Detecting Deceptive Speech: Humans vs. Machines

Julia Hirschberg
Columbia University
ICASSP 2018

Co-Authors and Collaborators

- ***<u>Sarah Ita Levitan</u>***
- ***<u>Michelle Levine</u>***
- ***Guozhen An***
- ***Andrew Rosenberg***
- ***Gideon Mendels***
- ***Rivka Levitan***
- ***Kara Schechtman***
- ***Mandi Wang***
- ***Nishmar Cestero***
- ***Kai-Zhan Lee***
- ***Yocheved Levitan***
- ***<u>Angel Maredia</u>***
- ***<u>Jessica Xiang</u>***
- ***Bingyan Hu***
- ***Zoe Baker-Peng***
- ***Ivy Chen***
- ***<u>Meredith Cox</u>***
- ***Leighanne Hsu***
- ***Yvonne Missry***
- ***<u>Gauri Narayan</u>***
- ***Molly Scott***
- ***Jennifer Senior***
- ***<u>James Shin</u>***
- ***Grace Ulinksi***

The Columbia Speech Lab

Deceptive Speech

- ***Deliberate choice to mislead***
  - ***Without*** prior notification
  - To gain some ***advantage*** or to avoid some ***penalty***
- ***Not***:
  - Self-deception, delusion, pathological behavior
  - Theater
  - Falsehoods due to ignorance/error
- ***Everyday (White) Lies*** very hard to detect
- But ***Serious Lies*** may be easier to detect

Why would Serious Lies be easier to identify?

- Hypotheses in research and among practitioners:
  - Our ***cognitive load*** is increased when lying because we …
    - Must keep story straight
    - Must remember what we **have** said and what we have ***not*** said
  - Our ***fear of detection*** is increased if…
    - We believe our target is difficult to fool
    - Stakes are high: serious rewards and/or punishments
- ***Makes it hard for us to control potential indicators of deception***

But Humans very poor at Recognizing these Cues: Aamodt & Mitchell 2004 Meta-Study ()

Group	**#Studies**	**#Subjects**	**Accuracy %**
Criminals	1	52	65.40
Secret service	**1**	**34**	**64.12**
Psychologists	4	508	61.56
Judges	**2**	**194**	**59.01**
Cops	**8**	**511**	**55.16**
Federal officers	**4**	**341**	**54.54**
Students	122	8,876	54.20
Detectives	**5**	**341**	**51.16**
Parole officers	**1**	**32**	**40.42**

Current Approaches to Deception Detection

- ***'Automatic' methods*** (polygraph, commercial products) no better than chance
- ***Train humans***: John Reid & Associates
  - Behavioral Analysis: Interview/Interrogation no empirical support
  - *Truth*: *I didn't take the money* vs. Lie: *I did not take the money* (**but non-native speakers rarely use contractions so….**)
- ***Laboratory studies***: Production and perception (facial expression, body posture/gesture, statement analysis, brain activation, odor,…)

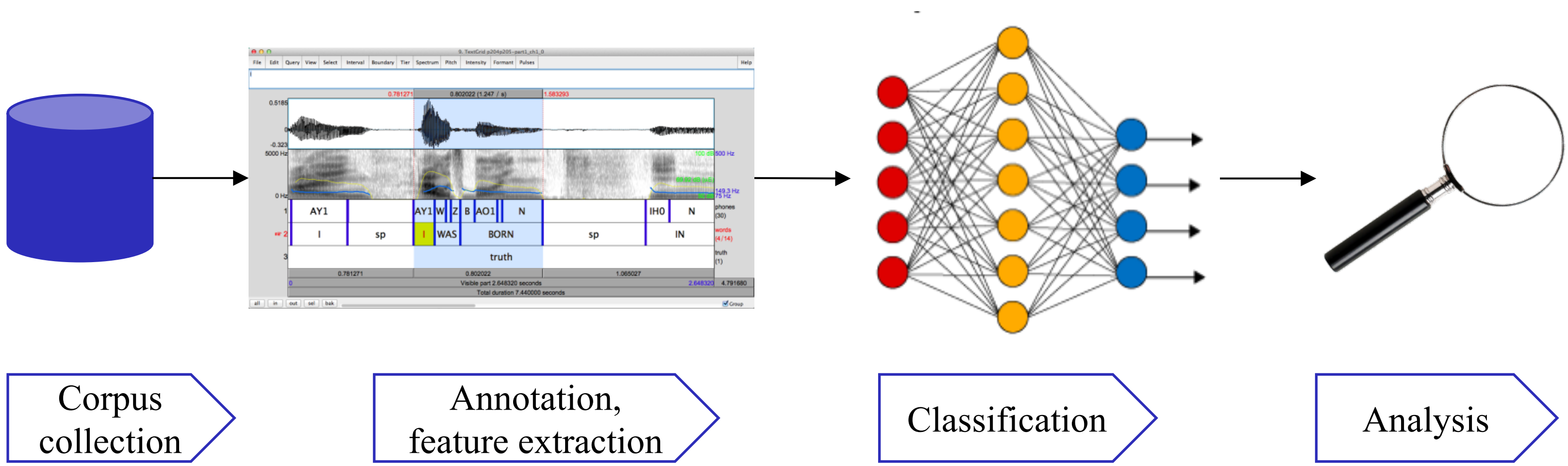
Our Approach

- Conduct ***objective, experimentally verified*** studies of spoken cues to deception ***on large corpora*** which predict better than ***humans*** or ***polygraphs***
- ***Our method:***
  - Collect speech data and extract ***acoustic, prosodic,*** and ***lexical cues*** automatically
  - Take ***gender, ethnicity, and personality factors*** into account as features in classification
  - Use ***Machine Learning*** techniques to train models to classify deceptive vs. non-deceptive speech

Questions We Hope to Answer

- Can we improve ***human deception detection***
  - By providing new knowledge and training materials
  - By providing classifiers to aid in deception detection
- Can we ***identify trust*** in humans – ***and mistrust***
- Can we ***control trust*** in machines: an ethical question…
  - When robots and avatars should be trusted
  - When they should not…

Deception Detection from Text and Speech



Outline

- ***Corpus collection***
- Classification of deception from text and speech
- Individual differences in deceptive behavior
- New goal: Acoustic-prosodic indicators of trust

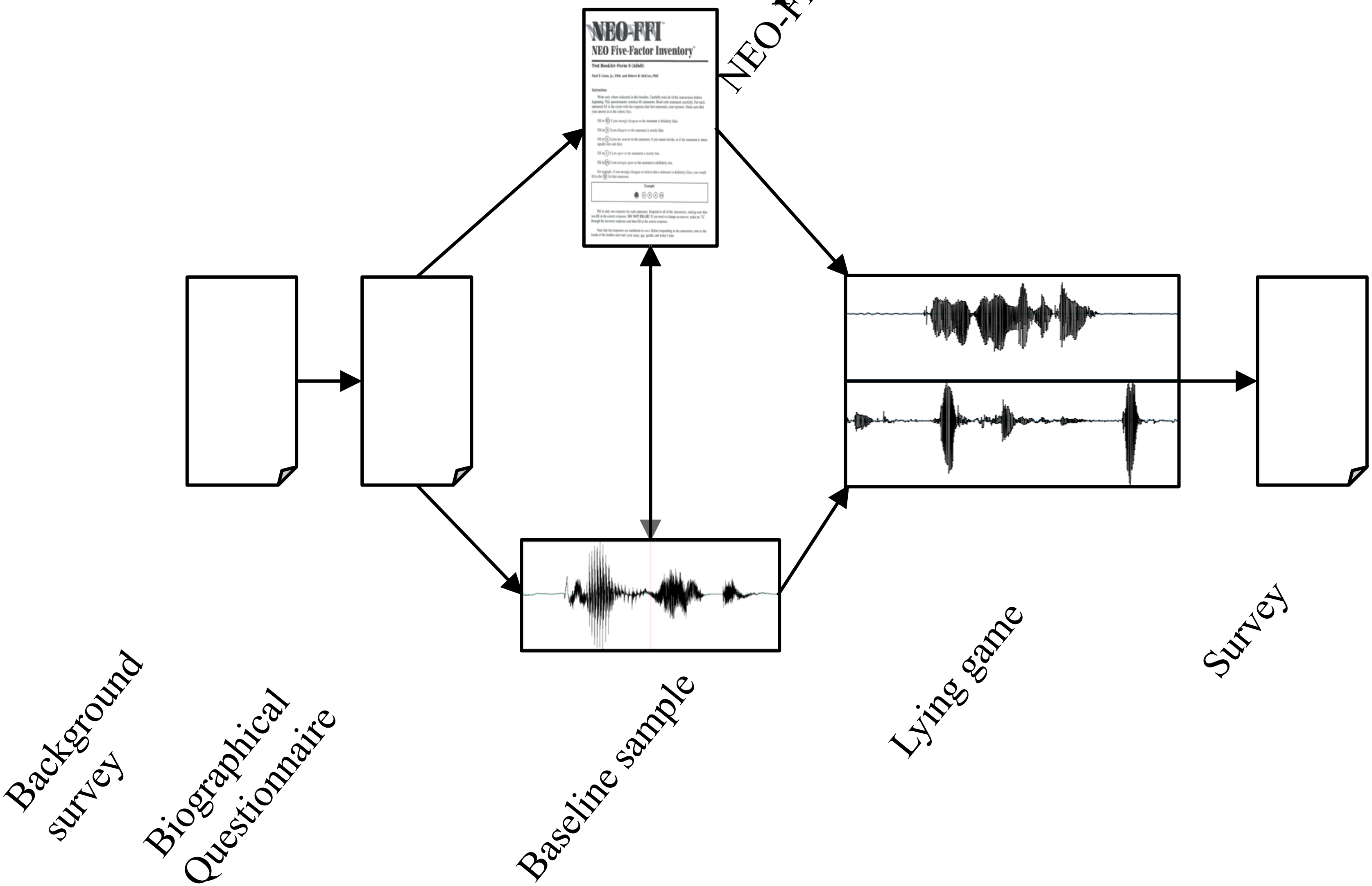
Columbia/SRI/Colorado Deception Corpus, 2003-5

