Measuring Pronunciation Differences

- What to measure
  - Segments
  - Syllables
  - Words
  - Phrases
- How to measure
  - Nominal/Catgorical
  - Numerical

It's possible to use a nominal/categorical measure (Séguy, Goebl) but then more complex units are almost always different. Challenge: how to define a numerical measure?

We're particularly interested in techniques for measuring differences in the pronunciation of comparable material—the form most dialect atlases have.

Segment distances

- Phones
  - Two segments are equal or different.
  - Distance between [ɪ] and [e] the same as between [ɪ] and [ɒ].
  - Rough, but easy to operationalize!
  - Rough measures reliable with large data sets.
- More sensitive measures later
- Challenge: How to lift segment distances to STRING DISTANCES.

String distances

- *Levenshtein distance* calculates the (least) cost of changing one string into another
- Example: *afternoon* is pronounced as [ˈæəftəˌnʉˑn] in the dialect of Savannah and as [ˌæftərˈnuˑn] in the dialect of Lancaster.

æəftənʉn	delete ə	1
æftənʉn	insert r	1
æftərnʉn	subst. ʉ/u	1
æftərnun		
		3

- All operations cost *one* unit (initially)
- Problem: how to guarantee the *least* cost

Levenshtein algorithm(æəftənʉn,æftərnun)

Create two-dimensional array

		æ	f	t	ə	r	n	u	n
æ									
ə									
f									
t									
ə									
n									
ʉ									
n									

Algorithm

Levenshtein distance(æəftənʉn,æftərnun)

		æ	f	t	ə	r	n	u	n
	0								
æ									
ə									
f									

- begin at upper left ($\Leftarrow 0$)
- to fill in a cell:

diag	above
left	min(above + delete, diag + replace, left + insert)

Algorithm

Levenshtein distance(æəftənʉn,æftərnun)

		æ	f	t	ə	r	n	u	n
	0	1	2	3	...				
æ	1	?							
ə	2								
f	3								
t	⋮								

Top horizontal row is always $1, 2, \ldots$ —cost of insertions

Left vertical column is always $1, 2, \ldots$ —cost of deletions

? is minimum(left+ins,above+del,diag+subst)

minimum$(1 + 1, 0 + 0, 1 + 1)$

Algorithm

Levenshtein distance(æəftənʉn,æftərnun)

		æ	f	t	ə	r	n	u	n
	0	1	2	3	...				
æ	1	0	$?_1$						
ə	2	$?_2$							
f	3								
t	⋮								
ə									
n									
ʉ									
n									

Algorithm

Levenshtein distance(æəftənʉn,æftərnun)

		æ	f	t	ə	r	n	u	n
	0	1	2	3	...				
æ	1	0	1						
ə	2	1	$?_1$						
f	3	$?_2$	$?_3$						
t	⋮								
ə									
n									
ʉ									
n									

Algorithm

Levenshtein distance(æəftənʉn,æftərnun)

		æ	f	t	ə	r	n	u	n
	0	1	2	3	...				
æ	1	0	1						
ə	2	1	1						
f	3	2	1						
t	⋮								
ə									
n									
ʉ									
n									

Algorithm

Levenshtein distance(æəftənʉn,æftərnun)

		æ	f	t	ə	r	n	u	n
	0	1	2	3	4	…			
æ	1	0	1	2					
ə	2	1	1						
f	3	2	1						
t	4			1					
ə	⋮				1				
n									
ʉ									
n									

Algorithm

Levenshtein distance(æəftənʉn,æftərnun)

		æ	f	t	ə	r	n	u	n
	0	1	2	3	4	...			
æ	1	0	1	2					
ə	2	1	1						
f	3	2	1	1					
t	4			1					
ə	⋮				1	2			
n					2		2		
ʉ								3	
n									3

- lower right corner contains Levenshtein distance, cost of least expensive set of transformations

Alignment

Levenshtein distance(æəftənʉn,æftərnun)

		æ	f	t	ə	r	n	u	n
	0	1	2	3	...				
æ	1	0	1	2					
ə	2	1	1						
f	3	2	1	1					
t	⋮			1					
ə					1	2			
n					2		2		
ʉ								3	
n									3

RuG

Alignment

Levenshtein distance(æəftənʉn,æftərnun)

		æ	f	t	ə	r	n	u	n
	0	1	2	3	...				
æ	1	0	1	2					
ə	2	1	1						
f	3	2	1	1					
t	⋮			1					
ə					1	2			
n					2		2		
ʉ								3	
n									3

æ	ə	f	t	ə	∅	n	ʉ	n
æ	∅	f	t	ə	r	n	u	n

Dialect distances

- Many sequence operations map [æəftənʉn] $\longrightarrow$ [æftərnun]. Levenshtein distance = cost of cheapest mapping.
- Using w words the distance between two dialects is equal to the average of w Levenstein distances.
  - automatically weights differences involving more frequent sounds more heavily
- All distances between n dialects are arranged in a $n \times n$ matrix for further analysis.

