Vancouver Phonology Group
April 9, 2016

Investigating perceptual biases, data reliability, and data discovery in a methodology for collecting speech errors from audio recordings

John Alderete, Monica Davies
Simon Fraser University

Collaborators: Stefan Frisch (South Florida) Alexei Kochetov (Toronto)

Thanks also to: *Holly Wilbee, Jennifer Williams, Gloria Fan, Rebecca Cho*

Abstract

This work describes a methodology of collecting speech errors from audio recordings and probes how some of its methodological decisions affect data quality and composition. Speech errors of all types (phonetic, phonological, syntactic, etc.) were collected by six data collectors from podcasts of natural and unscripted English speech. Analysis of these errors showed that (i) different listeners find different errors in the same audio recordings, but (ii) the frequencies of error types are similar across listeners; (iii) errors collected "online" in daily life without recourse to a recording are more likely to be affected by perceptual biases than "offline" errors collected from podcasts, and (iv) when properly trained, listeners are able to collect gradient phonetic errors that have not traditionally been included in speech error collections.

Speech error research

Common stock of problems, English and other languages (Shattuck-Huffnagel, Dell, Stemberger)

- Sound errors: source and error 'close', phonological similarity, syllable position constraint, word-onset effect; lexical bias, frequent sound bias, no category constraint
- Word errors: not necessarily 'close', category constraint, semantic bias, mixed errors

Groundwork for organization of production processes (Garrett, Fromkin, Dell, Stemberger)

- Lexical errors: lemma selection
- Word exchanges: function assignment
- Sound errors: phonological (and phonetic) encoding

Linguistic insights (Fromkin, Stemberger, Shattuck-Huffnagel)

- Psychological reality of phonemes, features, rimes, etc.; structural notions like ambisyllabicity and CV structure

Problems with observational techniques

Labor intensive

- Most large studies take several years to see results

Error prone

- Errors can be missed, incorrectly documented, and misclassified; about 1 in 4 need to be tossed out

Subject to perceptual biases, perhaps not representative

- Difficult to hear errors in certain contexts, certain structures
- Perceptual systems pull gradient patterns to poles
- Transcription methods not always best

Ambiguity of the data

- Many errors can be classified in more than one way, making enumeration of error types difficult

The turn from observational methods

- To mitigate some of these problems, many use experimental techniques: SLIP technique, tongue twister technique; also normal speaking, brain imaging

Goal: methodology that mitigates problems

Objective

Modify observational techniques to make them more efficient and less error prone

- Many data collectors: many talkers
- A rigorous training regime
- Separate data collection and classification
- Exclusive use of audio recordings: “offline” data collection

Progress in a year and a half

With team of five students, collected 8,500 errors of all types

Approx. 2,000 of 8.5k classified

Goals: Specific research questions

1. Given multiple data collectors, are there any important effects due to data collectors?

- Expect to find differences, but not clear if they will affect the sample

2. Is data collected online different from data collected offline?

- Expect to find differences, and that offline collection will result in a more accurate reflection of error levels and types.

3. Given offline collection method, what other uses might audio recordings have in language production studies?

- Gradient phonetic structures, existence and relevance for principles of speech production.

Context: Problems with data reliability

Sources of incorrect errors or unrepresentative sample of actual patterns

- Failure to record errors
  - errors are not recognized/remembered
  - errors occur too fast, past time envelope
- False Positives
  - error was misheard ('slip of ear', Cutler 1981)
  - incorrect recall or transcription of error
  - error doesn't meet definition "non-habitual deviation from speech plan"
- Misclassification
  - due to incorrect transcriptions
  - incorrect or insufficient contextual information
  - coding inconsistency

A reliability test: Ferber 1991

4 listeners, 45 minutes of German (online), also listened offline by Ferber

- Approx. 2/3 online errors heard by only one listener
- Important differences between two classes of listeners: one listener found no sound errors, two listeners found no word errors
- Many false positives: 24% (12/51 all online)

