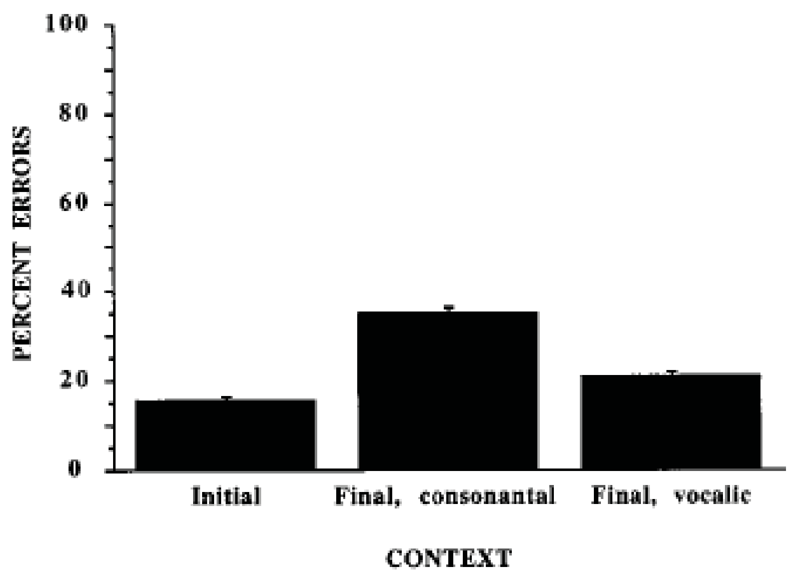
24.915/24.963 Linguistic Phonetics

Perception II: Distinctiveness and cue strength



Source: Redford, Melissa A., and Randy L. Diehl. "The relative perceptual distinctiveness of initial and final consonants in CVC syllables." The Journal of the Acoustical Society of America 106, no. 3 (1999): 1555-1565.

Readings for next week - Speech production:

- Keating (1990)
- Browman and Goldstein (1990)

Assignment:

- Recording and measurements for affricate voicing experiment.
- Talk to me about your final project

- Sorry, no office hours tomorrow (Wed 10/4)

Perceptual distinctiveness in phonology

- Perceptual distinctiveness plays a central role in recent phonological theory.
  - Related concept of cue strength: stronger cues to a contrast imply that the contrast is more distinct.
- There is hypothesized to be a general preference for more distinct (less confusable) contrasts:
  - Theory of Adaptive Dispersion – vowel inventories
  - Distribution of retroflex vs. apical alveolar contrasts (1st lecture)
  - Distribution of major place contrasts

Distribution of retroflexion contrasts in Gooniyandi

Summary:

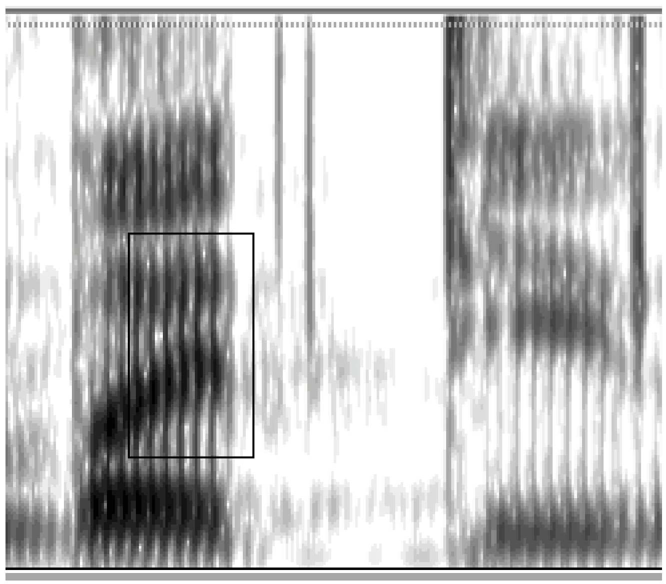
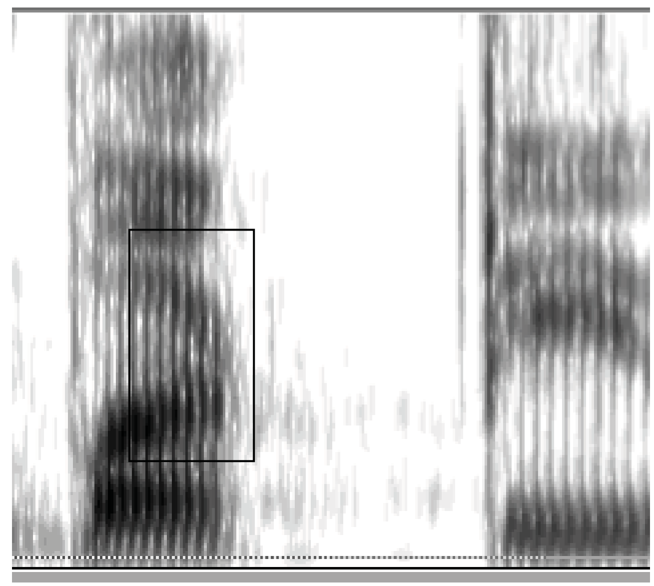
- Contrast between retroflex and apical alveolar **after vowels** V_#, V_V, V_C
- No contrast elsewhere #_, C_
- This pattern of distribution is common in Australian and Dravidian languages.
- Probable universal: If a language contrasts retroflexes and apical alveolars in contexts with no preceding vowel, then it also contrasts these sounds following vowels.
  - ➢ Why?

