Pronunciation models for conversational speech

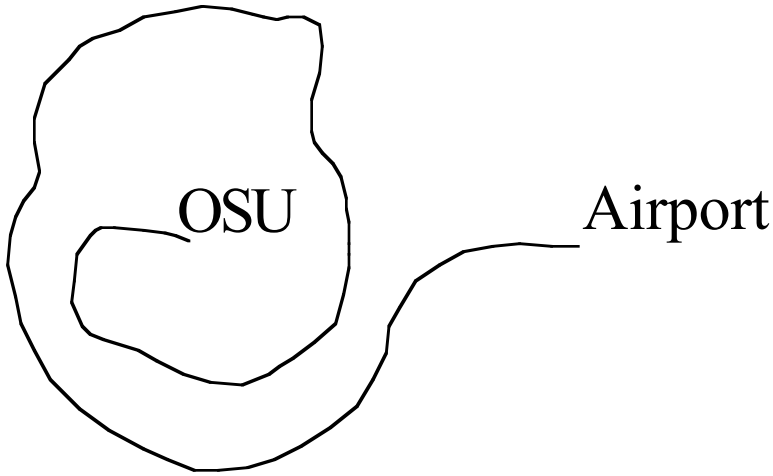
Keith Johnson
Linguistic
UC Berkeley

Using a pronunciation dictionary of clear-speech citation forms we find a segment deletion rate of nearly 12% in a corpus of conversational speech. The number of apparent segment deletions can be reduced by constructing a pronunciation dictionary that records one or more of the actual pronunciations found in conversational speech, however, the resulting "empirical" pronunciation dictionary often fails to include the citation pronunciation form. Issues involved in selecting pronunciations for a dictionary for linguistic, psycholinguistic, and ASR research will be discussed. One conclusion is that Ladefoged may have been the wiser for avoiding the business of producing pronunciation dictionaries.

(Supported by NIDCD grant #R01 DC04330-03).

I met Peter Ladefoged in 1986 or 87 ...

This is the route we took to his hotel in Columbus



Peter stories.

1. Swimming pool & champaign
2. shirt with missing pocket

3. Anglo-Saxon words
4. Cardinal 7 - job interview question
5. *Consonants and Vowels* - chatty radical

An argument against consonants and vowels (for Peter)