Context: Perceptual biases (Bock 1996, Pérez et al. 2007)

Content bias: errors are easier to detect if they affect meaning

Word onset: errors are less noticeable if they occur at the end of words than the beginnings of words

Predictability: errors also easier to detect when they occur in highly predictable environments (e.g., *… go smoke a **cikarette***) or primed with words associated with the error word

Bias for discrete symbolic categories: especially for sound errors, biased toward assigning sounds to discrete phonological categories

Compensation for coarticulation: phonetic environments may enhance certain contrasts and lead to selection of some discrete sounds over others

Attention bias: lower level errors (phonetic or phonological) are often more difficult to detect and therefore require greater attention, substitution harder detect than exchange (e.g., ***left lemisphere*** vs. ***heft lemisphere***)

Feature biases: sound errors with changes in some features are easier to detect that others, e.g., nasality and voicing easier than place; also more feature changes easier to detect than fewer

Context: Additional constraints and biases

Data collector bias: collectors differ in the rate of detection and types of errors (see perceptual biases)

Talker sampling bias: sample may be biased toward the errors of specific individuals, e.g., if just one data collector

Theoretical bias: purchase of a theory or specific hypothesis may affect the errors collected

Memory constraints: mentally very taxing to follow a conversation, track entities, and then detect and record errors

Native language bias: hard to detect errors that don't fall into the categories of your own language

Context: Common ways of addressing problems

Best practices: only write down errors that have high degree of confidence in, 30 seconds after speech act, make conscious effort to collect (no multi-tasking)

Intended words: Immediately question speaker if intended word unclear (not always useful)

Many data collectors: Dell & Reich (1981) used 200 undergraduate students to collect a large sample; see also Pérez et al. 2007

Collect lots of data: 3 data collections have upwards of 6,000 errors

Compare with experimental data (see Stemberger): if fits with trends in experimental data, stronger statement; otherwise look for tasks effects or problems with data collection

- **Objection**
  Most large corpora are collected online by one or two individuals; therefore prone to (1) perceptual biases and (2) reliability problems.

SFUSED: General methods

Multiple data collectors

- reduces collector bias, allows it to be studied
- currently six data collectors, generating about 200 errors a week

Offline collection from audio recordings

- Errors collected from podcasts on different topics
- Podcasts selected for having natural unscripted speech, usually Western Canada and U.S. (Midlands dialect 'Standard American')
- Examples: *Astronomy Cast, Go Bayside, Accidental Tech, Rooster Teeth*
- Multiple podcasts (8 currently) with different talkers, approx. 50 hours of each podcast
- Record dialectal and idiolectual features associated with speakers (because habitual, so not an error); listeners develop expectations about individuals
- Older approach combined online and offline, but no longer collect online

Training regime

- Undergraduate students, introduction to formal linguistics, phonetics and phonology
- Given phonetic training in transcription and tested for transcription accuracy
- Introduction to speech errors, definition and illustration of all types
- Training through listening tests: assigned pre-screened recordings, asked to find errors; learn by reviewing correct list of errors. Trainees that reach a certain level of accuracy continue.

SFUSED: Addressing the problems

- **Follow best practices (online)**
  - direct observation (no secondary sources)
  - record error within 30 seconds
  - conscious effort
- **Collector bias**
  - multiple collectors
    - so biases can be studied
    - also, if bias exists, reduced to specific collectors
- **Talker bias**
  - reducing effect of specific individuals
    - multiple collectors
    - multiple recordings
    - larger number of talkers
- **Theory bias**
  - most data collectors not versed in language production theory
    - those that are not assigned a project
- **Symbolic categories**
  - we allow for gradient categories to be recorded
- **Perceptual biases**
  - inherent problem
    - can at least study given that researchers tagged, so can look for effects
- **Missed errors**
  - inherent problem, and difficult one
    - but available of audio recording reduces this considerably
    - Error detection rates enables us to determine if many errors were missed
- **Incorrect errors**
  - also difficult problem
    - reduce it with phonetic training
    - Must have a second check at classification stage

