

# Automating Assessment of Therapist Fidelity: Attainable Goal or Holy Grail?

Dave Atkins, PhD  
Grin Trusz, BA  
Chris Dunn, PhD

Psychiatry and Behavioral Science  
University of Washington

datkins@uw.edu

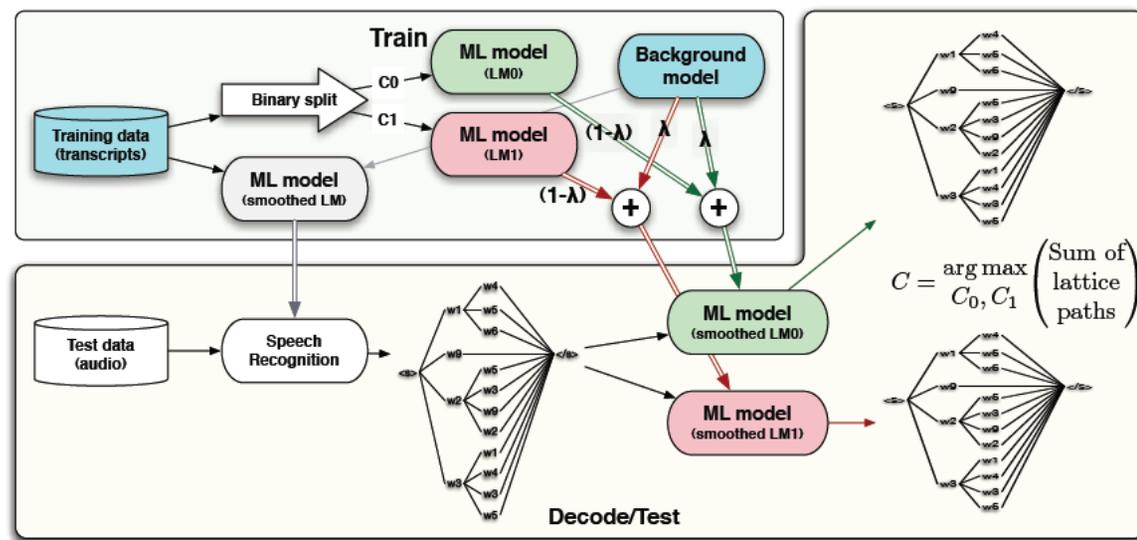


Fig. 3. Overview of the classification process without human transcripts through the use of ASR lattices.

# Overview

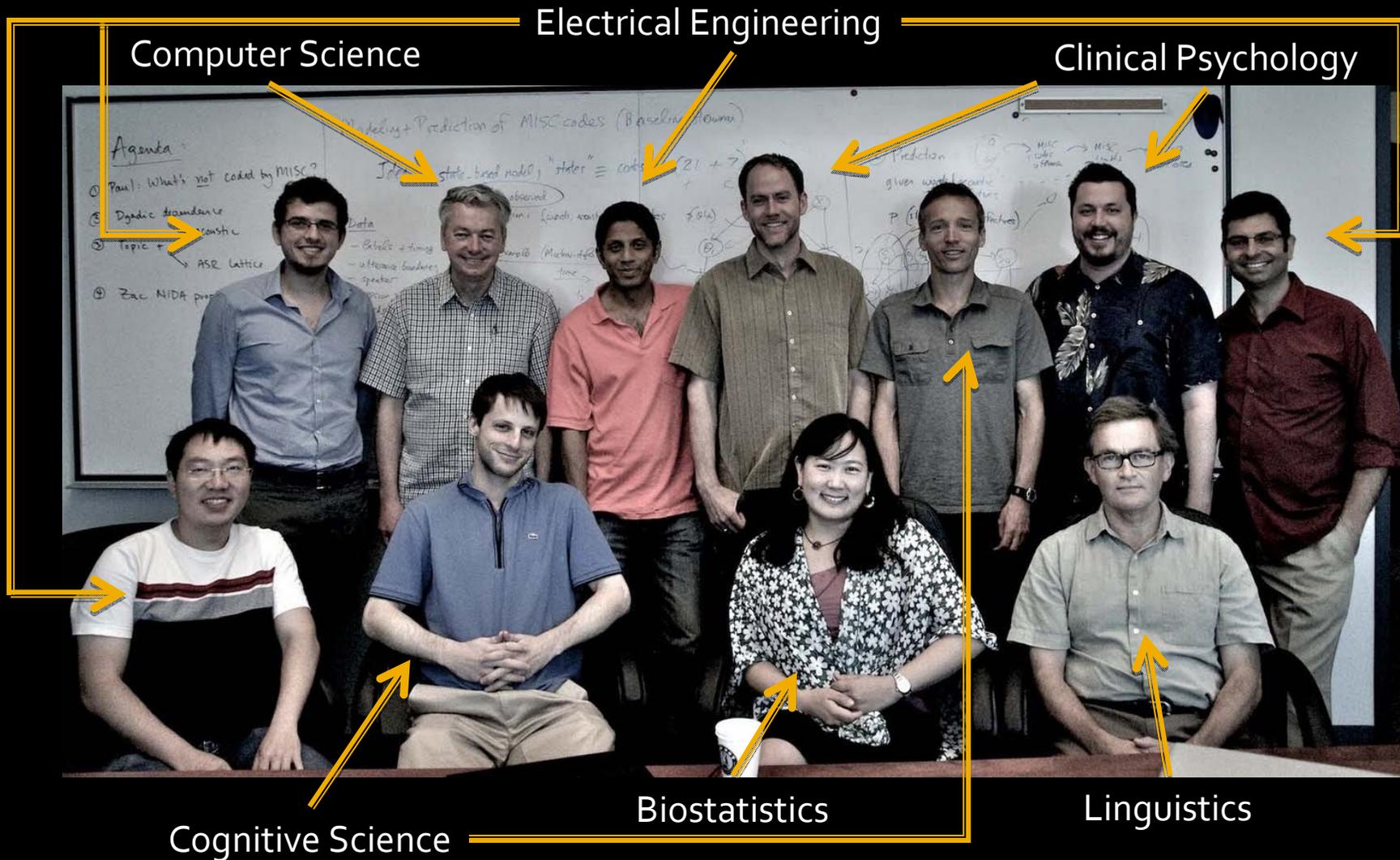
1. Provider fidelity to behavioral interventions as challenge to implementation