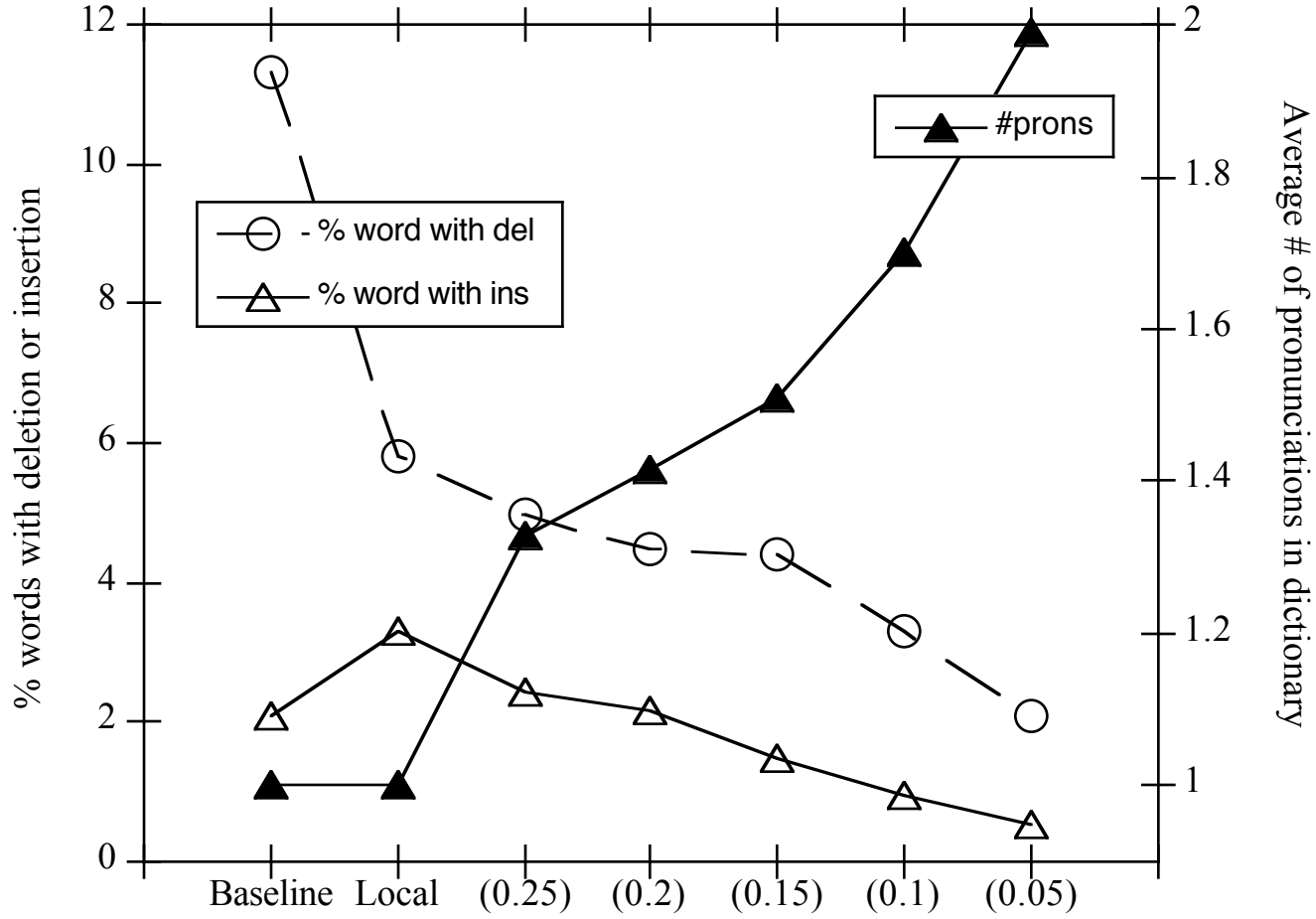
1. Develop a series of "empirical" pronouncing dictionaries

   a. use the ViC (aka "Buckeye") Corpus
      (http://vic.psy.ohio-state.edu)
      - 108,000 word tokens (17 talkers)
      - phonetically transcribed (by phoneticians)

   b. The dictionaries are:
      - Citation dictionary (CMU pronouncing dictionary)
      - Local favorite - most common pronunciation
      - Multiple entry - frequency threshold to enter dictionary - 25% of tokens use the variant, 20%, 15%, 10%, 5%

2. Find deletion and insertion rates
    - relatively massive variation
    - DTW align transcription with dictionary entry
      Johnson (2002) ARLO paper.

3. Measures
    - coverage: # of tokens in corpus that have a deletion or insertion (using best matching entry)
    - complexity: average # of pronunciations per word.


12
10
8
6
4
2
0
2
1.8
1.6
1.4
1.2
1
% words with deletion or insertion
Average # of pronunciations in dictionary
- % word with del
% word with ins
#prons
Baseline
Local
(0.25)
(0.2)
(0.15)
(0.1)
(0.05)

Error rate with local pronunciation remains high
(9% of words have deletion or insertion).

Adding variants to the dictionary reduces error rate
(as it must), but this is a steady slow reduction in error.

The cost of adding pronunciation variants -

- increased ambiguity
- increased number of word hypotheses
  (the exploding lattice problem)

Two questions:

1. What does “two variants per word” mean?

2. What is the relationship between word frequency and number of pronunciation variants?

1. What does it mean to average two variants per word?

Some words have a lot of variants.

Imagine a full coverage dictionary - include all variants. The average number of variants in the full coverage dictionary is only 3.2.