SFUSED: Raw collection phase

2015b_speechErrors.xlsx

research	sex	self?	date	example	intended	confidence	gradient	corrected?	podcast	file	time	notes
gf	m		2015-04-13	If it just feels like 'hey why not let's go out' I /fit xxx I think you know what? People on okaycupid	think			y	whc	whc012_2014-07-3	0:17:05	
gf	m		2015-04-13	yep and uh just send anything you've seen that you want us to /[s]alk about to we have concerns dot	talk				whc	whc012_2014-07-3	0:22:35	
gf	m		2015-04-13	"You're /[ʃ]ertainly shooting the girls this morning Dr. P"	certainly				whc	whc013_2014-08-(	0:02:41	speaker adopted a "surfer dude" accent while saying this phrase
gf	m		2015-04-13	yea yea kim's right. Kim's right. Uh she wou he probably got /har lost in the harsh martain /wil[t]erne	lost; wilderness			y	whc	whc013_2014-08-(	0:07:02	only first error was corrected; first error: speaker began saying 'harsh'
gf	m		2015-04-13	right, I feel like even if you're /[r]ith a small group of people	with				whc	whc013_2014-08-(	0:08:11	
gf	m		2015-04-13	Look, /I'm made a house out of all of our mash potatoes	I				whc	whc013_2014-08-(	0:18:17	
gf	m		2015-04-13	I feel like I'm already making a more efficient use of my time because I'm /washing the movie and an	watching				whc	whc014_2014-08-(	0:06:18	
gf	m		2015-04-13	I'd have sex with cave girl /anna xxx Emma Stone	Emma			y	whc	whc014_2014-08-(	0:07:41	
gf	m		2015-04-13	well how good /it xxx am I at doing something	am			y	whc	whc014_2014-08-(	0:11:18	
gf	m		2015-04-14	I think that you duh the short attention /[ʃ]pan can be manipulated by shiny objects	span				whc	whc014_2014-08-(	0:13:50	
gf	m		2015-04-14	kids ask your mom what a radio station is xxx cause /[gɛ] here comes another good one	(null)				whc	whc014_2014-08-(	0:16:51	
gf	m		2015-04-14	but now that's xxx that's why you're gonna have to name your /kit xxx kid like [Zæzbɒt]	kid			y	whc	whc015_2014-08-(	0:03:04	
gf	m		2015-04-14	I had a guy in college say /# don't wanna xxx you don't wanna be Jeffrey	you			y	whc	whc015_2014-08-(	0:08:19	
gf	m		2015-04-14	isn't it annoying when somebody jus xxx when somebody does one and you didn't give them /per[t]i	permission			y	whc	whc015_2014-08-(	0:08:43	speaker makes an error half way through the word and corrects it withou
gf	m		2015-04-14	oh I'm here /I'm[b] both an awesome clown entertainment personality and the brother of your /[s] x	I'm; mother			y	whc	whc015_2014-08-(	0:10:57	only second error was corrected; speaker may have started to say 'sister'
gf	m		2015-04-14	not even you 'cause steve was /free bagging you was taken? (other speaker: not free bagging, tea ba	tea			y	whc	whc015_2014-08-(	0:17:40	correction was prompted by other speaker
gf	m		2015-04-15	the /[tʃ] xxx the original transformers surprised me by being moderately watchable	original			y	whc	whc016_2014-08-:	0:00:16	
gf	m		2015-04-15	you can see this in our every day lives where we xxx /[w|r]e talk about celebrities in a certain way	we		trans		whc	whc016_2014-08-:	0:04:44	
gf	m		2015-04-15	hey man I want /you to xxx introduce you to my friend Carl	introduce			y	whc	whc016_2014-08-:	0:08:36	
gf	m		2015-04-15	yeah you /[k]uys can't see it but I'm playing with one of those tech deck fingerboards right now	guys				whc	whc016_2014-08-:	0:11:11	
gf	m		2015-04-15	if you've only /[f] xxx argued three times and she was like "we've only argued three times and it hasn	argued			y	whc	whc016_2014-08-:	0:15:26	correction was prompted by other speaker