2. Introduce interdisciplinary work on text-mining and speech signal processing toward automating provider fidelity

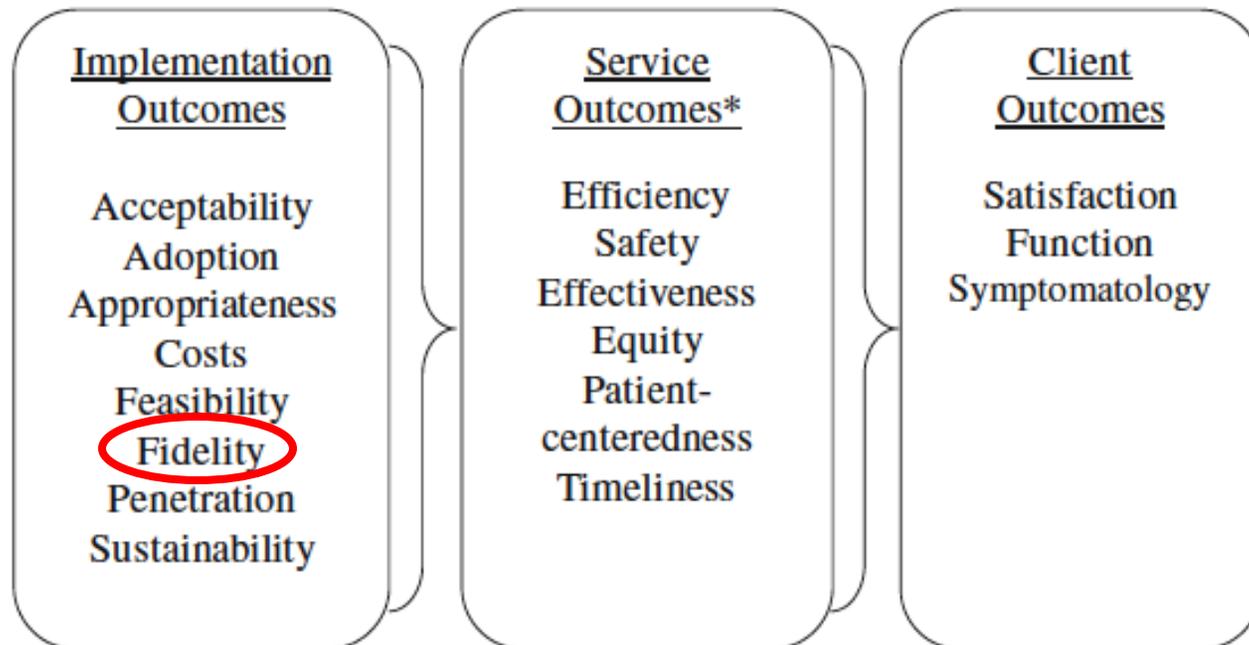
3. Discuss current status and future directions

# Interdisciplinary collaborators

Funding: NIAAA R01 AA018673; NIMH R01 MH056223-S1



# Therapist fidelity is key implementation outcome for behavioral interventions



\*IOM Standards of Care

**Fig. 1** Types of outcomes in implementation research

Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health*. Mar 2011;38(2):65-76.