Distribution of retroflexion contrasts

Outline explanation (Steriade 1995, 2001 etc):

- Contrasts preferentially appear in environments where they are more distinct perceptually.
- Apical alveolars are more distinct from retroflexes when they are preceded by a vowel.
- Therefore some languages only allow the contrast in this context.
- The primary cues to the contrast between retroflexes and apical alveolars are located in the VC transitions
  - lowered F3 and/or F4 before retroflexes
- Most retroflex stops are retroflexed at closure, but the tongue tip moves forward during closure.

The phonetics and phonology of retroflex consonants

Courtesy of Ashtu Killimangalam. Used with permission.

apical alveolar [t]　　　　retroflex [ʈ]

Malayalam

Neutralization of major place contrasts

	_V (_L)	**_#**	**_(N) _T (_F)**
Spanish Japanese		neutralization	assimilation
Diola Fogny			assimilation
Russian			

V = vowel; L= glides & liquids; N = nasals, T = stops, F = fricatives

Jun (1995), DeLacy (2002), Steriade (2001)

Neutralization of major place contrasts

- E.g. Spanish nasals
  - Place contrast prevocalically

    mata, nata, ɲata

    kama, kana, kaɲa
  - Neutralization word-finally - pre-pausally, the only nasal is [n] in Castilian, [ŋ] in many other varieties.

    adan 'Adam' albun 'album'
  - Neutralization before obstruents (assimilation)

kampo	'country'	*kanpo, etc
manto	'cloak'	*mamto, etc
maŋko	'one-handed'	*manko, etc

Major place neutralization

Diola Fogny nasals (Sapir 1957)

- Place contrasts before vowels [m, n, ɲ, ŋ]
- Contrast word-finally (after a vowel):

ɹum	'bite'	famfan	'lots'	buɹuŋ	'road'

- Neutralization before consonants - nasal must be homorganic with a following obstruent of nasal (medial or final)
  - Nasals delete before approximants.

kəguːmp	'ashes'	bunt	'lie'
kaŋg	'be furthest away'	maɲɟ	'know'

/ni-gam-gam/	→	niga**ŋg**am	'I judge'
/ku-bɔɲ-bɔɲ/	→	kubɔ**mb**ɔɲ	'they sent'
/na-tiːŋ-tiːŋ/	→	natiː**nt**iːŋ	'they sent'

Major place neutralization

- Pattern 3: E.g. Russian

anglijə	'England'	mgla	'fog'
gbaronu	'the baron'	kto	'who
tkan^{j}ja	'weaving'		

Neutralization of major place contrasts

	_V (_L)	_#	_(N) _T (_F)
Spanish Japanese		neutralization	assimilation
Diola Fogny			assimilation
Russian			

V = vowel; L= glides & liquids; N = nasals, T = stops, F = fricatives

Jun (1995), DeLacy (2002), Steriade (2001)

Major place neutralization

Unattested language type:

- Place contrasts / _ V: ma - na
- Neutralization /V_#: an, *am
- Contrast / _ C: anta – amta

Another unattested language type:

- Neutralization / _ V: na, *ma
- Contrast /_C: anta – amta
    - NB specific place contrasts may be permitted finally but not initially, e.g. English velar nasal [æm – æn - æŋ], but [mæ-næ- *ŋæ]

Major place neutralization

- Preferred environments for major place contrasts are contexts where the contrasting sounds are more distinct (Steriade 1999 etc).

_V	pa - ka	ma - na
	>	>
V_#	ap - ak	am - an
	>	>
V_T	ap˺ta - ak˺ta	amta - anta

- Evidence?