gf	m		2015-04-15	oh you're wearing /they'll xxx the like tennis racket on your feet?	the like			y	whc	whc017_2014-08-:	0:05:30	
gf	m		2015-04-15	if we can't /[tʃ]oot for an ideal with regard to the gorilla-arm ratio	shoot				whc	whc017_2014-08-:	0:10:21	
gf	m		2015-04-15	If you're about to say that you accidentally gave drugs to a /[tʃ] xxx donkey, you're remembering the	donkey			y	whc	whc017_2014-08-:	0:13:05	
gf	m		2014-04-16	I saw /bamboos for real xxx ah bamboos? I saw Koalas for real in uh xxx uh I saw them in Australia	koalas			y	whc	whc017_2014-08-:	0:14:25	
gf	m		2015-04-16	/set xxx leave a review that says I believe in my heart that the host of this show could fight one hund	says			y	whc	whc017_2014-08-:	0:17:39	speaker may have anticipated a word that appears later on in the sentenc
gf	m		2015-04-16	yeah where /where[t] point does personality develop?	what				whc	whc018_2014-08-:	0:01:59	speaker seems to blend 'where' and 'what' together
gf	m		2015-04-16	If it's not a direct necessity for survival. If it's only informed by your /ne[sɑsɛs] xxx your necessity for	necessity			y	whc	whc018_2014-08-:	0:05:18	
gf	m		2015-04-16	/I her# xxx I'm here for the petting zoo	I'm here			y	whc	whc018_2014-08-:	0:11:51	
gf	m		2015-04-16	There've been a lot of new scans and tests and studies on dogs that show they do have a very /high[j	high				whc	whc018_2014-08-:	0:16:11	
gf	m		2015-04-16	God I hope I meet /than alien	an	low			whc	whc018_2014-08-:	0:16:59	speaker says it in a whisper tone
gf	m		2015-04-16	and /they conclusion that they came to is that it existed in this weird riff between species	the				whc	whc018_2014-08-:	0:19:13	
gf	m		2015-04-16	that's dolphin for leave review. /They dolphins love iTunes	the				whc	whc018_2014-08-:	0:21:26	
gf	m		2015-04-16	we call it trickle up /epi xxx economics	economics			y	whc	whc019_2014-08-:	0:00:56	
gf	m		2015-04-17	nah you know what /[b|m]an, this thing is weird	man		ambig		whc	whc019_2014-08-:	0:02:13	
gf	m		2015-04-17	they're for fish commutes? That sounds like really /# inefficient.	it's				whc	whc019_2014-08-:	0:03:14	missing word
gf	m		2015-04-17	forty fish per minute uh shooting out across /though xxx the vastness of space into uh higher on the	the			y	whc	whc019_2014-08-:	0:06:19	
gf	m		2015-04-17	shove a /san in it xxx salmon in to that	salmon			y	whc	whc019_2014-08-:	0:06:58	
gf	m		2015-04-17	I don't think this is orginally built for /frui[tʃ] xxx uh for fruit transportation	fruit			y	whc	whc019_2014-08-:	0:14:29	
gf	m		2015-04-17	/you sent xxx your mom sends thousands of dollars through xxx that tube and then two lollipops con	your mom			y	whc	whc019_2014-08-:	0:18:24	
gf	m		2015-04-17	I agree with you. I love the whimsy /in[θ] xxx inside things	inside			y	whc	whc019_2014-08-:	0:21:10	
gf	m		2015-04-17	and there's a /[k]ood fucking reason we don't	good				whc	whc019_2014-08-:	0:21:18	
gf	m		2015-04-19	Such /[ʃ]alesmanship.	salesmanship				whc	whc020_2014-08-:	0:00:47	
gf	m		2015-04-19	And a lot of people were skeptical. And uh I kinda /rolled up# xxx wrote it up as BS	wrote it up				whc	whc020_2014-08-:	0:01:10	
gf	m		2015-04-19	which is xxx which could completely /revol[æ] xxx lutionize our world	revolutionize			y	whc	whc020_2014-08-:	0:02:36	speaker only corrects the second part of the word that contains the error

