Fidelity of automatically coded family speech of mothers, fathers, and 30 month-old children with and without hearing loss

Mark VanDam,[1] Paul De Palma,[2] & Noah Silbert[3]

[1]Department of Speech & Hearing Sciences,
Elson S. Floyd College of Medicine,
Washington State University, and
The Hearing Oral Program of Excellence (HOPE)

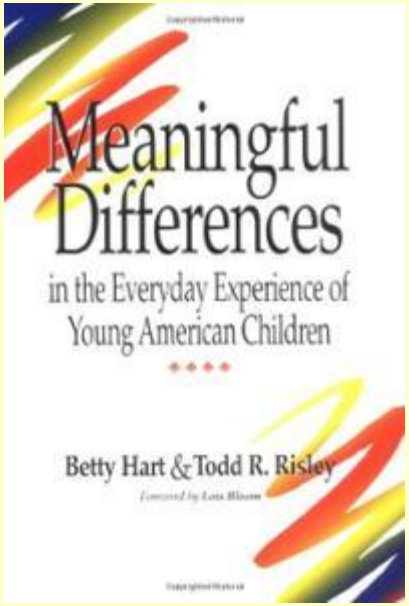
[2]Department of Computer Science, School of Engineering and Applied Science, Gonzaga University

[3]Department of Communication Sciences & Disorders
University of Cincinnati

Talk presented as part of the Paper Symposium "Studying Language Development Through Human and Automated Annotation of Infants' Natural Auditory Environments" at the *2016 International Conference on Infant Studies (ICIS)*, New Orleans, LA, May 27, 2016

FUNDING:
NSF/SBE-RIDIR: 1539133 (VanDam), 1539129 (Warlaumont), 1539010 (MacWhinney)
NIH/NIDCD: R01DC009569, DC009560-01S1(Moeller & Tomblin)
WSU Seed Grant: 124172-001 (VanDam)
Washington Research Foundation (VanDam)

Hart & Risley (1995) collected child speech data in natural, home environments of 42 families.



Such data is very expensive to collect and difficult to analyze and interpret.

It took H&R 3 yrs to collect and 6 yrs to interpret.

But now, useful child language data is collected using automatic speech processing (ASP) technology.

Researchers are using the Language ENvironment Analysis, LENA

Zimmerman, etal (2009) *Pediatrics*
Christakis, etal (2009) *Arch Pediatrics & Adol Med*
Oller, etal (2010) *PNAS*
Warren, etal (2010) *Journal Autism & Devel Disord*
Caskey, etal (2011) *Pediatrics*
Dykstra, etal (2012) *Journal of Autism*
VanDam, etal (2012) *Journal Deaf Studies & Deaf Educ*
Aragon & Yoshinaga-Itano (2012) *Sem Speech & Lang*
Suskind, etal (2013) *Comm Disord Quarterly*
Weisleder & Fernald (2013) *Psychological Science*
VanDam & Silbert (2013) *POMA*
Ambrose, etal (2014) *Ear & Hearing*
Warlaumont, etal (2014) *Psychological Science*
Canault, etal (2015) *Behavior Research Methods*
Gilkerson, etal (2015) *Journal of Speech-Language Hearing Research*
Odean, etal (2015) *Frontiers in Psychology*
Sosa (2015) *Journal of the American Medical Association – Pediatrics*
Suskind, etal (2015) *Journal of Child Language*

..... ++ and many more

Primary goals of this work are to (1) report goodness of *LENA* labeling technology and (2) compare machine performance for families with a typically-dev toddler and a toddler with hearing loss.

Data collection

LENA™
Every Word Counts

Labels on the acoustic signal:

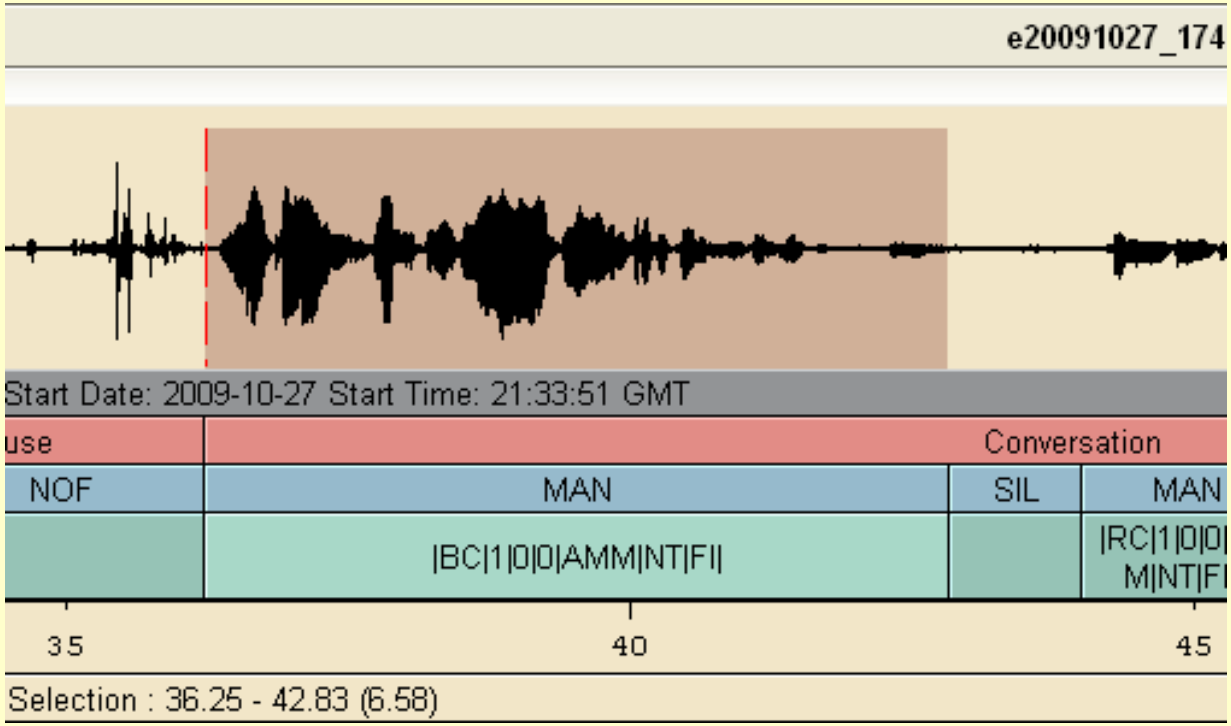
KEY-CHILD ***OTHER-CHILD***

ADULT-MALE ***ADULT-FEMALE*** ← **live human vocals**

SILENCE ***NOISE***

ELECTRONIC (TV, RADIO) ***UNCERTAIN / FUZZY*** ← **other acoustic events**

OVERLAPPING VOCALS



Automatic data collection results in very large database (VLDB) requiring <u>fully automated</u> data analyses.