# Assessing therapist fidelity presents unique challenges

What “data” do we need to collect to assess therapist fidelity?

Who has access to this data?

Gold-standard:

Observational coding of therapy

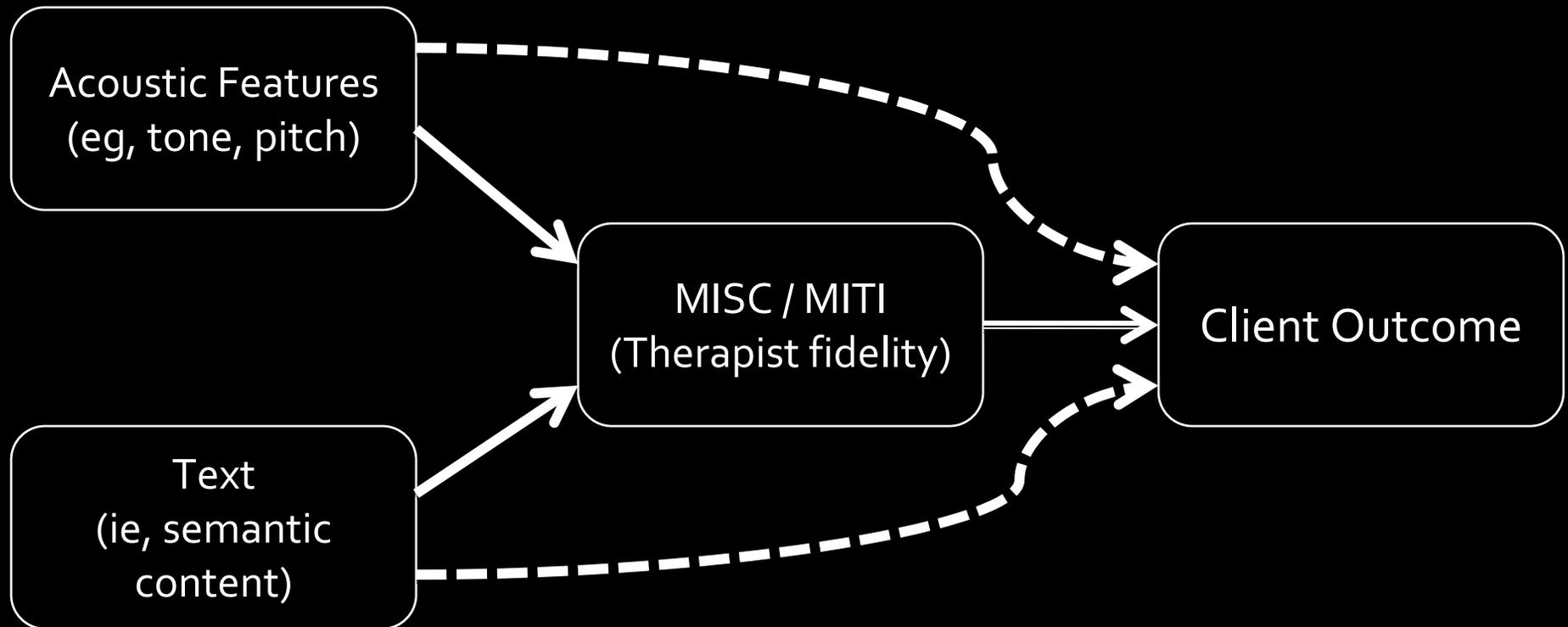
Proctor et al (2011): “Measures used in efficacy research will likely prove too cumbersome for real-world studies of implementation. ...detailed assessment of fidelity through coding of [tapes] would be too time-intensive for a multi-agency study assessing fidelity of treatment implementation.”

# Let's take a closer look at what we mean by therapist fidelity

Two minute section from a training tape of Motivational Interviewing (MI) for individuals with drug problems presenting in a primary care clinic.