Pronunciation Difference Measurement

- Rough (noted)
- Sensitive to stress (if mark occurs in transcription)
- Contextual effects *are* measured

t	u	$	b	æ	t	ə	z	*two batters*
t	u	$	b	æ	ɾ	ɚ	z	

- Sources of contextual effects are *ignored*
  if [t/ɾ] occurs in same ratio in two pairs of varieties, distance will be the same—regardless of specifics of conditioning
- Longer words contribute more heavily to pronunciation difference since their string distances are longer

Short Words vs. Long Words

- The simple Levenshtein distance can be normalized by the length of the word. For example, the sum of the operations is divided by the length of the longest alignment which gives the minimum cost. The longest alignment has the greatest number of matches.
- Example:

æ	ə	f	t	ə	∅	n	ʉ	n
æ	∅	f	t	ə	r	n	u	n
	1				1		1	

A total cost of 3 divided by a length of 8 gives a word distance of 0.38 or 38%.

- Using 125 words the distance between two dialects is equal to the average of 125 Levenstein distances.
- All distances between n dialects are arranged in a n $\times$ n matrix.

More Sensitive Segment Distances

- Levenshtein distance is also known as string distance and edit distance. Well-known example of DYNAMIC ALGORITHM (yet another name).
- Applications in other areas

  software engineering file differences

  bioinformatics differences between long strings of amino acids (DNA)

  translation aligning bilingual corpora

  ethnology tracking "folk processing" in bird calls
- Costs are often one for insertions and deletions, two for substitutions
  - seems wrong in assaying pronunciation dissimilarity

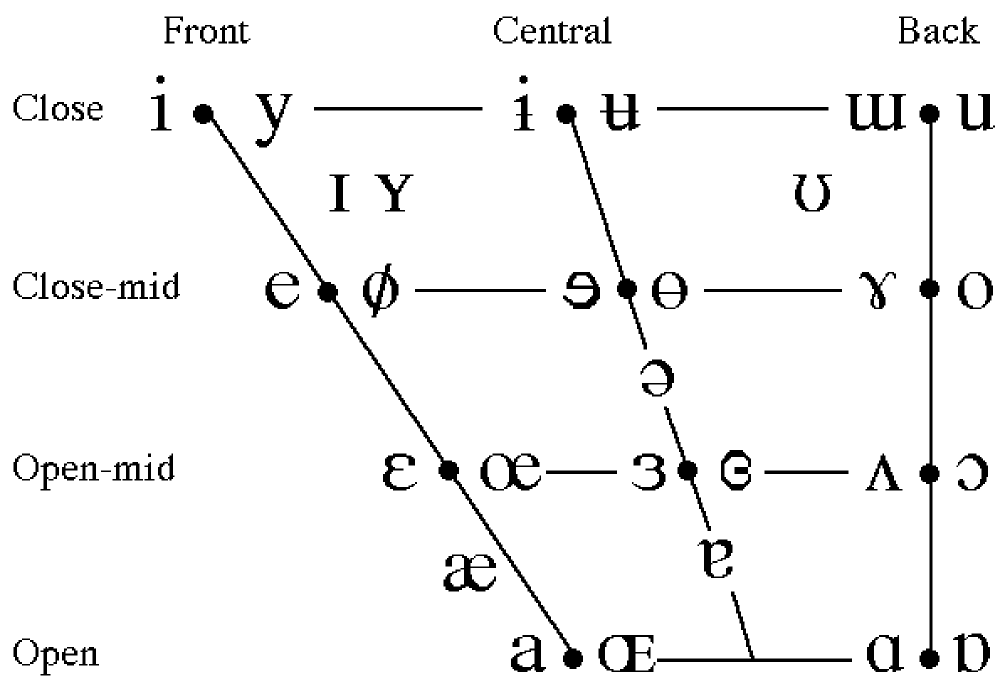
More Sensitive Segment Distances

- Phones
  - Two segments are equal or different.
  - Rough, but simple!
- Features
  - Finer diferentiation of segment distances.
  - Segment differences used to weight Levenshtein algorithm
  - Distance between [ɪ] and [e] smaller than between [ɪ] and [ɒ].
  - Distance between two bundles: sum of differences (simplest case).
    * Heeringa (2004) experimented with Euclidean combination, a $(1 - r)$ measure (r – Pearson's correlation coefficient) with little effect.