- ***7h of speech*** from 32 SAE-speaking subjects performing tasks and asked to lie about half
- ***Lexical and acoustic-prosodic features*** identified from psycholinguistic literature for classification
- Results
  - Classification accuracy (~70%) ***significantly better than human performance*** on our corpus
  - Considerable ***individual differences*** between speakers: Judges with certain ***personality traits*** performed better than our classifiers

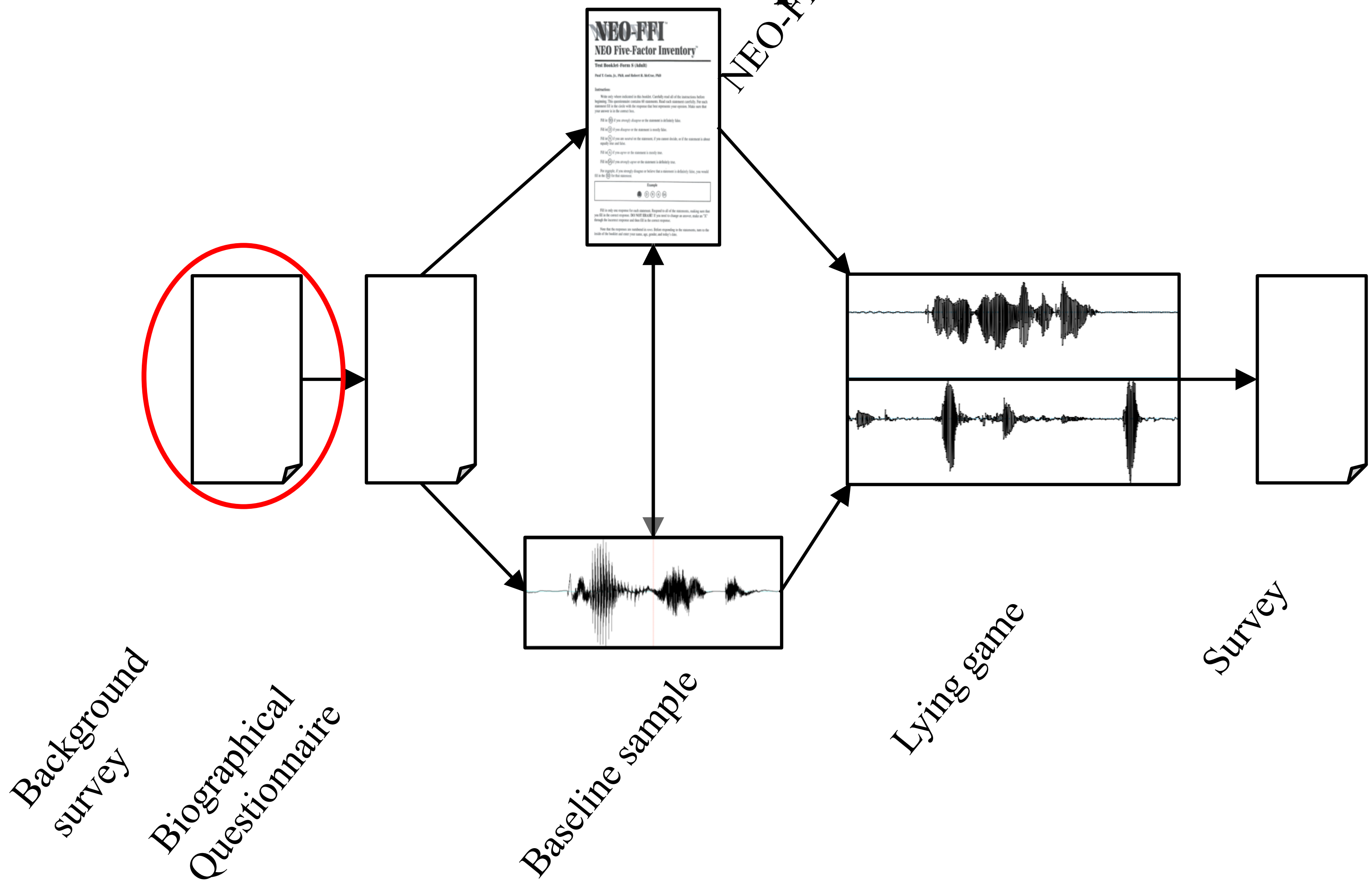
Columbia X-cultural Deception Corpus (2011--)

- New questions to ask:
  - Can ***personality factors*** help in predicting individual differences in deception?
  - Can ***people who detect lies better also lie more successfully***?
  - Do ***differences in gender and native language*** influence deceptive behavior? Judgment of deception?
- *New study*: Pair native speakers of Standard American English with Mandarin Chinese speakers, speaking English, interviewing each other

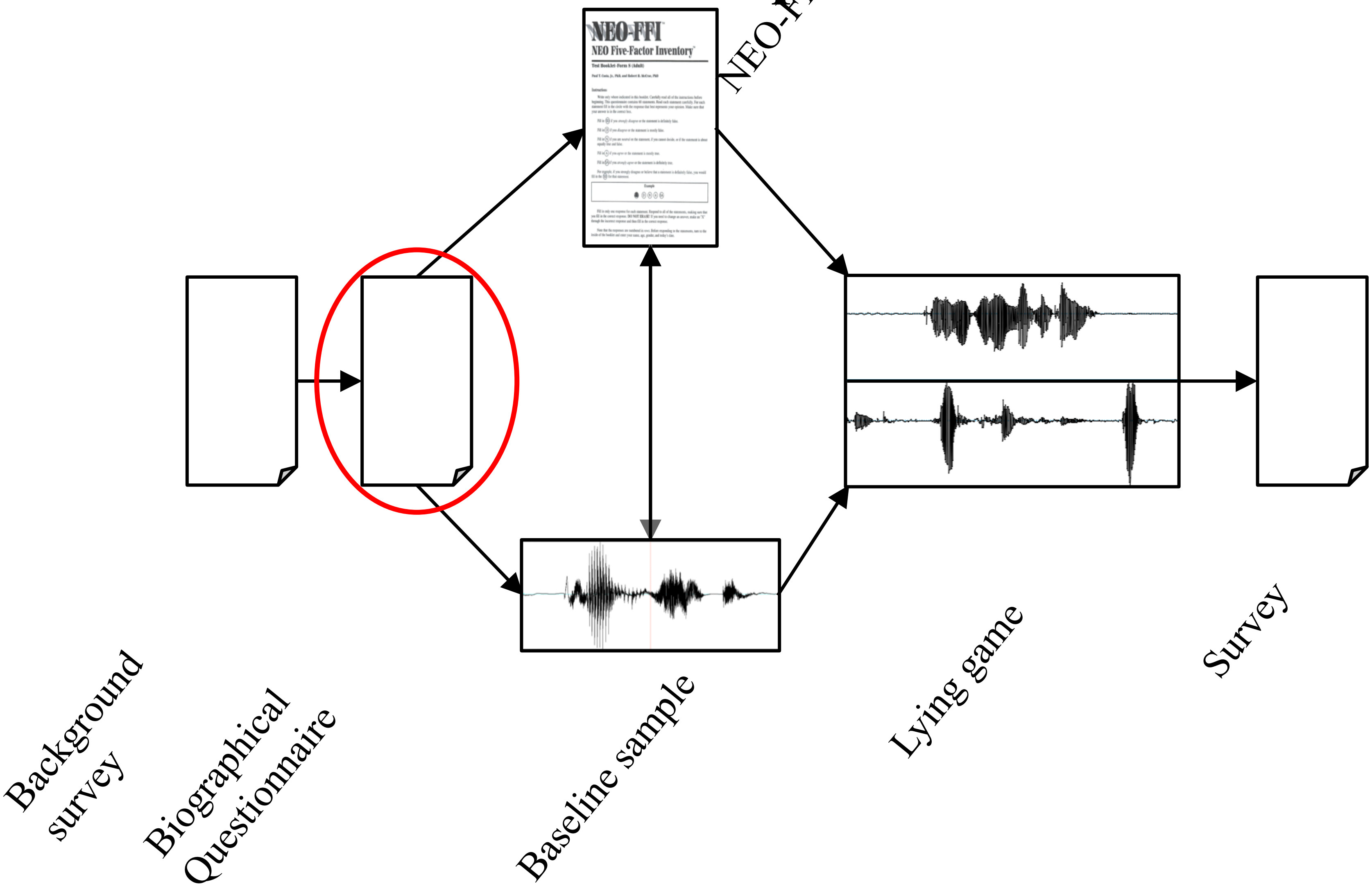
Our CXD Experiment



Our CXD Experiment



Our CXD Experiment



BiographicalQuestionnaire1.pdf.odt - LibreOffice Writer

File Edit View Insert Format Table Tools Window Help

Default FreeSans 11

Biographical Questionnaire

Participant No. _____ Date ___________

Instructions

Please carefully look through the questions. Write down the true answer to each question in the "True Answer" column. When you have finished that, for all the questions that have don't have "X"s in the "False Answer" column, make up an answer. Consult the additional sheet you have been given. You want to choose a lie that you are not as familiar with as the true answer.