(NIDA-funded, BI Study, Roy-Byrne, PI)

# General model: Two linguistic inputs leading to fidelity ratings and outcome



# Underlying semantic dimensions: Topic model example with couple tx

## CONTENT-BASED TOPICS

Family		Finances		Relationships		Sex		Transportation		Work	
Topic 64	.011	Topic 23	.011	Topic 50	.014	Topic 63	.010	Topic 1	.008	Topic 91	.010
mom	.096	money	.104	married	.080	sex	.131	car	.167	job	.197
mother	.060	dollars	.054	together	.053	sexual	.068	drive	.118	work	.139
dad	.056	buy	.050	relationship	.045	love	.026	park	.026	career	.027
sister	.040	hundred	.043	date	.044	part	.021	down	.018	money	.024
brother	.030	card	.031	live	.036	interesting	.018	street	.017	day	.022
call	.028	thousand	.021	met	.034	initially	.017	turn	.016	company	.019
day	.027	bought	.018	move	.028	touch	.016	traffic	.013	support	.019
father	.023	credit	.017	attracted	.024	back	.016	accident	.013	people	.018
deal	.018	pay	.015	remember	.023	desire	.016	home	.013	situation	.014
live	.014	fifty	.015	marriage	.022	physical	.015	bus	.012	hours	.013
parents	.014	cost	.014	thought	.018	intimacy	.014	drove	.010	week	.013
down	.014	give	.014	months	.017	life	.013	hours	.010	positive	.012
care	.013	car	.012	two	.016	talk	.013	direction	.010	happy	.011
stuff	.013	expensive	.012	decided	.015	relationship	.011	freeway	.009	business	.011
house	.012	five	.012	pretty	.015	pleasure	.010	walk	.008	interview	.010

## EMOTION-BASED TOPICS

Negative Emotional Content				Positive Emotional Content							
Topic 3	.009	Topic 46	.008	Topic 79	.010	Topic 14	.010	Topic 95	.007	Topic 10	.011
angry	.170	give	.023	upset	.113	good	.039	[laugh]	.383	good	.101
anger	.072	shit	.021	back	.030	thought	.036	guess	.031	nice	.078
hurt	.047	pissed	.021	mad	.030	[laughing]	.031	good	.017	thought	.070
frustrated	.037	point	.019	temper	.024	pretty	.029	work	.017	felt	.048
trying	.032	whatever	.019	talk	.021	people	.025	thank	.012	appreciate	.033
upset	.027	man	.018	crying	.020	talk	.020	give	.012	week	.031
mad	.020	fuck	.017	angry	.018	part	.019	wow	.012	remember	.026
point	.018	care	.016	sorry	.018	summer	.017	definitely	.012	day	.026
sad	.016	god	.016	understand	.017	enjoy	.0				
emotional	.016	black	.015	fact	.015	remember	.0				
part	.016	problem	.015	late	.015	fun	.0				
felt	.014	fine	.015	apologize	.013	u	.0				
whatever	.014	walk	.014	fine	.012	nice	.0				
express	.014	white	.013	ready	.012	great	.0				
respond	.014	cannot	.013	reason	.012	vacation	.0				

## THERAPY-RELATED TOPIC

TBCT							
Topic 60	.016	Topic 49	.013	Topic 83	.014	Topic 90	.0
problem	.262	solution	.088	listen	.072	pattern	.0
solving	.092	down	.067	communicatio	.053	point	.0
issue	.049	write	.062	paraphrase	.030	two	.0
solution	.035	idea	.051	give	.029	down	.0
work	.025	pros	.038	person	.027	become	.0
talk	.022	cons	.037	talk	.027	part	.0
two	.022	brainstorming	.034	trying	.026	conflict	.0
define	.020	good	.025	level	.023	response	.0
part	.019	two	.024	speaker	.022	process	.0
definitely	.018	problem	.023	practice	.019	recognize	.0
brainstorming	.015	possible	.021	point	.018	situation	.0
trying	.014	trying	.018	editing	.017	deal	.0
communicatio	.014	discuss	.017	x	.017	withdraw	.0
process	.013	talk	.017	floor	.016	style	.0
discuss	.012	agree	.016	summarize	.015	change	.0

