geq$ 0.41	-0.18
3	Colon $\geq$ 0.03	0.41	13	Anger $\geq$ 0.92	-0.19
4	Posemo $\geq$ 2.67	0.32	14	Time $\geq$ 3.71	-0.20
5	Voiced $\geq$ 584	0.32	15	Negate $\geq$ 3.52	-0.20
6	Relig $\geq$ 0.43	0.27	16	Fillers $\geq$ 0.54	-0.22
7	Insight $\geq$ 2.09	0.25	17	Time $\geq$ 3.69	-0.23
8	Prompt $\geq$ 0.06	0.25	18	Swear $\geq$ 0.61	-0.27
9	Comma $\geq$ 4.60	0.23	19	Swear $\geq$ 0.45	-0.27
10	Money $\geq$ 0.38	0.20	20	WPS $\geq$ 6.13	-0.45

Conscientiousness

- -swear
- -anger
- -negemotion

Conscientiousness model with all features					
#	**Positive rules**	α	#	**Negative rules**	α
1	Occup $\geq$ 1.21	0.37	11	Swear $\geq$ 0.20	-0.18
2	Insight $\geq$ 2.15	0.36	12	WPS $\geq$ 6.25	-0.19
3	Posfeel $\geq$ 0.30	0.30	13	Pitch-mean $\geq$ 229	-0.20
4	Int-stddev $\geq$ 7.83	0.29	14	Othref $\geq$ 7.64	-0.20
5	Nlet $\geq$ 3.29	0.27	15	Humans $\geq$ 0.83	-0.21
6	Comm $\geq$ 1.20	0.26	16	Swear $\geq$ 0.93	-0.21
7	Nphon $\geq$ 2.66	0.25	17	Swear $\geq$ 0.17	-0.24
8	Nphon $\geq$ 2.67	0.22	18	Relig $\geq$ 0.32	-0.27
9	Nphon $\geq$ 2.76	0.20	19	Swear $\geq$ 0.65	-0.31
10	K-F-nsamp $\geq$ 329	0.19	20	Int-max $\geq$ 86.84	-0.50

- Observed:
  - +insight, +backchannel, +longwords
  - +word, +posemotion
- Self-assessed:
  - +positive feelings

Openness to experience

- Poor performance from Ears data – prosody helped but no language features
- But good performance from Essay data
  - Open/creative/unconventional people
    - don’t talk about school
    - use longer and rarer words
    - don’t talk about friends

#	**Ordered rules**
1	(School $\geq$ 1.47) and (Motion $\geq$ 1.71) $\Rightarrow$ NOT OPEN
2	(Occup $\geq$ 2.49) and (Sixltr $\leq$ 13.11) and (School $\geq$ 1.9) and (I $\geq$ 10.5) $\Rightarrow$ NOT OPEN
3	(Fam $\geq$ 600.335106) and (Friends $\geq$ 0.67) $\Rightarrow$ NOT OPEN
4	(Nlet $\leq$ 3.502543) and (Number $\geq$ 1.13) $\Rightarrow$ NOT OPEN
5	(School $\geq$ 0.98) and (You $\leq$ 0) and (AllPct $\leq$ 13.4) $\Rightarrow$ NOT OPEN
6	Any other feature values $\Rightarrow$ OPEN

Interspeech 2012 Paralinguistic challenge dataset

- SPC
- Speech clips randomly extracted from Radio Suisse Romand French news broadcasts
- 640 10-second speech clips from 322 individuals
- Emotionally neutral, no familiar words to non-French speakers
- Professional (307 samples; journalists) or nonprofessional (333 - interviewees) samples.
- Personality assessed by 11 judges