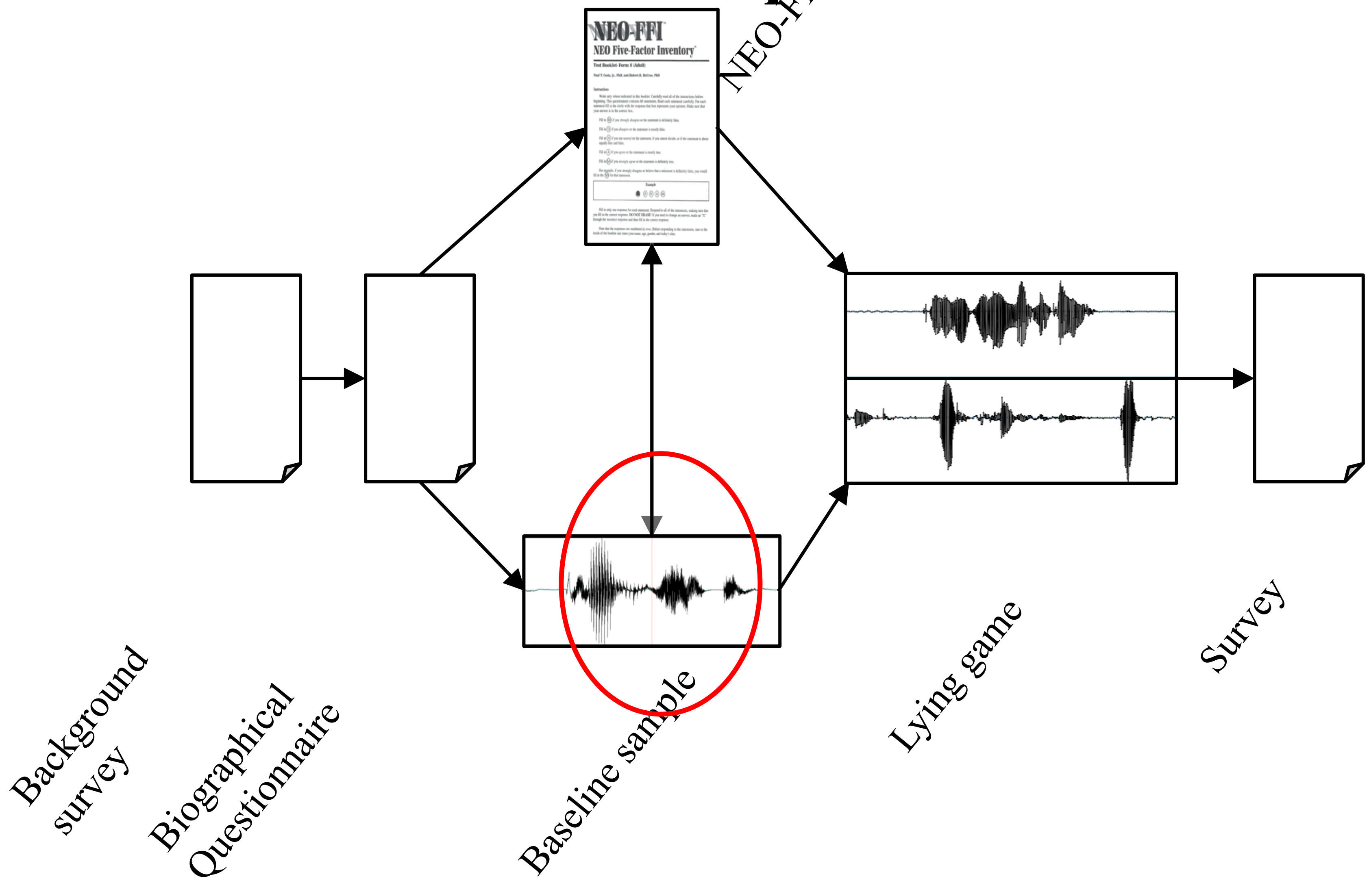
This experiment is completely anonymous- your name will never be linked to the data.

No.	Questions	True Answer	False Answer
1	Where were you born?		
2	How many years did you live in your first home?		
3	What is your mother's job?		
4	What is your father's job?		
5	Have your parents divorced?		
6	Have you ever broken a bone?		
7	Do you have allergies to any foods?		
8	Have you ever stayed overnight in a hospital as a patient?		
9	Have you ever tweeted? (posted a message on twitter)		
10	Have you ever bought anything on eBay?		
11	Do you own an e-reader of any kind?		
12	Who was the last person you were in a physical fight with?		
13	Have you ever gotten into trouble with the police?		
14	Who ended your last romantic relationship?		
15	Whom do you love more, your mother or father?		
16	What is the most you have ever spent on a pair of shoes?		
17	What is the last movie you saw that you really hated?		