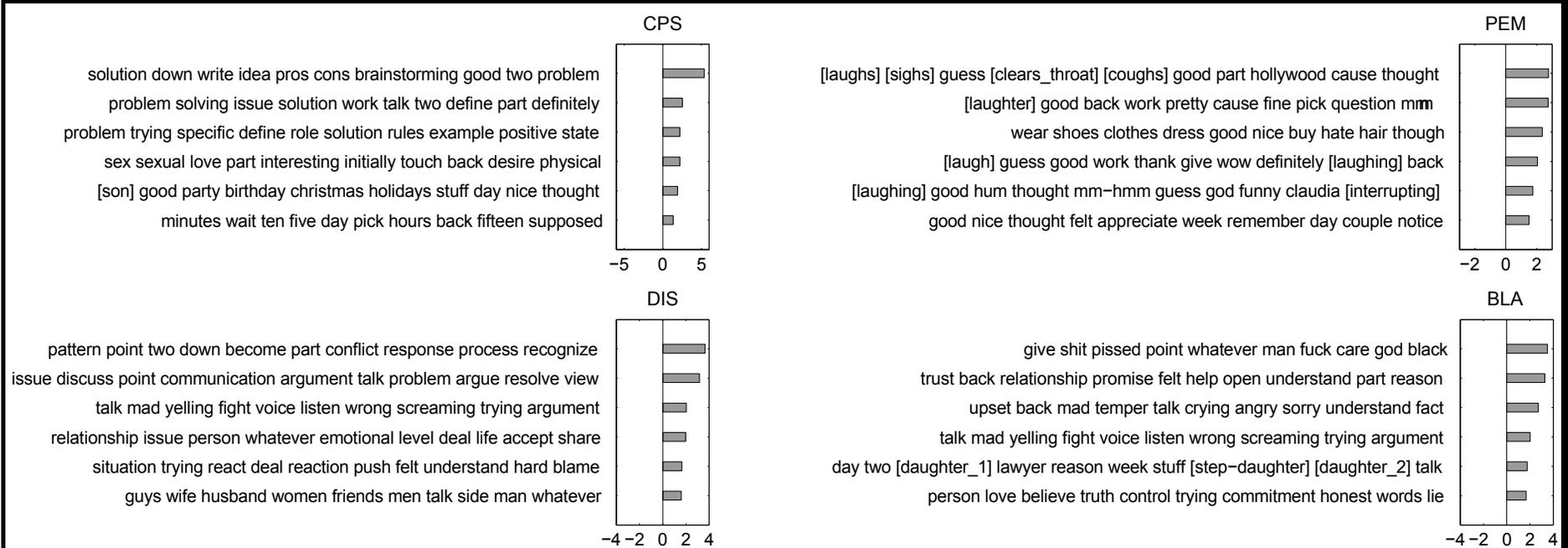
Topic models focus on finding underlying semantic content, or the "gist" of what is said (or written)

THERAPIST: ...and the model<sup>77</sup> is one in which you define<sup>60</sup> the problem<sup>60</sup>, you brainstorm<sup>49</sup> ideas<sup>49</sup>, you talk<sup>77</sup> about the pros<sup>49</sup> and cons<sup>49</sup> of the options<sup>77</sup>, and then the last part<sup>60</sup> of it, or the next to last part<sup>60</sup> of it is to integrate<sup>77</sup> the ideas<sup>49</sup> that you decided<sup>77</sup> were good<sup>49</sup> ideas<sup>49</sup> into a strategy<sup>77</sup> that you're going to um ... that doesn't mean that we know for sure it is going to work<sup>18</sup>. In fact<sup>19</sup> that is the reason<sup>77</sup> why the last stage<sup>60</sup> of the model<sup>77</sup> is to set specified<sup>49</sup> times that we should reevaluate<sup>77</sup> and see how your solution<sup>49</sup> is working<sup>18</sup> or not working<sup>18</sup> and you adjust<sup>18</sup> it accordingly<sup>4</sup>. Right? That was the model<sup>77</sup>, so um, are you um, contemplating<sup>57</sup> some problems<sup>60</sup> with the model<sup>77</sup> or are you um, suggesting<sup>77</sup> some problems<sup>60</sup> with the items<sup>49</sup> that we went through with that, that hasn't<sup>88</sup> been articulated<sup>77</sup> yet?

WIFE: Well, I mean I'm, I'm looking at the problem<sup>60</sup> and the potential<sup>60</sup> solutions<sup>60</sup> and, short<sup>12</sup> of me doing something, then nothing will change<sup>22</sup>.