Measuring perceptual distinctiveness

- Perceptual distinctiveness is related to confusability: less distinct sounds are more confusable.
- One way to measure confusability is via an identification task:

- Play a number of pre-vocalic consonants [pa, ta, …] and post-vocalic consonants [ap, at,…] that differ in place of articulation.
- Ask subjects to label the stimuli (words, orthographic nonce words).
- Observe rates of confusion (e.g. stimulus [ap] is identified as [at]).
- Is place confused more often in the post-vocalic context?

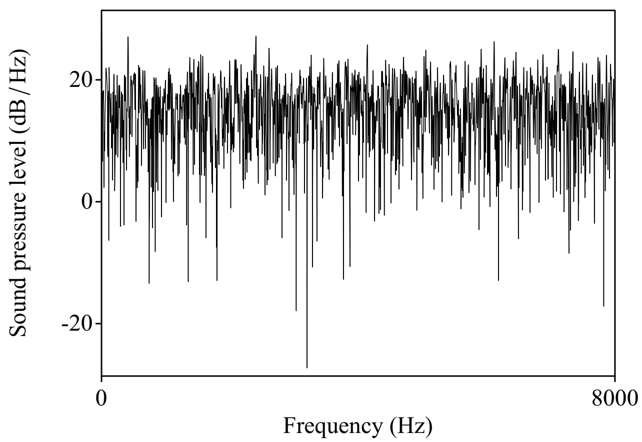
Redford & Diehl (1999)

- Compares the confusability of a variety of consonants (including [p,t,k]) in prevocalic and post-vocalic contexts.
- An identification experiment with natural stimuli.
- Stimuli: CVC syllables
  - All combinations of C = [p, t, k, f, θ, s, ʃ], V = [i, a, u] (7×3×7).
  - Frame sentences ‘Say CVC some more’, ‘Say CVC again’
    - 3 conditions: #<u>C</u>V, V<u>C</u>#C, V<u>C</u>#V
  - Read by two male and two female speakers of American English.
  - Presented in a low level of pink noise (15 dB SNR).
  - Noise is often added to stimuli in studies of confusability in order to encourage errors.

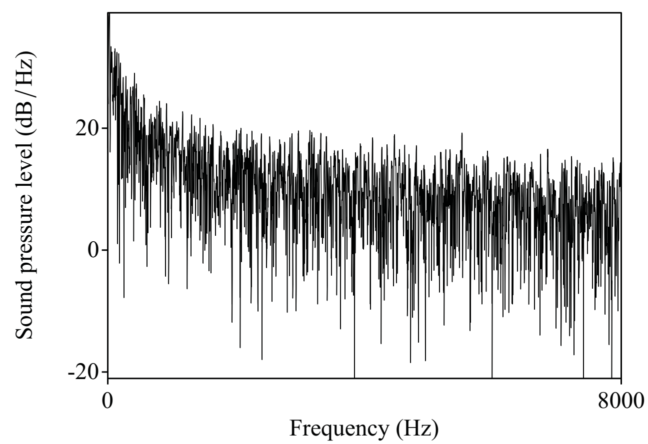
Addition of noise

- Varieties of noise:
  - White noise - flat spectrum
  - Pink noise - spectrum slopes down at 3dB/octave.
  - ‘Cocktail party’ noise, a.k.a. multi-talker babble

white



pink



- Level specified as Signal-to-Noise Ratio - difference between average intensity of speech and noise in dB.
  - Examples have 0dB SNR, Redford & Diehl used 15 dB SNR.

Subjects & Procedure

- 7 female, 7 male, native speakers of American English
- Stimuli presented over headphones.
- Subjects heard each stimulus once in random order at 3.5 s intervals.
- Subjects asked to record the target words in orthography.

Results

- Confusion matrices

- #<u>C</u>V

	p	t	k	f	θ	s	ʃ	Other	None	
p	**2295**	15	23	9		1			9	
t	6	**2254**	5	24	3	1	1	5	53	T
k	39	14	**2251**	4	6	2		2	34	a
f	384	24	20	**1678**	113	65		14	54	r
θ	129	44	12	670	**1274**	197	3	7	16	g
s	10	25	1	84	87	**2106**	27	3	9	e
ʃ	1		2	2	3	38	**2302**	2	2	t