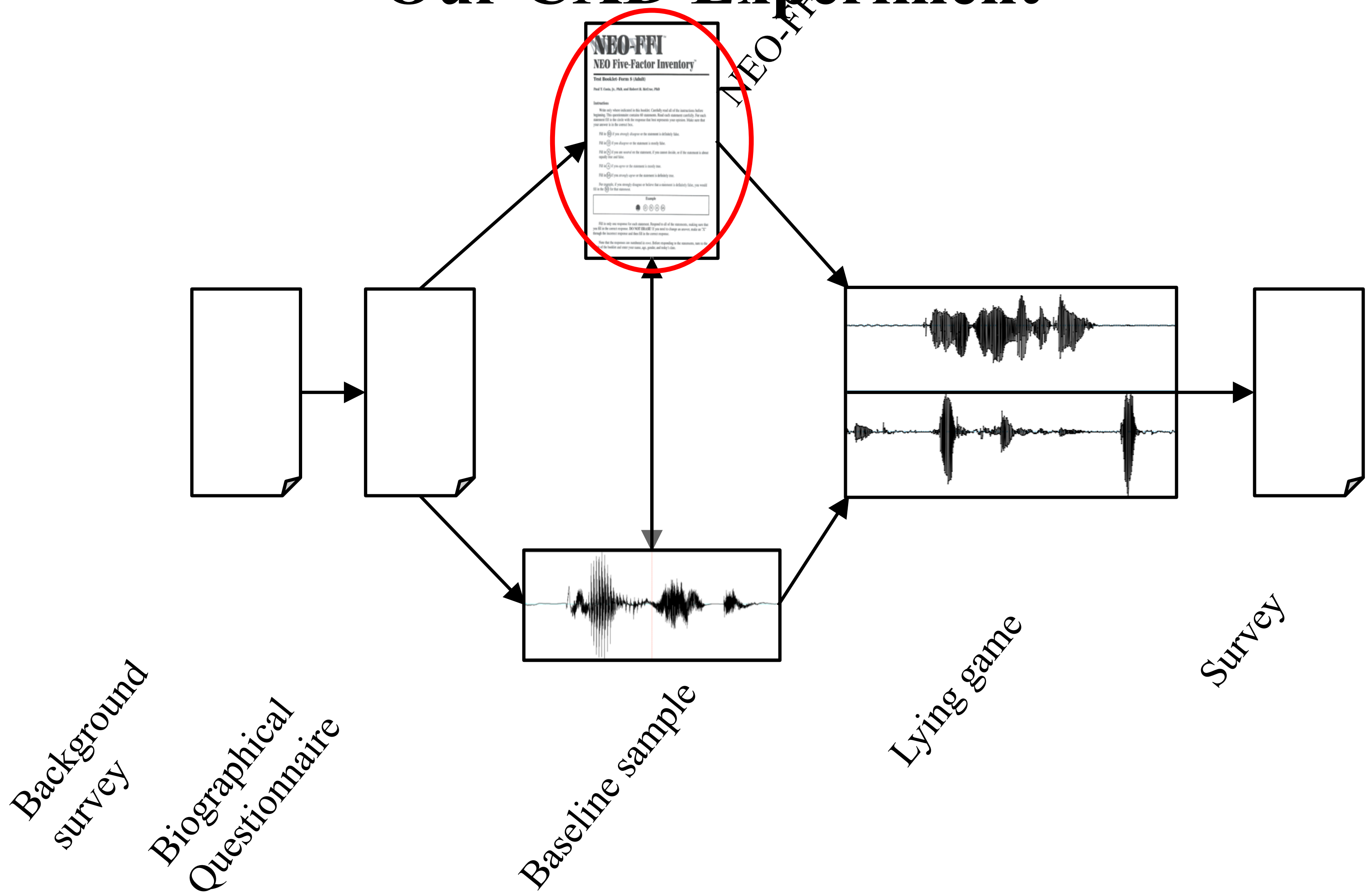
Find

Page 1 / 1 Default English (USA) INSRT STD

Our CxD Experiment



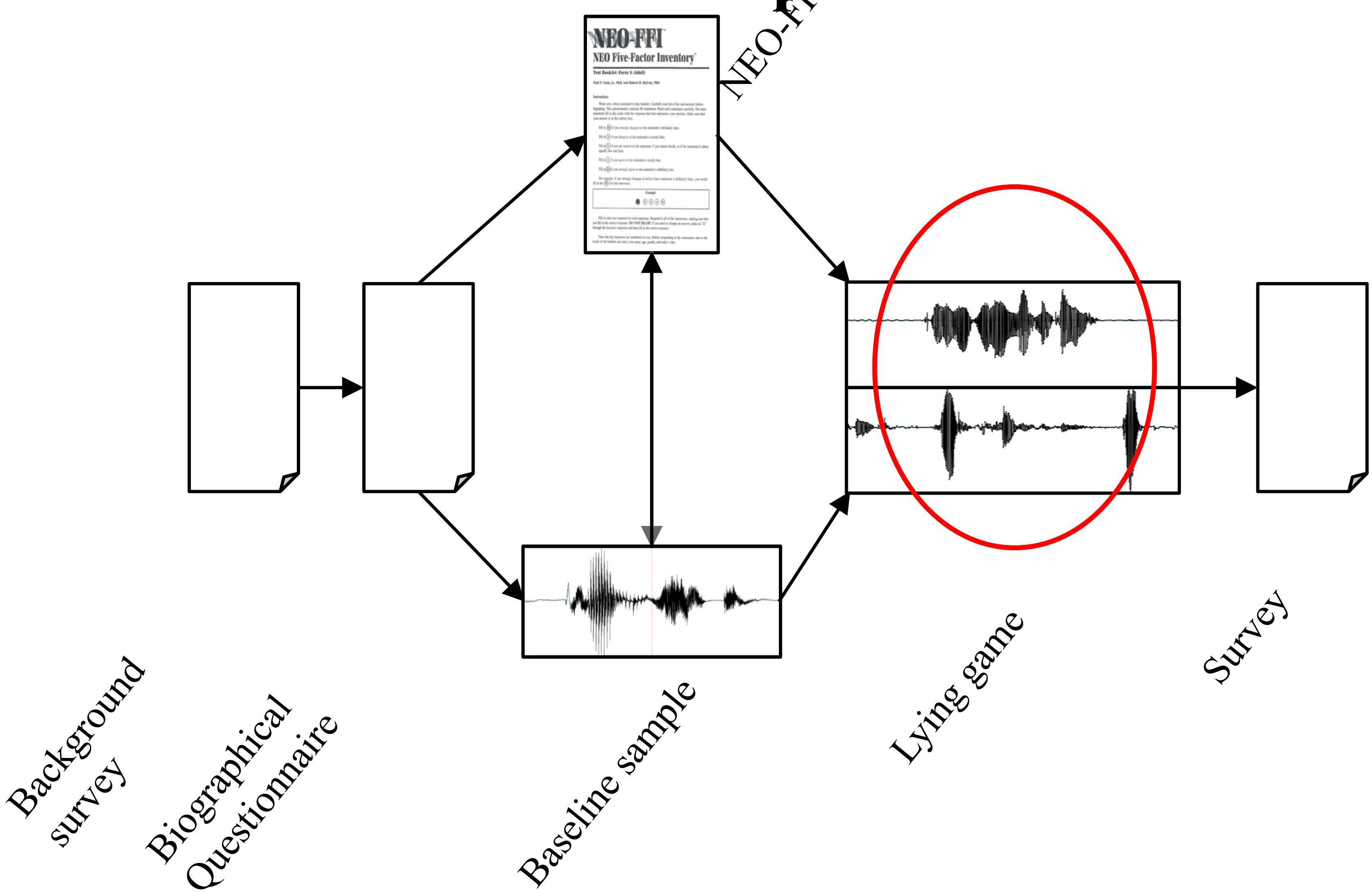
Our CXD Experiment



The Big Five NEO-FFI (Costa & McCrae, 1992)

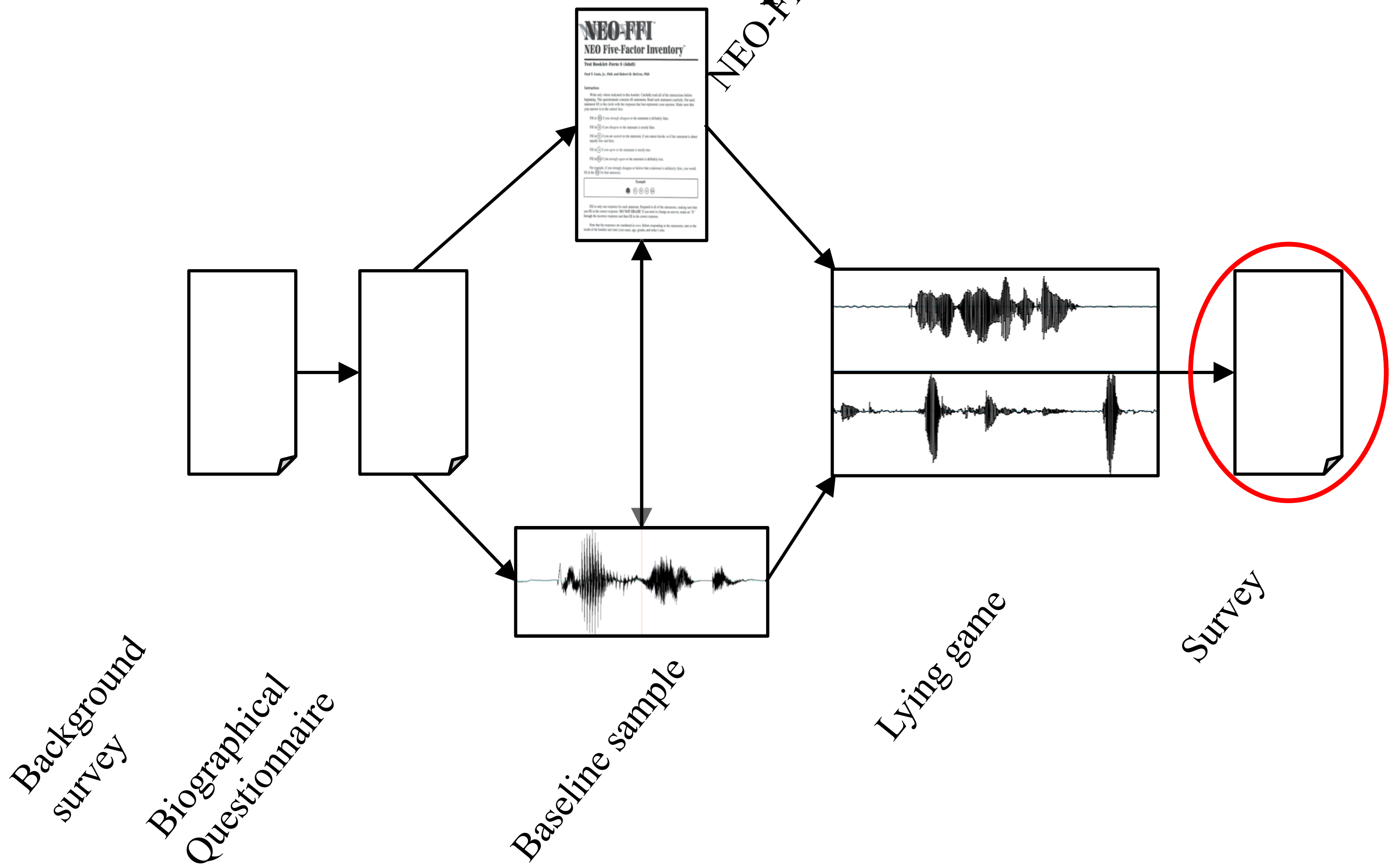
- **Openness to Experience:** "I have a lot of intellectual curiosity."
- **Conscientiousness:** "I strive for excellence in everything I do."
- **Extraversion:** "I like to have a lot of people around me."
- **Neuroticism:** "I often feel inferior to others."
- **Agreeableness:** "I would rather cooperate with others than compete with them."

Our CXD Experiment



Our CXD Experiment

Our CXD Experiment



Motivation and Scoring

- ***Monetary motivation***
  - **Success for interviewer**:
    - Add $1 for every correct judgment, truth or lie
    - Lose $1 for every incorrect judgement
  - **Success for interviewee:**
    - Add $1 for every lie interviewer thinks is true
    - Lose $1 for every lie interviewers thinks is a lie
- ***Good liars tell the truth as much as possible*** when lying, so how do we know what's true or false for follow-up questions?
  - **Interviewees press T/F keys after every phrase**

Columbia X-Cultural Deception (CXD) Corpus

- ***340 subjects, balanced by gender and native language (American English, Mandarin Chinese)*:122 hours of speech**
- ***Crowdsourced transcription***, speech alignment
  - TF keypress alignment
- Segmented into
  - **Inter-pausal units (IPUs)**
  - **Speaker turns**
  - **Question/answer sequences** (Q/A and Q/A+ follow-up)

“Where were you born?”

True or False?

"Where were you born?"

“What is the most you have ever spent on a pair of shoes?”

True or False?

“What is the most you have ever spent on a pair of shoes?”