SFUSED: Classification phase

Add new variables (largely following Stemberger 1993)

- Raw speech errors then given a set of 45 fields that classify the error

Example

- "I read officially it's **/[k]upposed** to ^come out this afternoon"
  - *intended:* **supposed**

Classification:

- level: **sound**, **contextual** (from *come),* **nonlexical**
- type: **anticipation**
- intruder: **[k]**
- supplantedIntended: **[s]**
- supplantedIntendedSyllableRole: **onset**
- supplantedIntendedSyllabicPosition: **word-initial**, etc.

Variables:

- Right lemma?, lexical, contextual, level, gradient, confidence, contentWord, intruder, supplantedIntended, syllableRole, syllable position, intendedPartOfSpeech, distance, clause-bound, word-bound
- use a set of attributes enable us to test specific hypotheses about the structure of speech errors, e.g., phonological similarity

Verification

- All errors are relistened and checked for validity and accuracy

SFUSED: Illustration of the interface

Layout: Main Data Input | View As: | Preview | Edit Layout

SFU Speech Error Database | **Completed records so far:** 290 | **This is record No.** 1500

Major Class Variables

classLevel: Sound
classType: Substitution
classDirection: Perseveration
classContextual: (•) Y ( ) N
classLevelAlternate:

Example Variables

I'm about halfway through judging by the maps at the ^selection /[ʃkrin]

Intended Word: screen
Intended in IPA: 'skrin
Error Word: [ʃkrin]
Error in IPA: 'ʃkrin
Corrected?
Clipped Word? ( ) Y (•) N
Identical Neighboring Seg? ( ) Y (•) N ( ) NotAppl
Confidence Intended
Confidence Transcription
Word Bounded Error? ( ) Y (•) N ( ) NotAppl
Syllable with error has (sound errors only):
(•) Main Stress ( ) Second. Stress ( ) No Stress

Word Variables

Right Lemma? (•) Y ( ) N ( ) NotAppl
Lexical Word? N
Error POS: NotAppl
Intended POS: Noun
Intended Open/Closed? (•) Open ( ) Closed ( ) See Notes

Sound Variables

SupplantedIntended
s
Syllable Role
C - Appendix
Word Position
Initial

Intruder
ʃ
Syllabic Role in Error: C - Appendix
Syllabic Role in Source:
Word Position in Error: Initial
Word Position in Source:

Record Variables

Record Completed? (•) Y ( ) N
Confirmed By [x] ja [ ] md [ ] hw
DateLastModified 2015-05-17 8:18:49 PM
Found by Researcher dw
Found Date 2014-05-22
File Name
Time of File
Online/offline on
Talker Self?
recordSex m

Specific Class Variables

Blend Type:
Gradient Type:
Prosody Type:

Check all patterns for Complex Processes:
[ ] Sub [ ] Add [ ] Del [ ] Shift [ ] Stress [ ] Grad [ ] Exch

Weird Morpho or Syntax Error we want to come back to: ( ) Y ( ) N

Notes

Experiment 1: Effect of data collector

Questions

- How reliable is our error collection methodology?
- Are data collectors collecting the same types of errors?
- What errors are heard by all?
- Which are heard by specific collectors?