# Topic model output has fairly strong association with behavioral codes

Transcripts →  
 Topic model →  
 Output (low dimensional summary) →  
 Regression model for codes



Average  $R^2 = \sim 0.85$

# Speech signal processing is natural method to analyze tone.

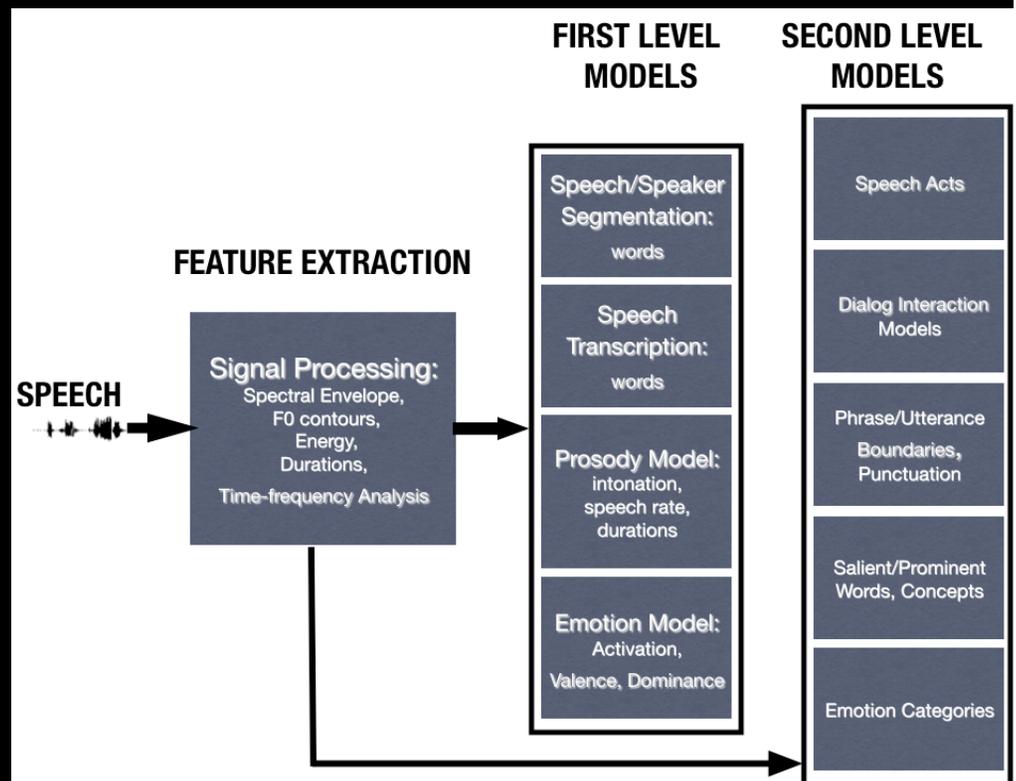
Cowie et al. (2001): "...the same words may be used as a joke, or as a genuine question seeking an answer, or as an aggressive challenge [e.g., 'I suppose you think England is going to win the World Cup']" (p. 33).

Acoustic features can be decoded from audio signal:

F<sub>0</sub> : Fundamental frequency (arousal)

Energy : Loudness

Jitter : Tremble of vocal chords



# Acoustic features are also predictive of behavioral codes in session

Toward automating a human behavioral coding system for married couples' interactions using acoustic features

Matthew P. Black<sup>a,\*</sup>, Athanasios Katsamanis<sup>a</sup>, Brian R. Baucom<sup>b</sup>, Chi-Chun Lee<sup>a</sup>, Adam C. Lammert<sup>a</sup>, Andrew Christensen<sup>c</sup>, Panayiotis G. Georgiou<sup>a</sup>, Shrikanth S. Narayanan<sup>a,b</sup>

<sup>a</sup>Signal Analysis & Interpretation Laboratory (SAIL), University of Southern California, 3710 McClintock Ave., Los Angeles, CA 90089, USA<sup>\*</sup>

<sup>b</sup>Department of Psychology, University of Southern California (USC), 3620 McClintock Ave., Los Angeles, CA 90089, USA

<sup>c</sup>Department of Psychology, University of California, Los Angeles (UCLA), 1285 Franz Hall, Los Angeles, CA 90095, USA

Based purely on acoustic features (ie, no words), average classification accuracy between 70%-75%

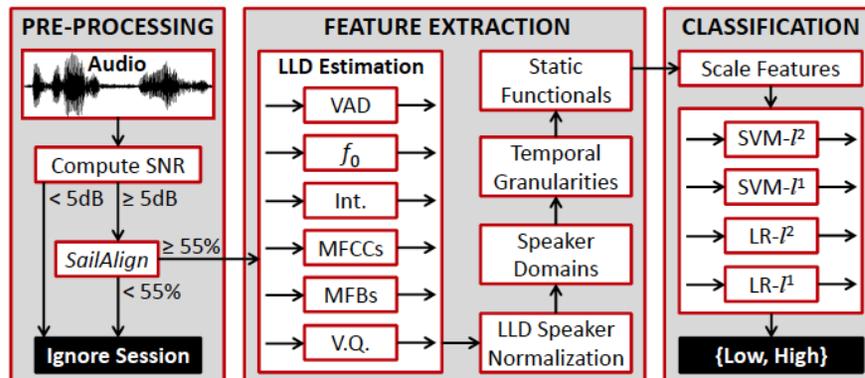


Figure 2: A system block diagram, illustrating the methodology taken in this paper, from pre-processing the data and extracting acoustic features to classifying extreme instances of a particular code as low/high.