Outline

- Corpus collection
- ***Classification of deception from text and speech***
- Individual differences in deceptive behavior
- Acoustic-prosodic indicators of trust

Features We Extract

- ***Currently:***
  - **Text-based**: n-grams, psycholinguistic, Linguistic Inquiry and Word Count (LIWC) (Pennybaker et al), word embeddings (GloVe trained on 2B tweets)
  - **Speech-based**: openSMILE IS09 (386)
  - **Gender**, **native language**
  - **Five personality** dimensions (NEO-FFI)
- ***Next***: **Syntactic features** (complexity) and all combined

Corpus Segmentation

- Inter-pausal unit (**IPU**)
- **Turn**
- **Question-level** (first answer, first+follow-up answers)

Unit	**Interviewee**	**Interviewer**	**Total**	**Avg. length** (sec)		**Avg. # wo**rds	
IPU	111,479	81,536	193,015	1.4		4.5	
Turn	43,706	41,753	85,459	4.7		12.4	
Q-level	7,418	7,418	14,836	one 3.2	chnk 20.9	one 7.9	chnk 56.5

Machine Learning Experiments

- What are the best classification models?
- What are the optimal segmentation units?
- Which feature sets are most useful?

Machine Learning Experiments

- ***What are the best classification models?***
  - **Statistical machine learning – Logistic Regression, SVM, Random Forest**
  - **Neural networks – DNN, LSTM, hybrid system**
- What are the optimal segmentation units?
- Which feature sets are most useful?

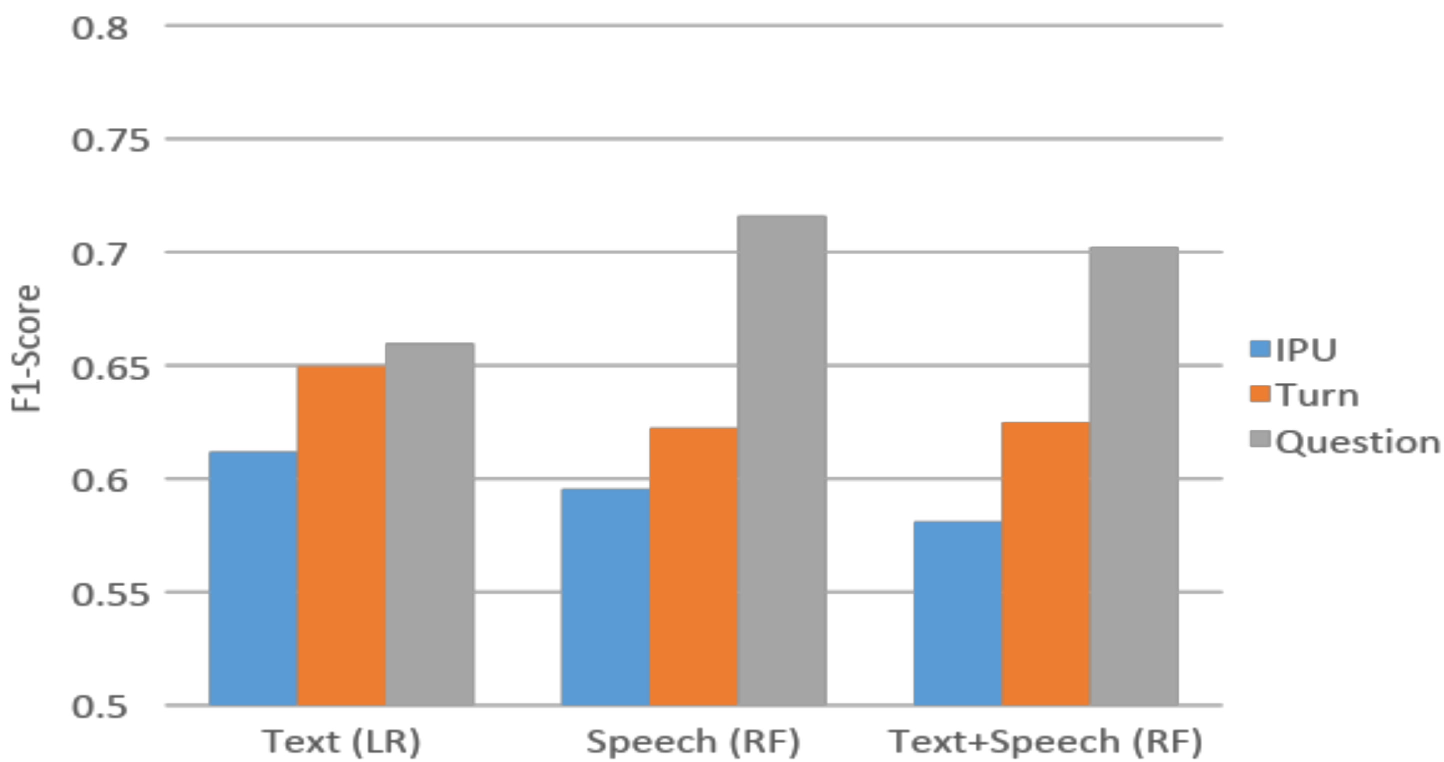
Machine Learning Experiments

- What are the best classification models?
- ***What are the optimal segmentation units?***
  - **Inter-Pausal Unit (IPU), speaker turns, Q/*A* , Q/*A+follow-up As***
- Which feature sets are most useful?

Machine Learning Experiments

- What are the best classification models?
- What are the optimal segmentation units?
- ***Which feature types are most useful?***
  - **Text-based: n-grams, psycholinguistic-based, LIWC, GloVe word embeddings trained on 2B tweets**
  - **Individual difference features: gender, native language, personality**
  - **Speech-based: openSMILE IS09 acoustic-prosodic features (e.g. f0, intensity, speaking rate, VQ)**

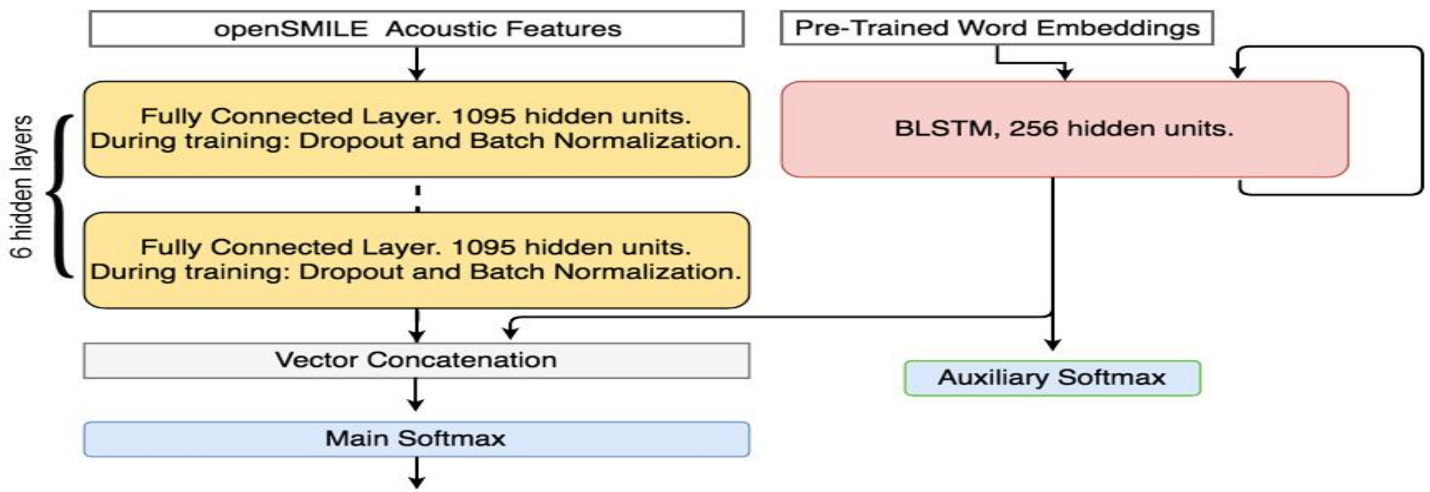
Segmentation – Longer is Better: IPU, Turn, Question



Mendels, Levitan et al. 2017, "Hybrid acoustic lexical deep learning approach for deception detection "