Reliability of LENA labels, previous findings

ASR agreement for segments humans labeled as

‘adult’ = 82%

‘child’ = 76% & 73%

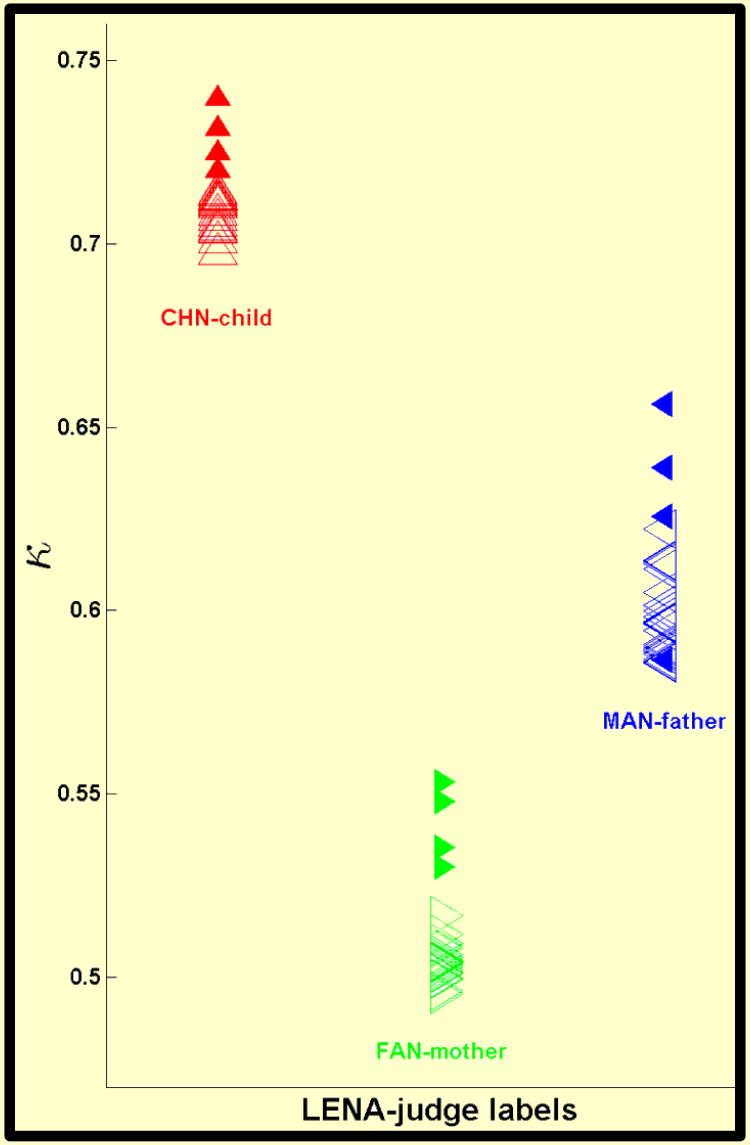
Human agreement for segments ASR labeled as

‘adult’ = 68%

‘child’ = 70% & 64%

Xu *etal* 2009; Christakis *etal* 2009; Warren *etal* 2010; Zimmerman *etal* 2009; Oller *etal* 2010; Soderstrom & Wittebolle 2013; Canault et al 2015; Weisleder & Fernald 2013

Reliability of LENA labels with typ-dev kids



ASR—human agreement

	%	κ
CHN-child	85.9	.709
FAN-mother	59.6	.505
MAN-father	60.8	.598

VanDam & Silbert, 2013; 2016

The present work, method and design

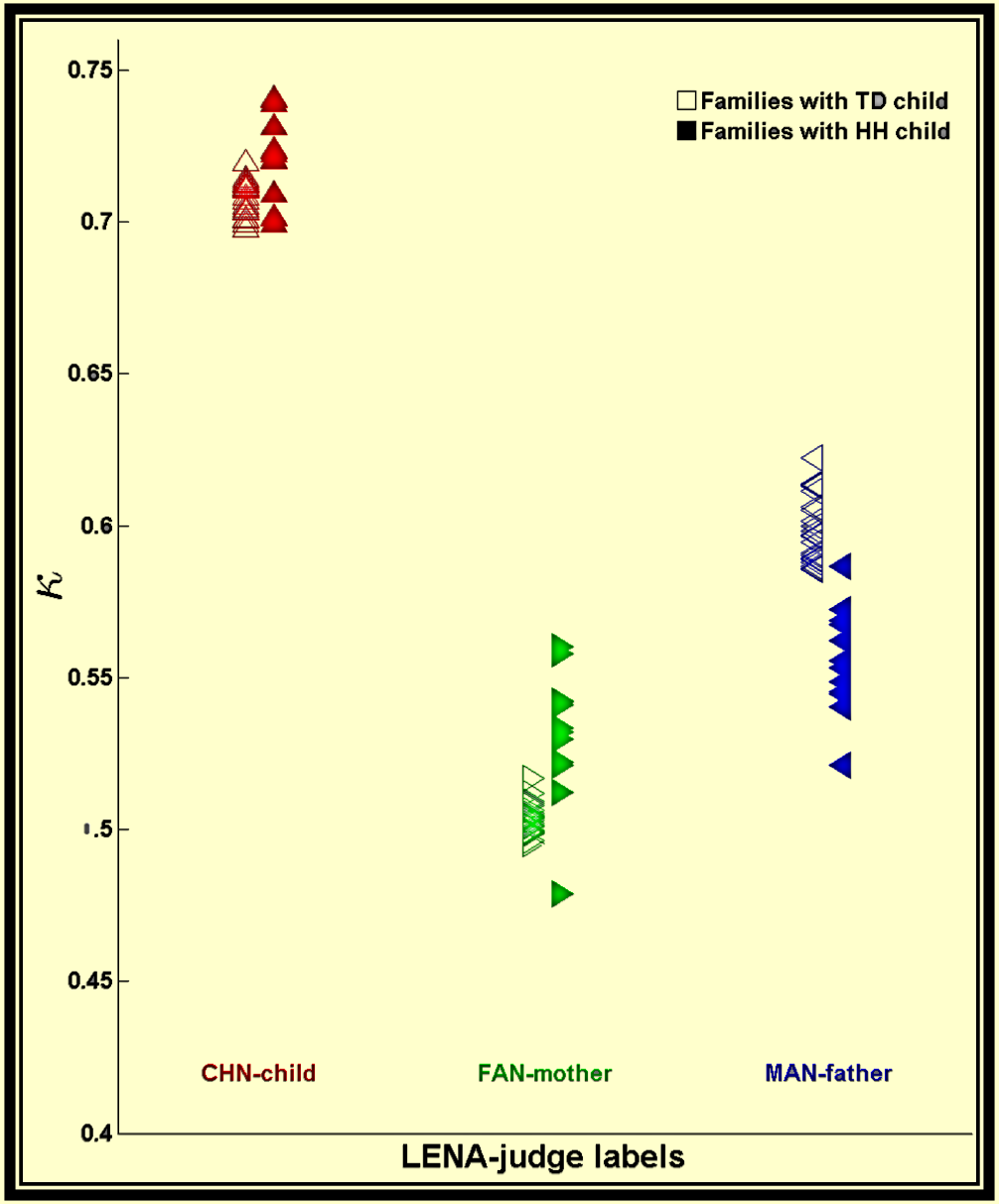
Stimuli: 2340 tokens from 26 families with TD child (mean age=29.2 mos).
2340 tokens from 26 families with HH child (mean age =28.9 mos).
30 tokens each from *adult-female (FAN), adult-male (MAN), child (CHN)*
26 x 30 x 3 = 2340