Okay, so:

1. Semantic content in therapeutic dialogue is related to fidelity
2. Prosody and pitch is as well

But, these are still one step removed from automated assessment of fidelity

# What is the actual possibility of generating near real-time fidelity?

“That’s aggravating, very aggravating”: Is it possible to classify behaviors in couple interactions using automatically derived lexical features?

Panayiotis G. Georgiou<sup>1</sup>, Matthew P. Black<sup>1</sup>, Adam C. Lammert<sup>1</sup>, Brian R. Baucom<sup>2</sup>, and Shrikanth S. Narayanan<sup>1,2</sup>

<sup>1</sup> Signal Analysis and Interpretation Laboratory (SAIL), Los Angeles, CA, USA

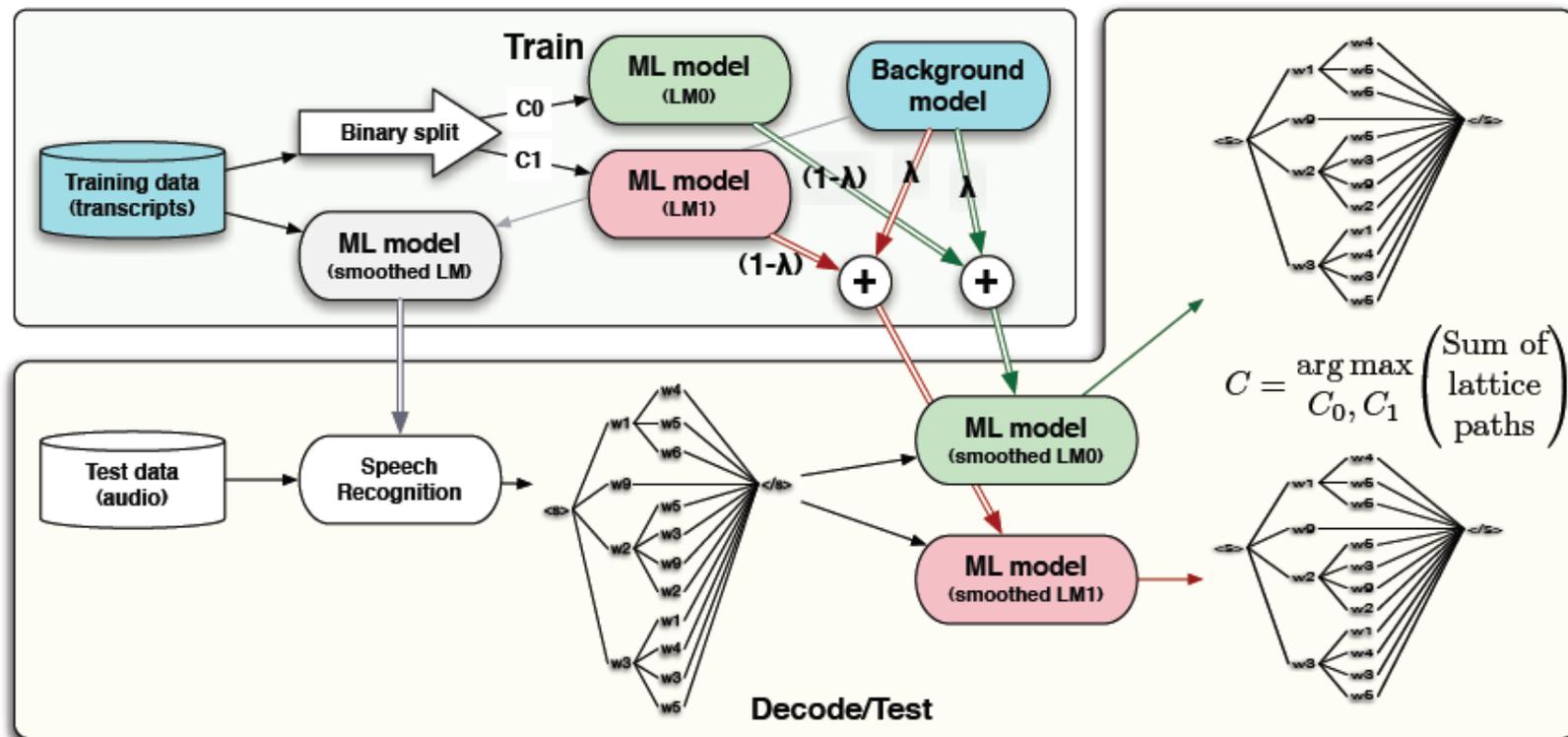
<sup>2</sup> Department of Psychology, Univ. of Southern California, Los Angeles, CA, USA  
<http://sail.usc.edu>