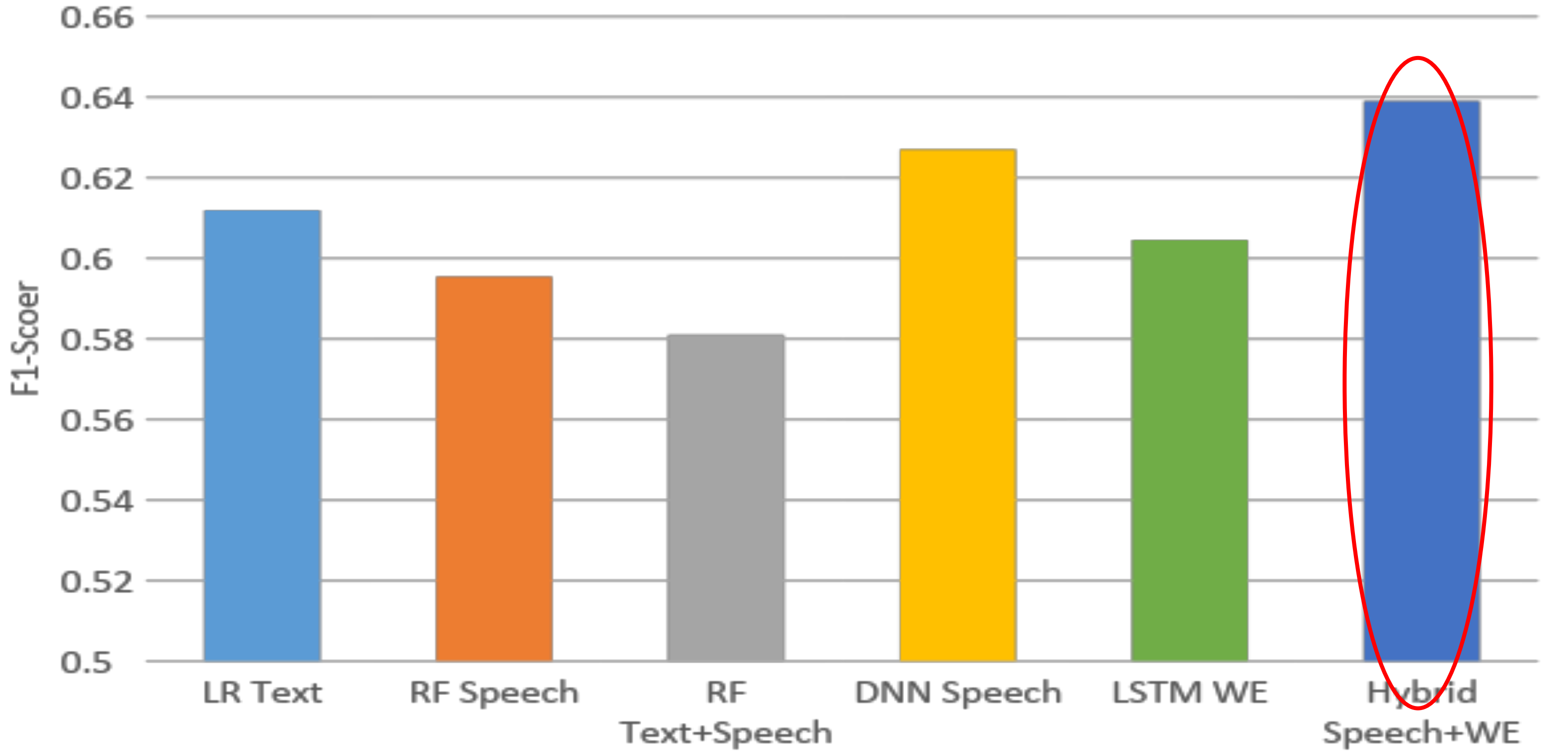
Deep Learning Approaches

- BLSTM-lexical
- DNN-openSMILE
- Hybrid: BLSTM-lexical + DNN-openSMILE



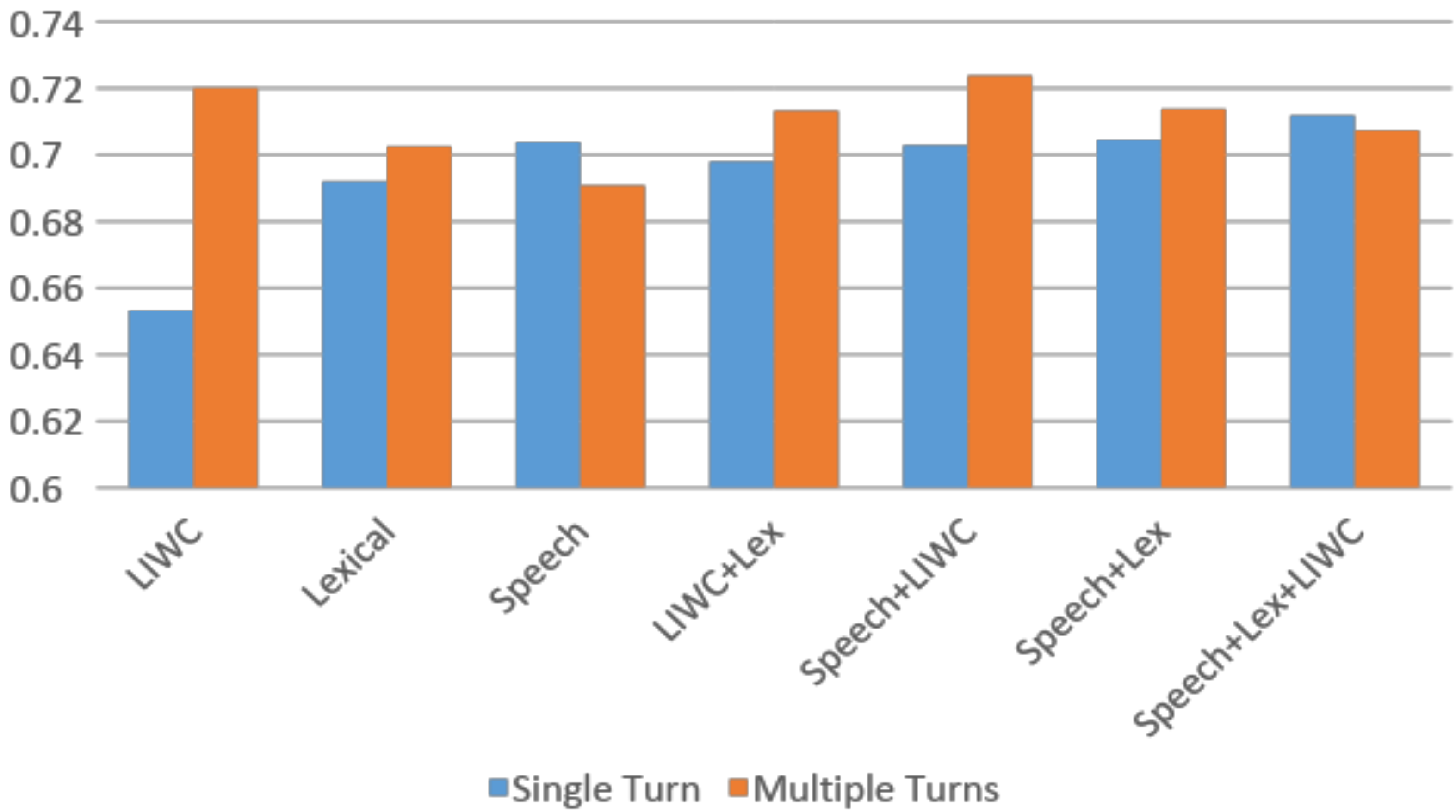
Mendels, Levitan et al. 2017, "Hybrid acoustic lexical deep learning approach for deception detection "

IPU Classification: Hybrid is Better



Mendels, Levitan et al. 2017, “Hybrid acoustic lexical deep learning approach for deception detection ”

Question-Level Random Forest; Context is Better



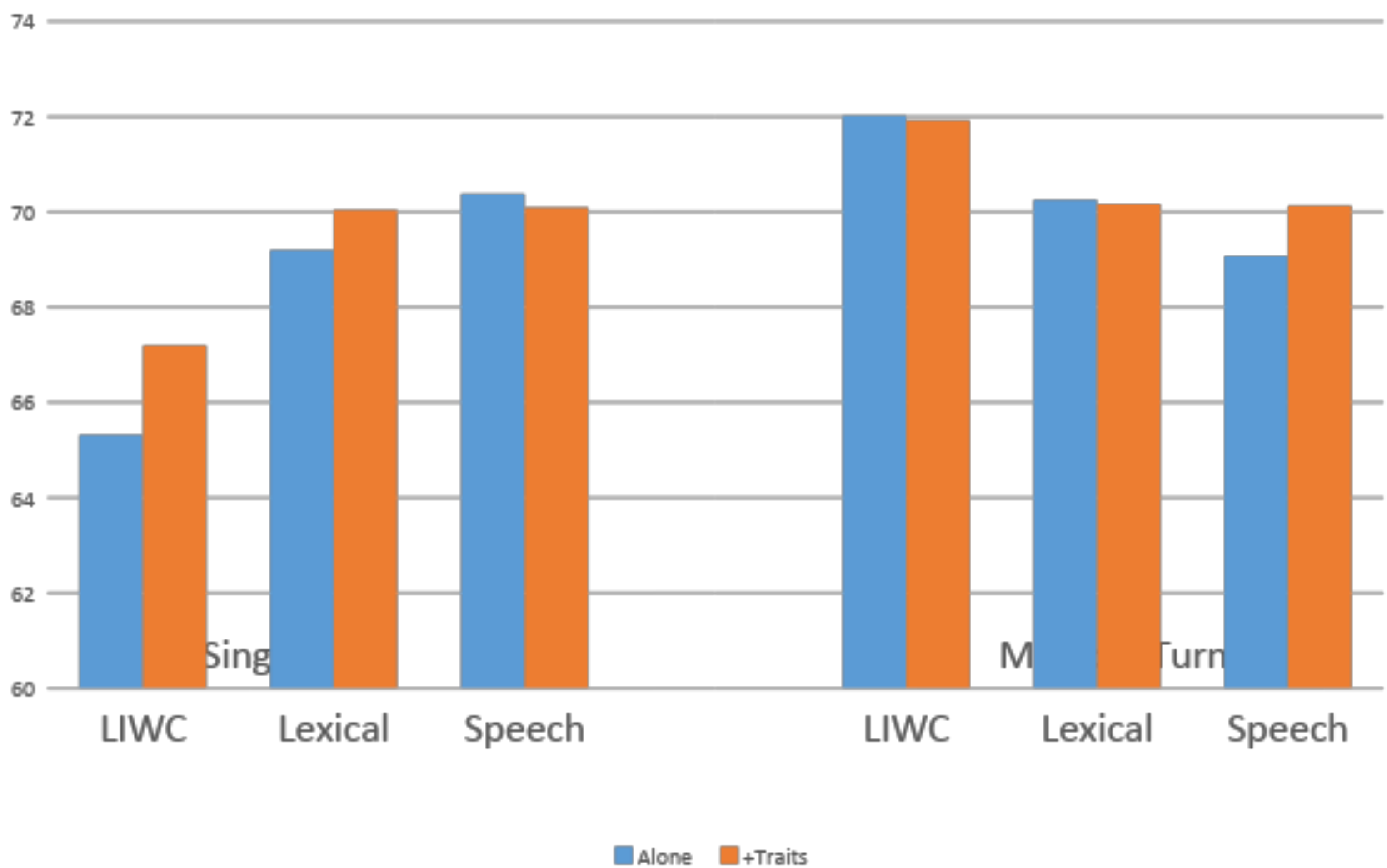
Analysis – Acoustic Features

Feature	t	p	sig
Duration	-0.63	0.53	
Pitch Max	4.37	1.28E-05	*
Pitch Mean	0.56	0.58	
Intensity Max	3.45	0.0006	*
Intensity Mean	1.33	0.18	
Speaking Rate	-1.69	0.09	
Jitter (f0 var.)	-1.31	0.19	
Shimmer (Ampl var.)	-1.39	0.17	
NHR	0.35	0.73	