Methods

- Listened to nine 30-40 min. podcasts (three different series), 380 minutes total
- Each podcast listened by three trained listeners (two split the nine)
- Made error submissions separately, then at least two listened to all errors and checked error validity

Basic findings

Total 'errors' found	380	*error every minute*
Total 'errors' incorrect	94 (25%)	
Total correct errors	286 (75%)	*error every 1 minute 20 secs*

Experiment 1: Accuracy and rate

	Total 'errors'	False positives	% correct	Minutes per correct error
Listener 1	50	16	68%	4.85
Listener 2	85	18	79%	3.21
Listener 3	177	33	81%	2.64
Listener 4	206	32	84%	2.18

Implications

- Can require 70% accuracy as criterion for continuation in data collection
- Individual collectors far about actual MPE, but impossible to expect collectors to find all errors
- Large number of false positives: need verification stage in classification

Experiment 1: Consistency in correct errors

Out of 286 correct errors	
Heard by just one person	193 (67%)
Heard by specifically two people	53 (18%)
Heard by all three people	40 (14%)
Heard by more than one	93 (32%)

- About 2/3rds of the data were heard by just one person
- Raises the issue of whether there are differences in the types of errors heard by individual collectors?

Experiment 1: Error levels by collector

	Sound	Word	Other	Total
Listener 1	17 (48.6%)	14 (40.0%)	4 (11.4%)	35
Listener 2	38 (55.9%)	15 (22.1%)	15 (22.1%)	68
Listener 3	89 (61.4%)	40 (27.6%)	16 (11.0%)	145
Listener 4	100 (57.8%)	46 (26.6%)	27 (15.6%)	173

Definitions

- Sound errors: any sublexical error, usually most difficult to detect
- Word word: substitution, deletion, addition, exchange of a word
- Other: morphological, syntactic, morpho-syntactic, marginal categories

Observations

- Percentage of sound errors is broadly consistent, btw 50-61%
- Only outliers are word and other categories

Experiment 1: Salience measures for all errors

	Total	Corrected errors		Gradient errors		Exchanges
Listener 1	34	19	55.9%	4	11.8%	
Listener 2	67	34	50.7%	8	12%	1
Listener 3	145	54	37.2%	25	17.2%	
Listener 4	173	73	42.2%	25	14.5%	1

Observations

- Corrected errors (by speaker) are a red flag to listener; some variation, but listeners with more practice (3, 4) heard more uncorrected errors
- Gradient errors (cf. phonetic errors): also difficult to hear because involve subphonemic structure; good consistency
- Exchanges: only 1 out of 286 errors; very salient

Experiment 1: Salience measures for sound errors

	Total	Errors in stressed syllables		Errors in initial segment	
Listener 1	17	14	82.4%	7	41.2%
Listener 2	38	29	76.3%	13	34.2%
Listener 3	89	73	82%	31	34.8%
Listener 4	100	77	77%	44	44%

Observations

- Errors in stressed syllables are presumably more salient, and also very common; relatively similar percentages (76-82%); still large number of errors heard in unstressed syllables (20%)
- Errors in initial segments are also more common and apparently more salient; 34-44%, with relative consistency

Experiment 1: What errors heard by all?

	Errors heard all three *N* = 40	All errors *N*=286
% corrected by talker	77.5% (31/40)	34.6%
% sound errors	57.5%	58.0%
% gradient errors	2.5% (only one)	17.1%

Examples of uncorrected errors

rock → [b]rock
Right, the **/[b]rock** just gets xxx consumed when it strikes the ground

try → [k]try
… you write grant proposals, you /[k]try to get the time …

technically → [d]echnically
… did come out before Left4Dead /**[d]echnically**, so xxx yeah

photographic –> photogaphic
… to look through **/photogaphic** plates and go oh you know

Experiment 1: Take home messages

1. Many false positives: 25%

Require a second check on error validity and accuracy

2. Data collectors find different errors

Approx. 2/3rds of all errors heard by just one data collector

3. Specific errors differ, but error types and contexts broadly consistent

- Sound errors 50-61%, closer to 60% for experienced listeners
- Corrected errors: 37-56%, closer to 40% for experienced listeners
- Gradient errors: 12-17%
- Errors in stressed syllables: 76-82%
- Errors in initial segments: 34-44%