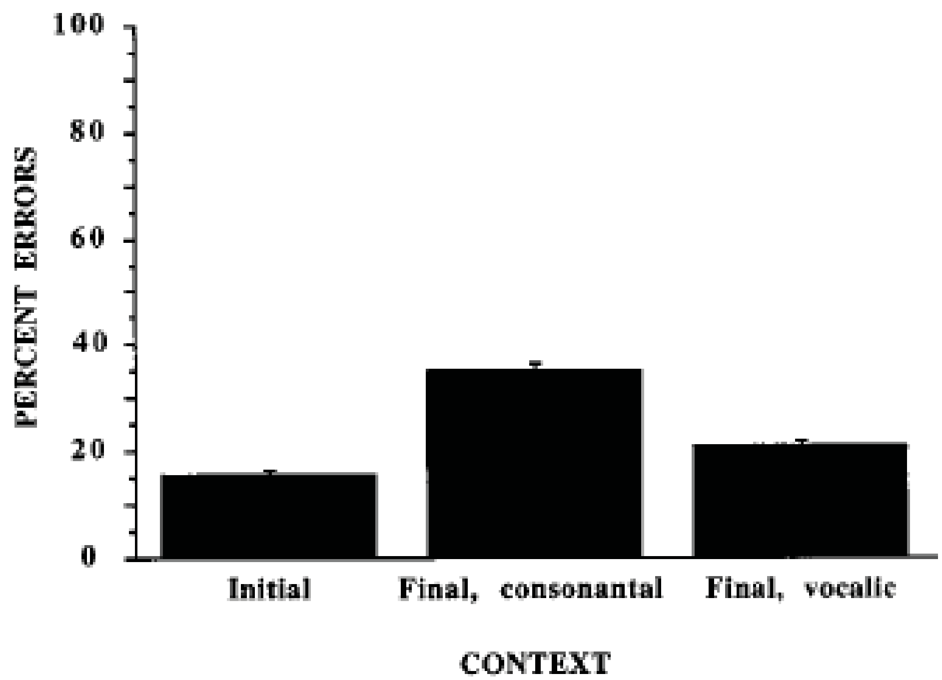
- V<u>C</u>#C

	p	t	k	f	θ	s	ʃ	Other	None	
p	**1007**	37	38	32	42	2		15	3	
t	76	**950**	38	11	72	2		20	7	T
k	32	13	**1063**	15	14	1		30	8	a
f	431	35	90	**366**	184	19	6	37	8	r
θ	249	52	104	122	**589**	23		23	14	g
s	17	29	9	41	180	**748**	81	30	41	e
ʃ	1			3	7	31	**1119**	11	4	t

Source: Redford, Melissa A., and Randy L. Diehl. "The relative perceptual distinctiveness of initial and final consonants in CVC syllables." The Journal of the Acoustical Society of America 106, no. 3 (1999): 1555-1565.

Results

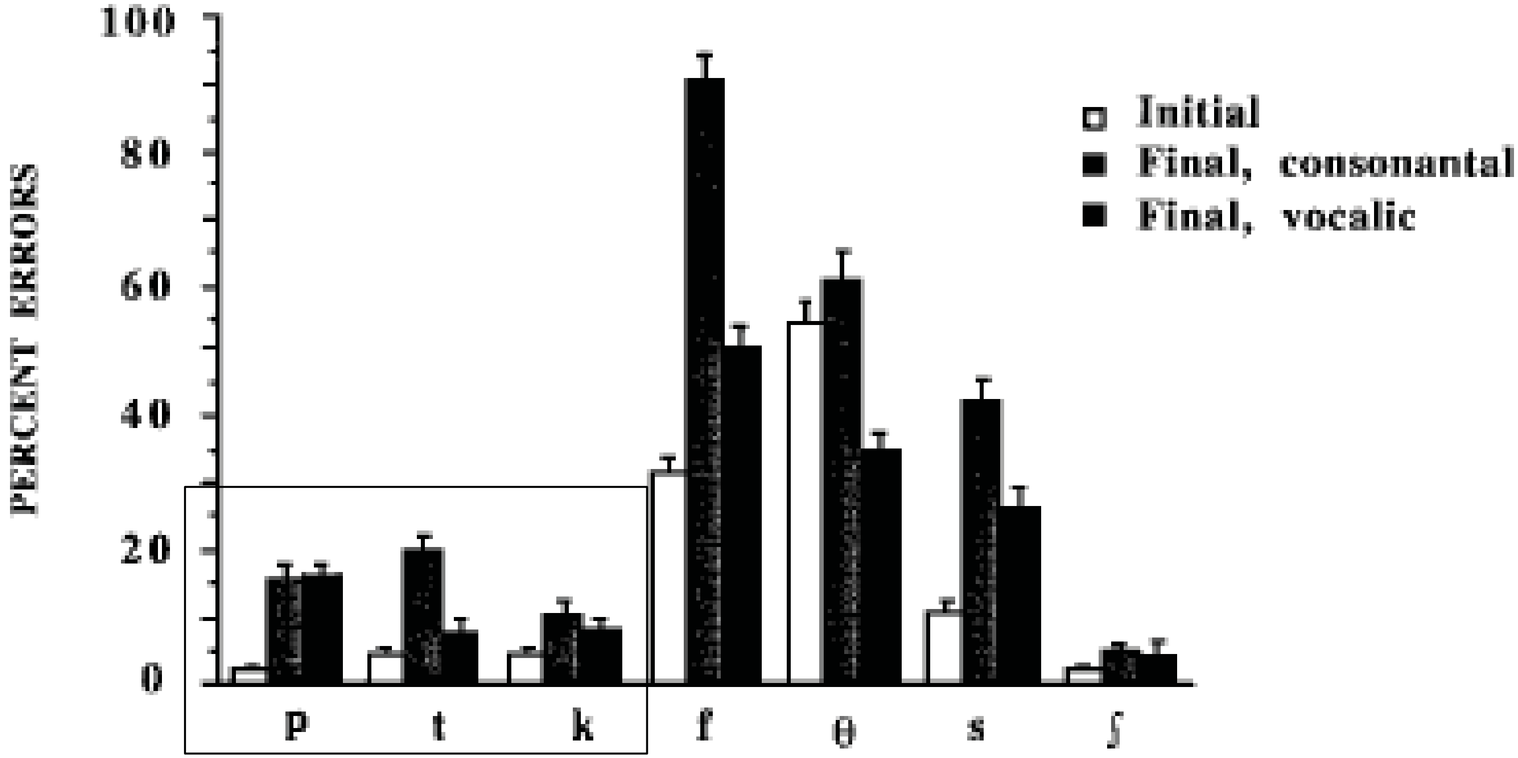
- error rate - percentage of erroneous identifications.
  - only manner and place errors were counted, not voicing errors.