Judges: 24 judges for TD stimuli; 13 judges for HH stimuli
about 2hrs of listening per judge
all judges listened to the same stim tokens (per group)
86,000+ auditory decisions/classifications

Task: 4AFC: *mom, dad, child, other*

Analysis: percent correct, Cohen's kappa, (regression) factor analysis

Results: reliability of LENA labels with HH kids



ASR—human agreement, TD families

	κ
CHN-child	.708
FAN-mother	.503
MAN-father	.599

ASR—human agreement, HH families

	κ
CHN-child	.721
FAN-mother	.530
MAN-father	.552

TD and HH distributions are different

	t	*p*
CHN-child	-3.79	$<10^{-3}$
FAN-mother	-6.24	$<10^{-6}$
MAN-father	9.85	$<10^{-11}$

Limitations

1. Data are messy. Algorithm artifacts (whisper, singing, range parameters) may influence machine output, but we can only speculate.

2. Other factors are known to play a role: spectral envelope/mean/tilt, shimmer (amp entropy), jitter (*f*0 entropy), SNR, nasalance, vocal quality (creak, fry), etc.

3. Individual differences surely exist (judges, stimuli).

4. Algorithm is a black-box. Alternative processing is not yet available.

Conclusions

1. **Machine labeling of child, mother, and father segments are not equally well done.**

2. **Machines and humans get fairly similar results—enough to be useful.**

3. **Machine labels of human talkers in HH families is better for moms and kids, but worse for dads. Not really sure why this would be.**

Future directions

1. Families with kids with hearing loss. Do machines or humans deal with hearing loss in the same way?

Machine and humans seem to treat TD and HH data similarly; very long duration (>970ms) may be unique to TD kids; interestingly, some of the *f*0 or *f*0-contour does not seem to be unique to TD kids.

2. Other disorders, including ASD, SLI, older/younger kids.

3. More modeling of the factors that influence decision making

4. Alternative processing options

Q & A

FACTORS:

1. **duration**
2. ***f*0 – mean**
3. ***f*0 – min**
4. ***f*0 – max**
5. ***f*0 – rise**
6. ***f*0 – fall**
7. **amp, RMS**
8. **amp, rise**
9. **amp, fall**