Analysis – Lexical Features

Truth	Lies	Neutral
Negation	Clout	Laughter
Function words	Informal	Comparison
Certain	Word count	Anger
Cognitive processes	Words per second	Power
	Past tense	Present tense
	Specificity	Future tense
	Hedges	Complexity
	Imagery	
	3rd person pronouns	

Classification with Gender and Native Language: 1st Answer and Chunks: Personal Info Helps with Less Context



Results for Classification

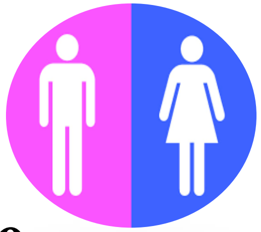
- ***Deception classification experiments***
  - **>.72 F1-Score when using question segmentation to predict li**es
  - **Note that human interviewers' F1 is 0.43**
  - **Deep learning** approaches for IPU segmentation probably most promising route in future
    - **Many more features** to examine together

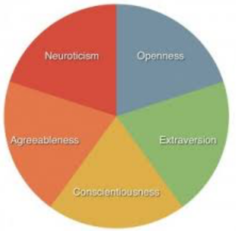
Outline

- Corpus collection
- Classification of deception from text and speech
- ***Individual differences in deceptive behavior and detection of deception***
- Acoustic-prosodic indicators of trust

Some Individual Differences

- ***Extroversion*** is correlated with success at deception, for **English male** speakers
- ***Native English speakers*** perform better at deception when paired with a **native Chinese** interviewer



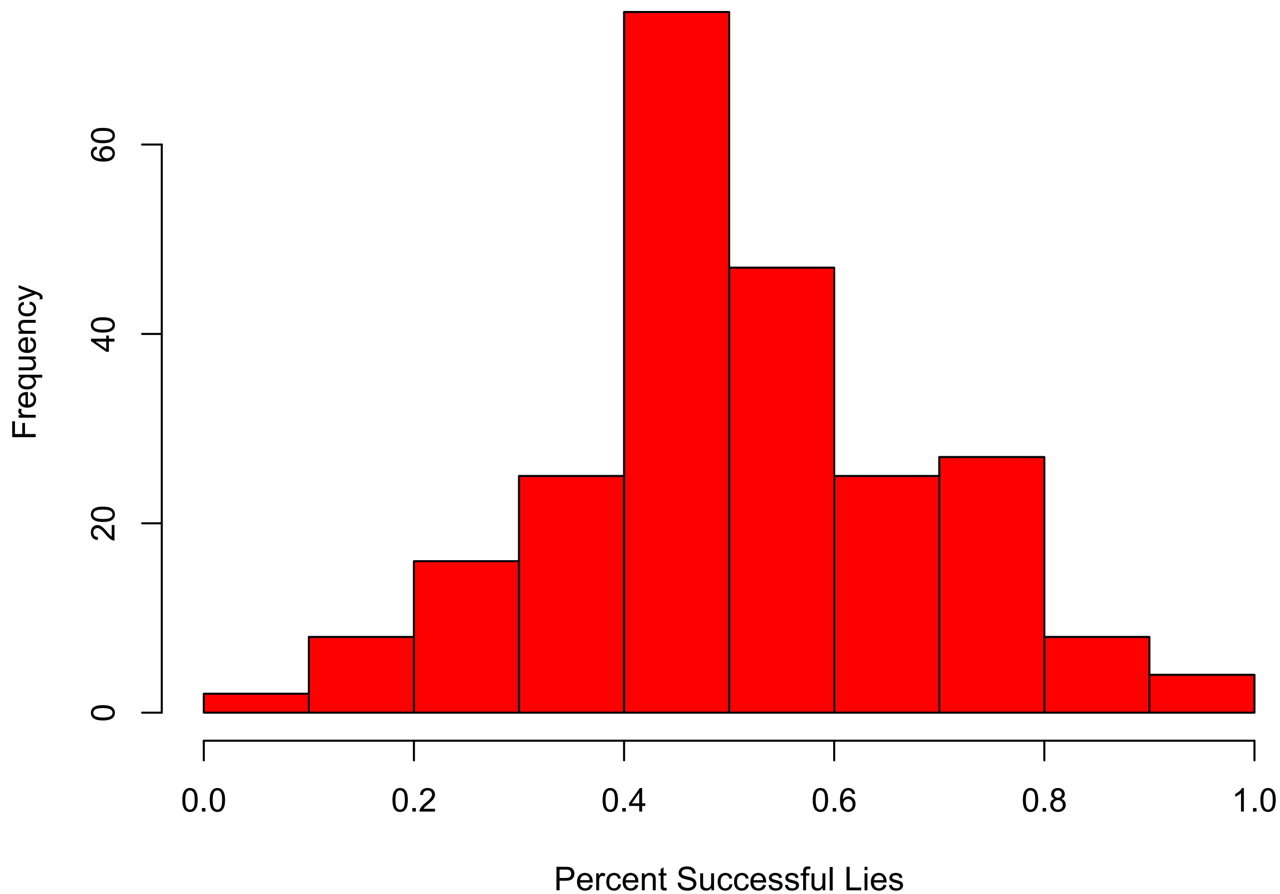


Individual Differences in Deception and Truth-telling

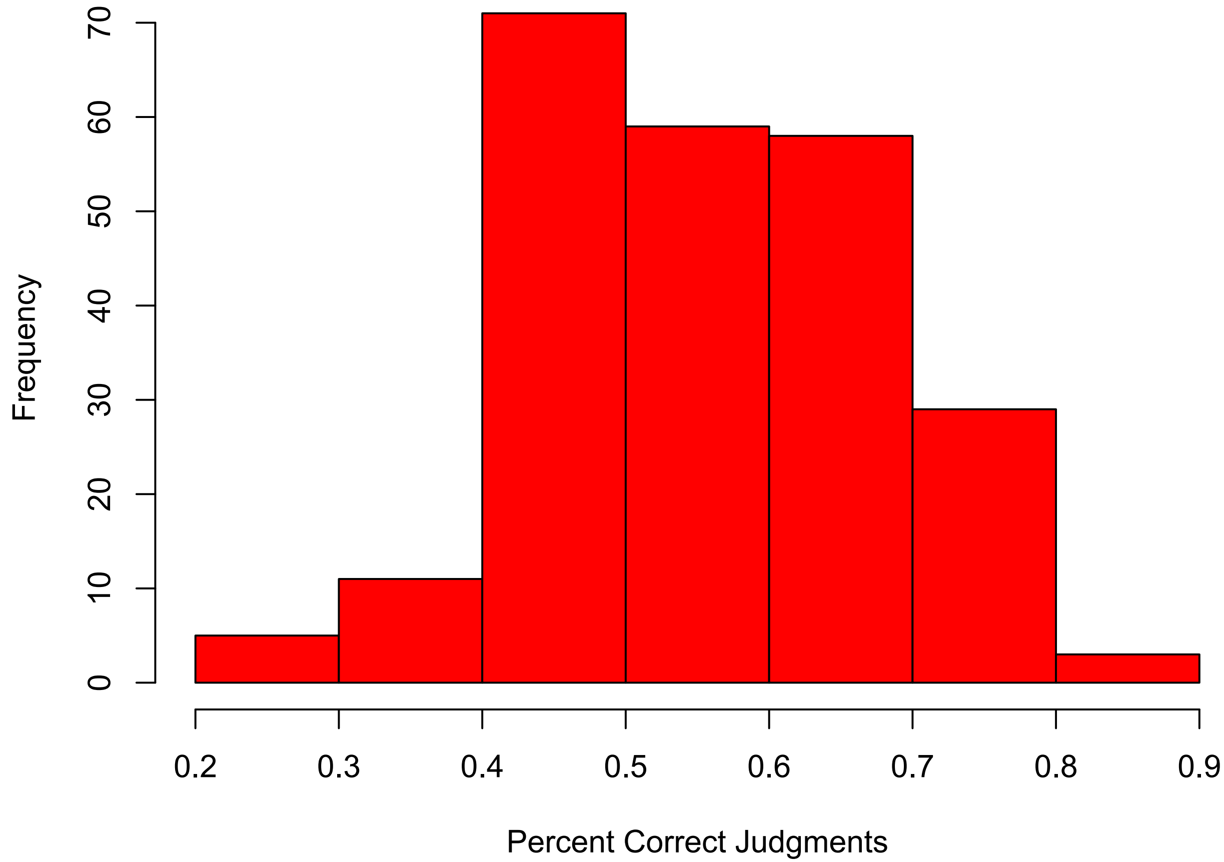
Male	*Female*	*English*	*Chinese*
Positive emotion (T)	**Jitter** (T)	**Intensity mean** (F)	**Speaking rate** (T)
Interrogatives (F)	**Perceptual process** (F)	**Swear** (F)	**Certainty** (T)
	Future tense (F)		**Feel** (F)
			Causation (F)

Levitan et al. 2018, "Linguistic indicators of deception and perceived deception in spoken dialogue"

Differences in Deceptive Ability: How well did interview*ees* lie?