Experiment 2: Online vs offline

To Do: *I'm going to redo this section with more data; please suggest new measures*

Expectation

- Offline method of collecting speech errors will be more reliable and more representative of the natural occurrence of errors

Method ("online" on-the-spot collection)

- Collectors: SFU linguistics undergraduates, after introduction to linguistics and phonetics and phonology (volunteers proceed into paid positions)
- Training: given one hour introduction to speech errors, including exposure to all major types of errors. Asked to find 1 hour period during their normal daily routine to collect these errors.
- Instructed in best practices (be sure of error, within 30 seconds, make conscious effort)
- Some errors collected at times other than 'set aside time', though typical practice was do this in a set time
- Students submitted raw errors with specific set of variables over email, monitored and vetted by first author

Experiment 2: Error levels by method

	Sound		Word		Other		Totals
Online	97	48.7%	85	42.7%	17	8.6%	199
Offline	166	58.1%	75	26.2%	45	15.7%	286

Observations

- Method had significant effect on levels (x^2 = 16.29, P=0.0003): ratio of sound/word errors changes quite a bit
- Interpretation: offline method enabled listeners to detect more difficult to hear sound and 'other' errors; percentage of word errors much higher with online method, perhaps because more salient

Experiment 2: Salience measures for all errors

	Corrected		Not corrected		Totals
Online	98	49.2%	101	50.8%	199
Offline	99	34.6%	187	65.4%	286

	Exchange errors		Totals
Online	11	5.5%	199
Offline	1	0.4%	286

Observations

- Method also affected the percentage of errors corrected errors (x^2= 9.82, *P*=0.0017): 1/2 for online, cf. 1/3 for offline
- Interpretation: listeners relied more on talker correction for detected errors; likely to actually undershoot this difference because context under-reported in online errors
- Easy to hear exchange errors also have much higher rate with online collection

More on exchanges

	Offline	online
sound	2	31
morpheme		7
word		24
phrase		2
total	719	1092
Total/N	**.28%**	**5.8%**

Much higher percentage of exchanges in online, across all levels.

> Easy to hear errors have higher rate of occurrence in online method; probably skewed, because careful listening of offline method much lower.

Experiment 2: Salience measures for sound errors

	Stressed syllables		Unstressed syllables		Totals
Online	79	81.4%	18	18.6%	97
Offline	131	78.9%	35	21.1%	166

	Error in initial segment		Error not initial segment		Totals
Online	41	42.3%	56	57.7%	97
Offline	71	42.8%	95	57.2%	166

Observations

- Method didn't affect counts in stressed vs. unstressed syllables
- Method did affect counts in initial-non-initial contexts
- Perhaps not the best measures of salience in sound errors; need deeper analysis of error types, features, complexity

Experiment 2: Take home messages

- Online method seems to be skewed towards easier to hear errors (word vs. sound errors, exchanges, corrected errors)
- But some measures of sound errors showed no effect or the opposite effect
- Perhaps need more sensitive measures: features, examine specific sound errors (anticipations vs. perseverations, etc.)

Possible additional questions

- Do online and offline differ in N-gram probability, e.g., *smoke a cigarette*? Expect difference between online and offline, since errors in highly predictable contexts are easier to hear.
- What about differences in the featural differences in sound errors, e.g., voicing (easier) vs. place (harder)?

Your comments here …

Overall advantages of offline method

- Many studies: "**I need more context**"!
- **Greater accuracy** through recourse to a sound recording: can remove false positives and better understand the context of an error
- Can assess **error detection rates** by data collector: minutes per error; can posit threshold on MPE to ensure greater coverage of the errors in a recording
- Errors collected offline likely to **better reflect natural occurrence of errors** in human communication; seems to be easier to detect hard to hear errors
- Supports **phonetic analysis** (prosody, pause durations, segmental acoustics, etc.)
- Easier to find **gradient errors** that fall between discrete phonological categories
- Only real drawback**: can't question speakers** on context, e.g., intended word.