5827 total words in the dictionary
5037 have 5 or fewer variants
3254 have 1 variant
292 have 10 or more variants
27 have 50 or more variants
“yknow” was transcribed 232 different ways.

[y ih n ow]	340 (22.4%)
[y ah n ow]	129 (8.5%)
[y ih nx ow]	96 (6.3%)
[y eh n ow]	68 (4.5%)
[y ih n ah]	60 (4.0%)
[y ahn . ow]	54 (3.6%)
[. ih n ow]	38 (2.5%)
[y . . ow]	33 (2.2%)
[y ih nx ah]	31 (2.0%)
[. ih n ah]	27 (1.8%)
[y . n ow]	27 (1.8%)
[y eh nx ow]	25 (1.6%)
[y ah nx ow]	24 (1.6%)
[y . . ah]	19 (1.3%)
[y ih n ao]	17 (1.1%)
[y ihn . ow]	17 (1.1%)
[. . n ah]	16 (1.1%)
[y ah n ah]	15 (1.0%)
[iy . n ah]	14 (0.9%)
[. . n ow]	13 (0.9%)
[y eh n ao]	13 (0.9%)
[y eh n ah]	13 (0.9%)
[y ah nx ah]	12 (0.8%)
[y ih nx ao	11 (0.7%)
[iy . n ow]	10 (0.7%)
[. . n eh]	8 (0.5%)
[n y ih nx ow]	8 (0.5%)
[y ih n eh]	8 (0.5%)
[y ih . ow]	8 (0.5%)

There were 1516 occurrences of -yknow- with 232 different pronunciations

[n y ih n ow]	7 (0.5%)
[y ah n ao]	7 (0.5%)
[y eh nx ao]	6 (0.4%)
[y eh nx ah]	6 (0.4%)
[ih ah n ow]	6 (0.4%)
[. ih nx ow]	6 (0.4%)
[y uw n ow]	6 (0.4%)
[. eh n ao]	6 (0.4%)
[. eh n ow]	5 (0.3%)
[y iy n ow]	5 (0.3%)
[y . . own]	5 (0.3%)
[. . n y ow]	5 (0.3%)
[. . n y ah]	5 (0.3%)
[y ah nx ao]	5 (0.3%)
[y ih . ah]	4 (0.3%)
[. ih n eh]	4 (0.3%)
[y ihn n own]	4 (0.3%)
[. ih nx ah]	4 (0.3%)
[y iyn n own]	4 (0.3%)
[n y ah n ow]	4 (0.3%)
[hh ih n ow]	4 (0.3%)
[y . nx ah]	4 (0.3%)
[iy . n eh]	4 (0.3%)
[. ih n ao]	4 (0.3%)
[y . n ah]	4 (0.3%)
[y ahn . .]	4 (0.3%)
[y uw nx ow]	4 (0.3%)
[y ah . ow]	4 (0.3%)
[. . en ow]	3 (0.2%)
[y ah . ah]	3 (0.2%)
[y . . ao]	3 (0.2%)
[sh ih n ow]	3 (0.2%)
[y uh nx ow]	3 (0.2%)
[. . en ah]	3 (0.2%)
[y uh n ow]	3 (0.2%)
[. eh n ah]	3 (0.2%)
[y iy n ah]	3 (0.2%)
[n ih n ow]	3 (0.2%)
[y ih nx eh]	3 (0.2%)
[y ahn . ah]	3 (0.2%)
[y ih ahn .]	3 (0.2%)
[. . n ih ah]	2 (0.1%)
[n y ih nx ah]	2 (0.1%)
[y eh n ah n]	2 (0.1%)