Differences in Deception Detection: How well did interview*ers* judge deception?



Results

- There are **gender and cultural/native language differences in deceptive behavior**
- There are **differences in ability to deceive and in ability to detect deception**
- ***Understanding these differences can improve deception classification by machines and perhaps by humans…***

Outline

- Corpus collection
- Classification of deception from text and speech
- Individual differences in deceptive behavior
- ***Acoustic-prosodic indicators of trust***

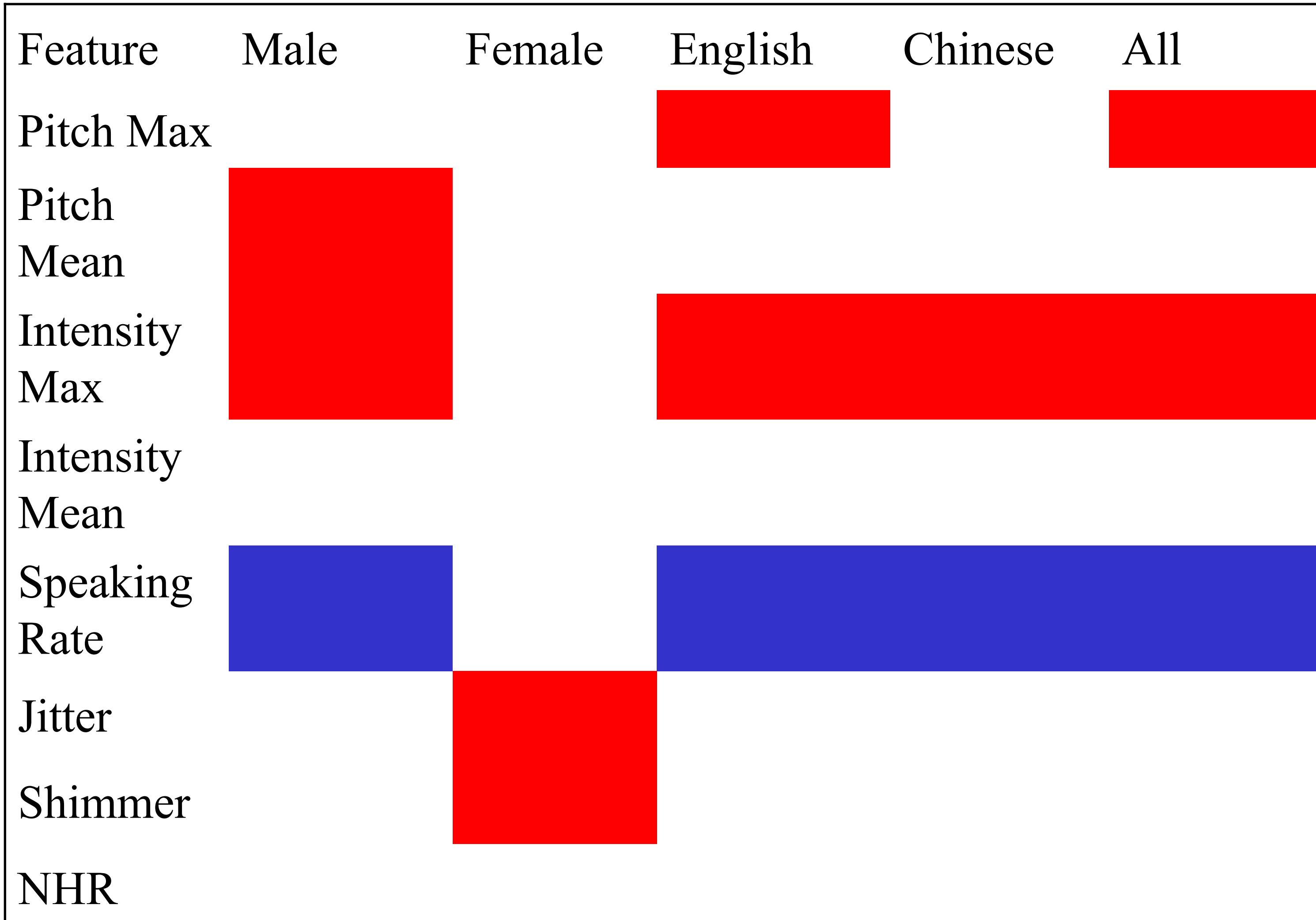
Features of Perceived Lies and Truth

Feature	t	p	sig
Pitch Max	2.35	0.02	*
Pitch Mean	1.65	0.1	
Intensity Max	2.625	0.009	*
Intensity Mean	-0.785	0.43	
Speaking Rate	-3.785	0.0002	*
Jitter	-1.815	0.07	
Shimmer	-1.905	0.06	
NHR	0.58	0.56	

Group-Specific "Trust" Indicators: Speakers: Native Language Matters

s	Male	Female	English	Chinese	All
Pitch Max					
Pitch Mean					
Intensity Max					
Intensity Mean					
Speaking Rate					
Jitter					
Shimmer					
NHR					

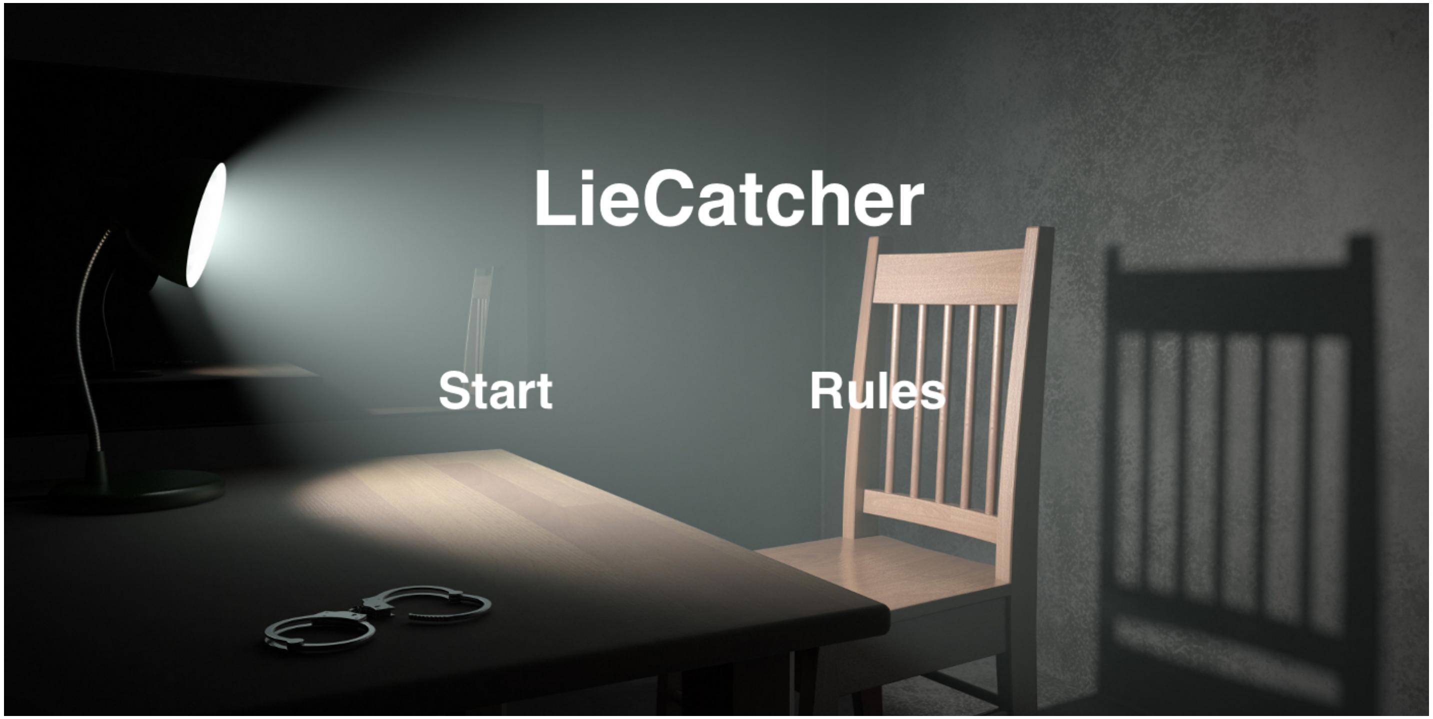
Group-specific "Trust" Indicators for Interview*ers*: Gender Matters



What's Next?

- ***Classifiers*** to detect trustworthy voices and ***TTS systems*** to create them
- Even ***better deception classifiers***
- ***Tools to train humans*** in deception detection
- A ***fun game***…

Games with a Purpose



Levitan et al. 2018, “LieCatcher: Game framework for collecting human judgments of deceptive speech”

“Did you ever cheat on a test in high school?”

TRUE or FALSE?

"Did you ever cheat on a test in high school?"

"Did you ever cheat on a test in high school?"

TRUE or FALSE?

"Did you ever cheat on a test in high school?"

Thank you!