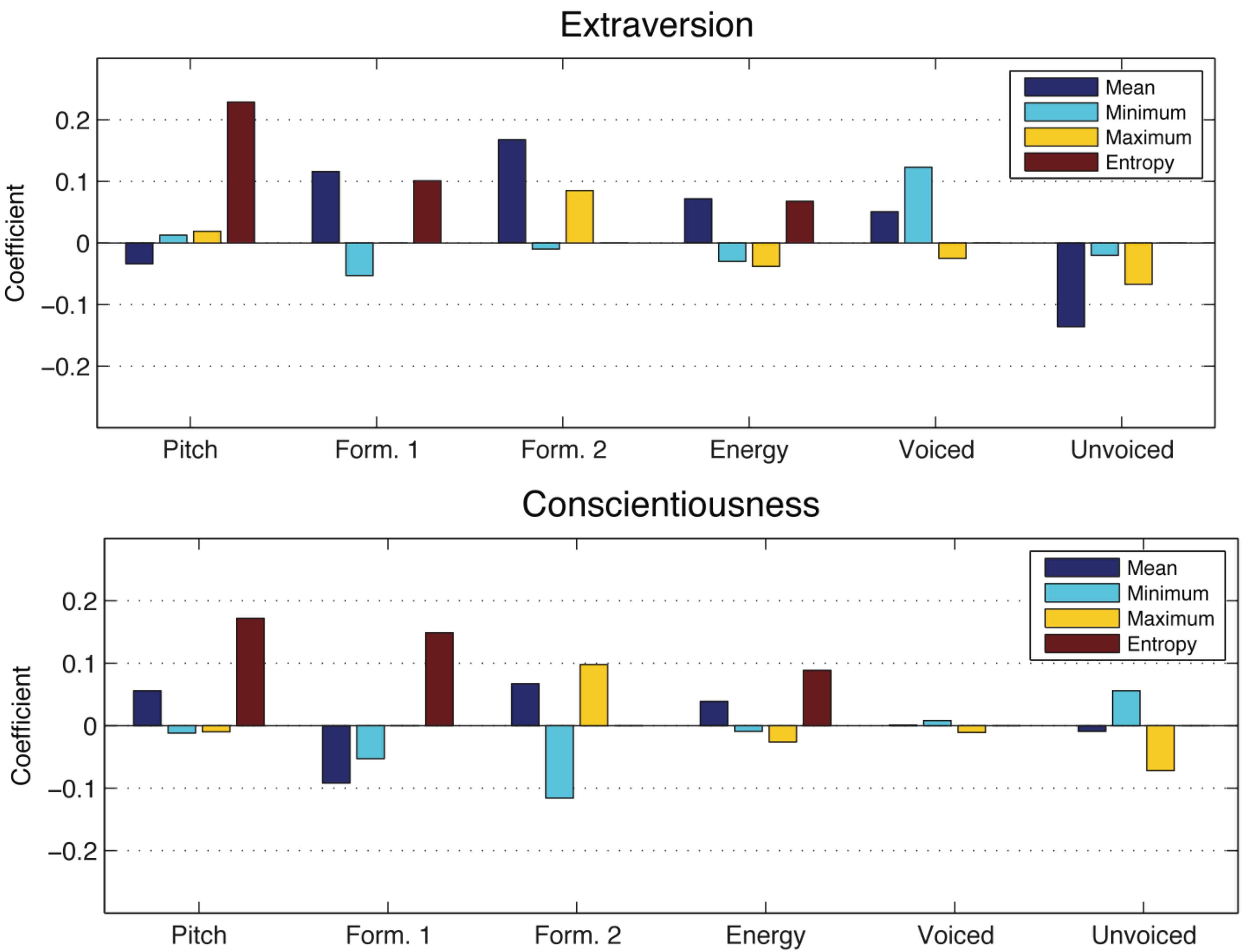
Personality labeled by BFI-10

ID	Question
1	This person is reserved
2	This person is generally trusting
3	This person tends to be lazy
4	This person is relaxed, handles stress well
5	This person has few artistic interests
6	This person is outgoing, sociable
7	This person tends to find fault with others
8	This person does a thorough job
9	This person gets nervous easily
10	This person has an active imagination

Extroversion: Q6 – Q1 Agreeableness: Q2 – Q7
Conscientiousness Q8 – Q3 Neuroticism Q9 – Q4
Openness: Q10 – Q5

Accuracy

Extraversion	73.5 ± 3.4
Agreeableness	63.1 ± 3.7
Conscientiousness	71.3 ± 3.5
Neuroticism	65.9 ± 3.7
Openness	60.1 ± 3.8


Extraversion
Coefficient
0.2
0.1
0
−0.1
−0.2
Mean
Minimum
Maximum
Entropy
Pitch
Form. 1
Form. 2
Energy
Voiced
Unvoiced
Conscientiousness
Coefficient
0.2
0.1
0
−0.1
−0.2
Mean
Minimum
Maximum
Entropy
Pitch
Form. 1
Form. 2
Energy
Voiced
Unvoiced