Dialect distances

- Refinement: feature bundle distances or acoustic distances as operation weights!
- We assure that the minimum cost is based on a alignment in which
  - a vowel matches with a vowel
  - a consonant matches with a consonant
  - the [j] or [w] matches with a vowel
  - the [i] or [u] matches with a consonant
  - the schwa matches with a sonorant

Discrete segment distances



Vowel distances in the Almeida & Braun system: distances of 1 point:
ɛ vs. æ (height), ɛ vs. ɜ (advancement), ɛ vs. œ (round).

Feature-based Segment Distances

Positive More sensitive distinctions

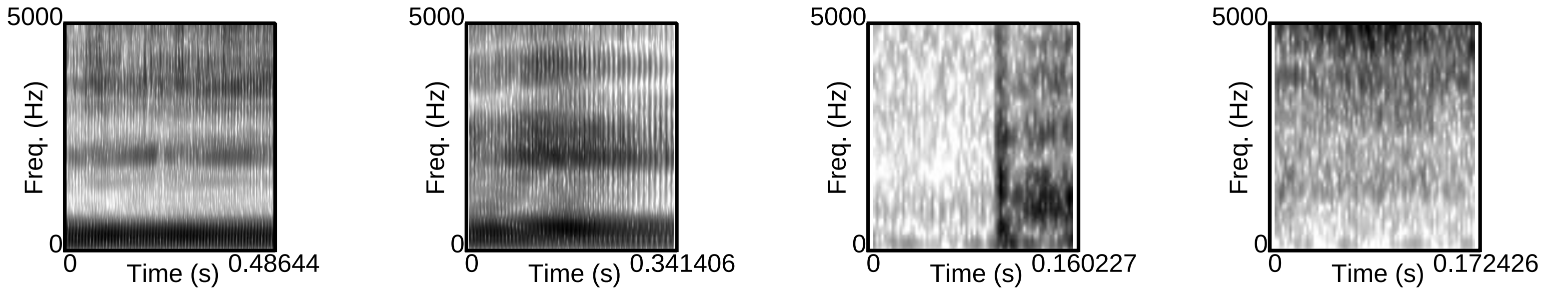
- In theory should yield no worse measurements

Negative

- Embarassment of riches (may systems)
- Many more parameters (therefore weaker)
- Most feature systems developed to facilitate phonological description, not to provide foundation for description of dialectal similarity

Acoustic segment distances

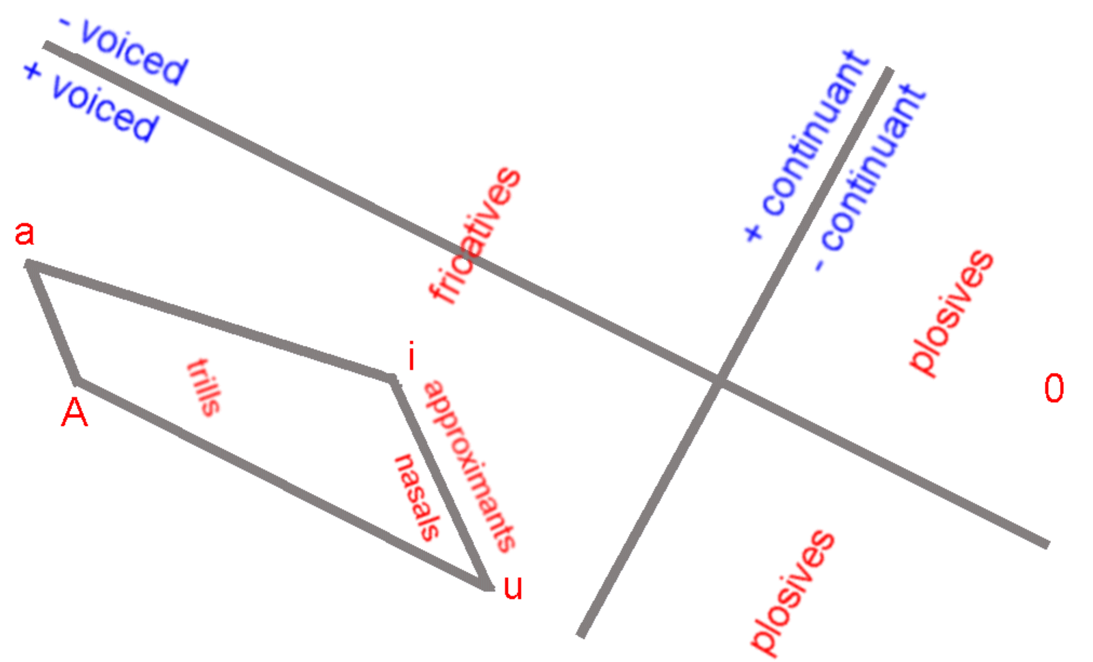
- Feature systems mostly not based on physical measurements.
- Samples of all IPA segments are found on the audio tape *The Sounds of the International Phonetic Alphabet* (1995).
- Calculate distances between the samples using their spectrograms or formant tracks.
- Intensity is processed, durations are made equal.