Copyright 2004 by Randy Glasbergen.  
[www.glasbergen.com](http://www.glasbergen.com)



**“We need better voice recognition software.  
Instead of a pay cut, everyone was given a paper cut.”**

# Preliminary research shows basic “proof of concept” for automated fidelity



**Fig. 3.** Overview of the classification process without human transcripts through the use of ASR lattices.

# Future Work / Ideas

Just finishing year 1 of NIAAA R01, so much more to come with applying linguistic methods to MI

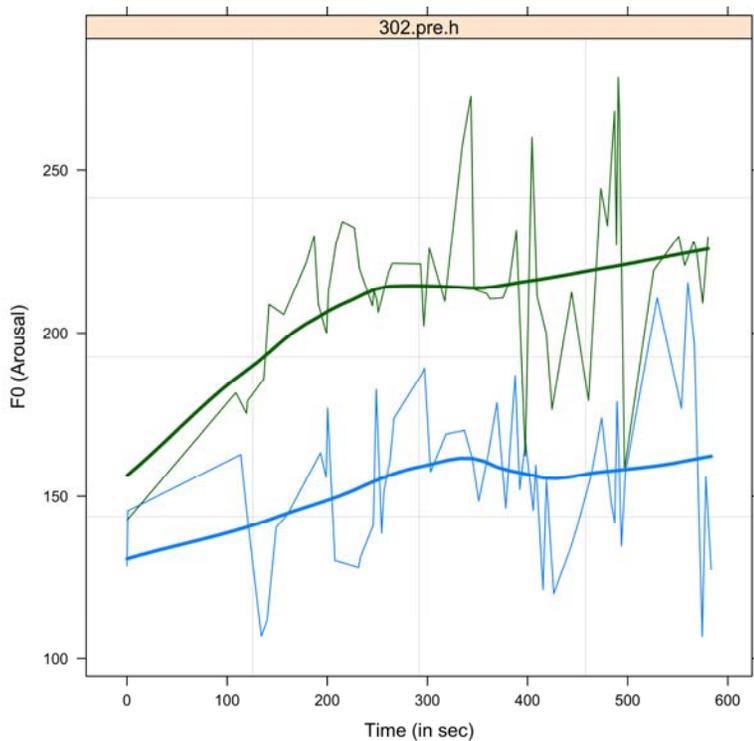
Exploratory Aims related to automated coding in currently submitted NIMH proposal focused on implementing evidence-based anxiety program at local HMO (Simon, PI)

Developing R34 proposal to NIDA (Zac Imel, PI) to explore developing a tool to use in training and supervision of MI

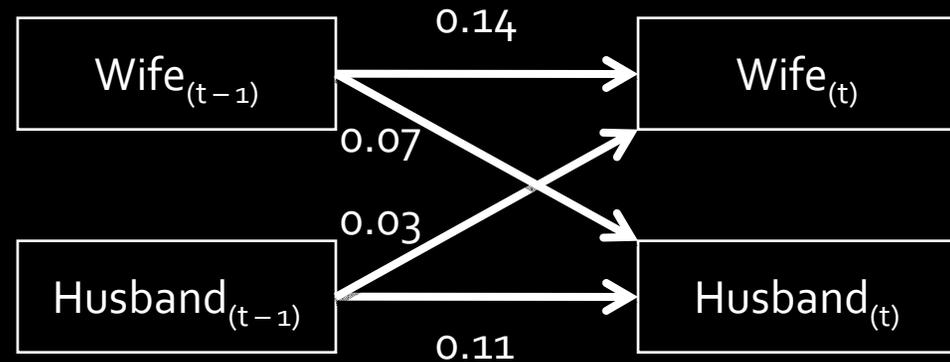
Contact Info: Dave Atkins, [datkins@uw.edu](mailto:datkins@uw.edu)



# Do acoustic features pick-up therapy relevant content?



Longitudinal actor-partner model for spousal linkage of arousal during 10 minute communication assessment



Additional findings:

1. Pathways are moderated by couple distress
2. Baseline Fo is associated with long-term outcome