Gradient errors: Motivation for research

Phonological vs. phonetic errors

Growing interest in fine-grained phonetic detail of speech errors (Buchwald, Goldrick, Frisch, Pouplier), but natural speech error corpora don't transcribe this level because it's impossible to do online.

Phonetic analysis as a check on perceptual bias: Frisch and Wright 2002 found z > s errors greater than z > s errors, cf. Stemberger 1991, found the opposite 'anti-frequency bias'; likely that high incidence of s > z due to perceptual bias towards classifying any voicing as a voiced category

Phonetic errors: deviations in some continuous parameter greater than two standard deviations from normal range, but not within normal range of another category; may not be pre-articulatory, speakers may not even be conscious of error (not investigated)

Phonological errors: outside normal range of intended category, and inside range of a different category; prearticulatory, speakers are conscious of the error, but can't help it.

Questions:

- Are phonetic and phonological errors discrete classes (different mechanisms) or do they fall on an continuum? How do phonetic errors differ from phonological errors?
- What factors predict frequency and structure of gradient errors? (Frisch: stability).
- Is it even possible to record gradient errors in spontaneous speech? How frequent are they, and which structures tend to show gradience?

Gradient errors: Coding and pilot results

Transcription conventions

Follow protocol used for child language acquisition (see Stoel-Gammon 2001)

- [A|B] for ambiguous sounds between two poles A and B, but closer to A.
- [A-B] for transitional segments (other than affricates and diphthongs), that move from one structure A to another B
- Intrusion: weak sounds, short in duration, that are clearly audible but don't have the full status of full consonants or vowels
- Use IPA for all transcriptions, not limited to sounds of English

Of 286 errors found in Experiment: 49 (17.1 %) gradient errors

34 ambiguous sounds (mostly vowels, nasal sounds)

12 transitional segments

3 intrusive segments

Projection: 15% of 4,000 = 600 phonetic errors, 10,000 = 1,500 phonetic errors

Out of 1811: 66 ambig, 29 transitional, 3 intrus, 4 other

Gradient errors: Illustrations

Gradient intrusion/epenthesis

GB012: "To everyone, including the **/cam[^{b}]ra** lens" (camera)

RT010: "… to make sure that no one else has ever **/stu[ð]id[^{i}]** fu …" (studied)

Ambiguous segments [X|Y], X Y poles, X is 'closest' category

AC70: "Science is a ^collaborative **/[æ|ɛ]f** xxx effort xxx very few people work in any sort of isolation"

GB012: "I'm hoping you **were /watch[w|r]ing** this when you were a teenager"

GB012: "I will say you are definitely the first guest to **/[k|g]ompare** anything saved by the Bell to either Robert Altman or Paul Thomas Anderson"

Transitional segments [X-Y], starts out as X, moves to Y

AC70: "a night assistant just automatically gets you your data for you and ships it to you either **/o[r-v]er** the internet or perhaps on a DVD"

GB12: "Cuz now they're observing their **/f[eɪ-æ]** xxx fake family, and that's when he's like . . ."

Gradient errors: Future prospects

Likely to get a lot of data

- If pilot results are right, should have several hundred examples of specific structures
- Better training: difficult to instruct data collectors about casual speech and coarticulation (which are habitual, false positives)

Potential differences between phonetic and phonological errors

- Focus on specific structures, e.g., front vowels, s-z, etc.
- Salience measures, e.g., corrected, stress, etc., fully conscious?
- Perceptibility: are phonetic errors recognized as errors like phonological errors?
- Context effects: blends, anticipations and perseverations?
- How fall in continuous parameter space?
- Are the poles in phonetic errors phonologically similar?

Correlate error patterns with non-error patterns

- Compare normal speech of same speaker with errorful speech
- Compare normal articulations of a population (e.g., Buckeye corpus)