i e p s

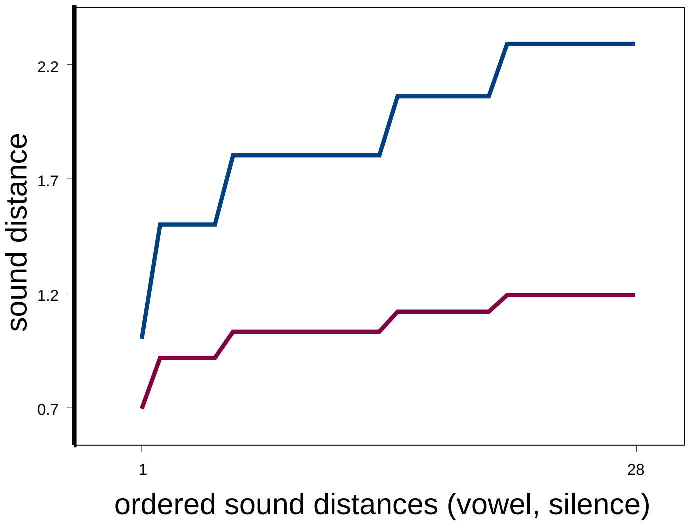
RuG

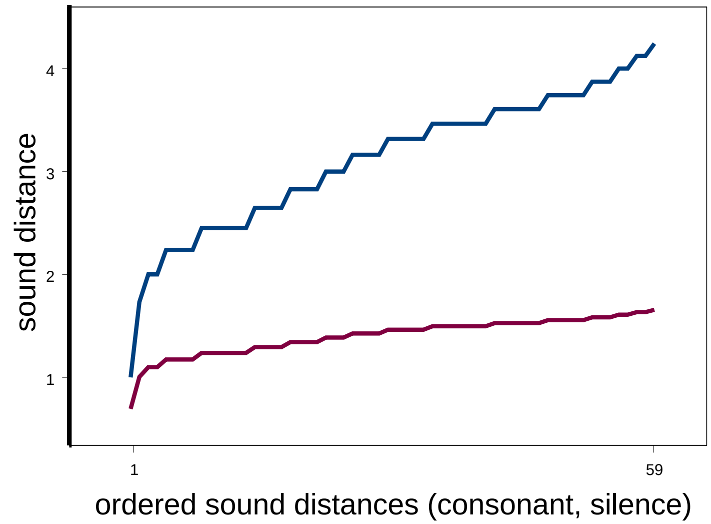
Acoustic segment distances (Heeringa)



Distances among 88 segments (28 vowels, 59 consonants, silence) calculated using the Barkfilter and reduced from 88 dimensions to two dimensions with multidimensional scaling.

Linear and logarithmic segment distances





Linear (upper) and logarithmic (lower) Almeida & Braun distances of 28 IPA vowels (left) and 59 IPA consonants (right) with respect to silence. Distances are sorted from low (left) to high (right). Greater distances are reduced more than smaller ones by using the logarithm.

Pronunciation Differences

- Given database of pronunciations of comparable material, we can obtain various measures of pronunciation difference.
- Since we'll characterize the distance(s) numerically, we can analyze the results using numerical techniques.
- Levenshtein distance appropriate for dialect atlas material with comparable pronunciations indicated, but inappropriate for corpus material, i.e. material without indication of which pronunciations are to be compared.

Next Steps

- LAMSAS pronunciations
  - No mapping to acoustic samples (too complex)
  - No logarithmic correction to flatten large diferences
- Quality of results