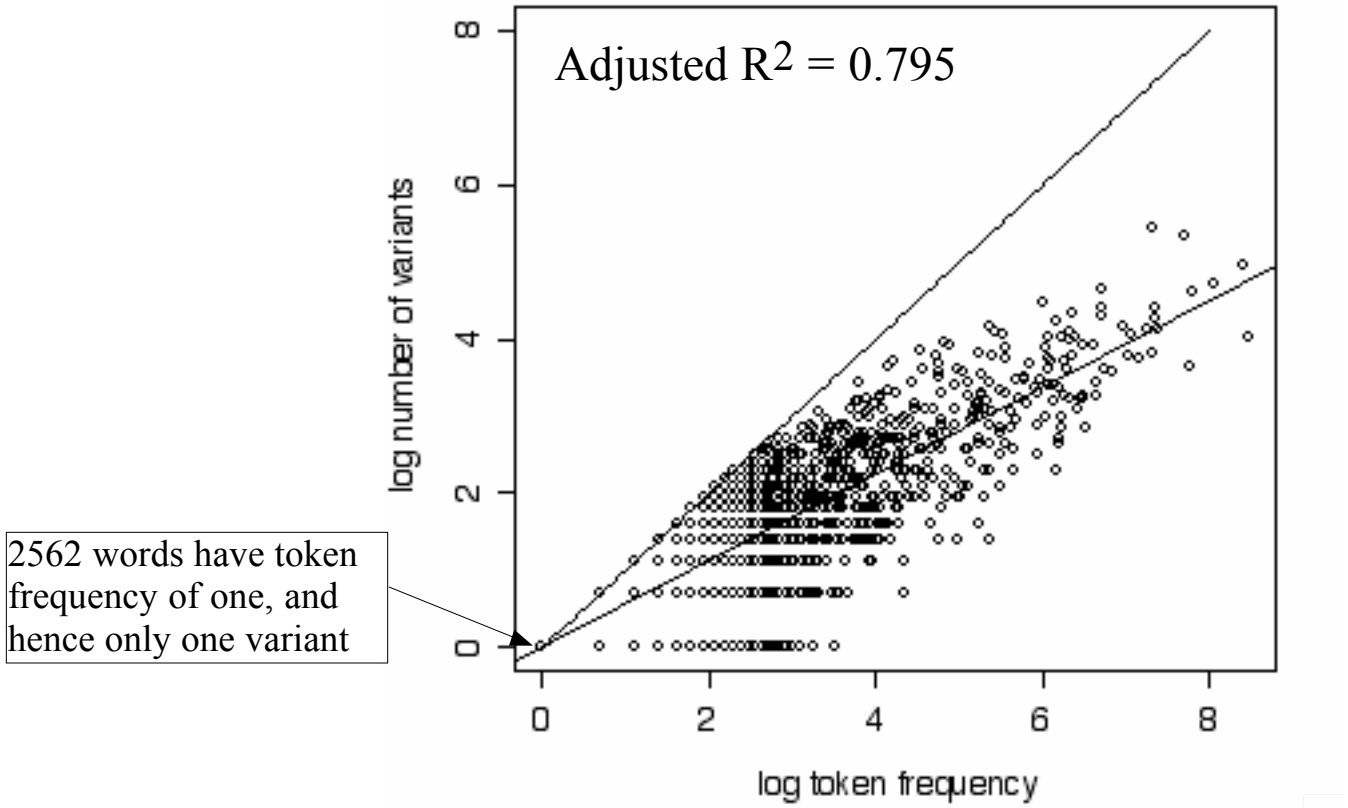
The 27 words that have 50 or more variant pronunciations

about, and, because, but, didn't, don't, everything, gonna, I, in, it, just, of, something, that, that's, the, think, to, was, well, what, when, with, yeah, yknow, you

High frequency words

2. Word frequency and the number of variants.

$$\ln(\#\text{variants}) = 0.56 * \ln(\text{freq})$$



What do listeners do?

1. Listen for consonants and vowels - look up words in the mental lexicon the way you look up a word in a dictionary.

- must list lots of pronunciations for frequent words.
- recognition lattice explodes (with abstract segments)

2. Listen for words
(alá Ladefoged in *Consonants and Vowels*).

- word-form matching provides redundancy
- a way out of the exploding lattice problem

Also (assuming detailed exemplars) predicts:

- word-specific phonetics
- lexical defusion of sound change