Source: Redford, Melissa A., and Randy L. Diehl. "The relative perceptual distinctiveness of initial and final consonants in CVC syllables." The Journal of the Acoustical Society of America 106, no. 3 (1999): 1555-1565.

Redford & Diehl (1999)

- Obstruent contrasts in place and manner are more distinct before vowels than before consonants.



Source: Redford, Melissa A., and Randy L. Diehl. "The relative perceptual distinctiveness of initial and final consonants in CVC syllables." The Journal of the Acoustical Society of America 106, no. 3 (1999): 1555-1565.

Major place neutralization

- What explains the differences in distinctiveness of place contrasts across these contexts?

_V	pa - ka	burst, release transitions
	>	
V_#	ap - ak	burst, closure transitions
	>	
V_T	ap˺ta - ak˺ta	closure transitions

- More cues in V_# compared to V_T (for stops)
- stronger cues _V compared to V_#
  - Release transitions provide stronger cues than closure transitions (Fujimura et al 1978)

Cue strength

- Measuring cue strength
  - vary multiple cues, model contribution of each cue to perceptual judgments (cf. VOT and F0)
  - conflicting cue experiments: construct stimuli where different properties cue different percepts (e.g. voiced, voiceless).
    - The stronger cue is the one that dominates in perception.
- Fujimura et al (1978) employed the 'conflicting cue' methodology to investigate the relative strength of cues to stop place in VCV.
- Closure transitions vs. Release burst+release transitions
  - E.g. cross-spliced [ab-] from [aba] and [-da] from [ada]
    - Do listeners perceive [aba] or [ada]? (Example)
  - Stimuli created from Japanese utterances.

Conflicting cue experiments - Fujimura et al

- Results: Release cues dominate for English and Japanese listeners

Figure removed due to copyright restrictions.
Source: Tables 2 & 3, Fujimura, Osamu, Marian J. Macchi, and Lynn A. Streeter. "Perception of stop consonants with conflicting transitional cues: A cross-linguistic study." Language and speech 21, no. 4 (1978): 337-346.

- Is this simply because burst+transitions > transitions?
- No: the same result was obtained when the stimuli were played backwards

Figure removed due to copyright restrictions.
Source: Tables 2 & 3, Fujimura, Osamu, Marian J. Macchi, and Lynn A. Streeter. "Perception of stop consonants with conflicting transitional cues: A cross-linguistic study." Language and speech 21, no. 4 (1978): 337-346.

Major place neutralization

- What explains the differences in distinctiveness of place contrasts across these contexts?

_V	pa - ka	burst, release transitions
	>	
V_#	ap - ak	burst, closure transitions
	>	
V_T	ap̚ta - ak̚ta	closure transitions

- burst + release transitions > burst + closure transitions because release transitions provide stronger cues than closure transitions.
- Why is pre-obstruent context worse than word-final position for nasal place contrasts?
  - Overlap with the following consonant?

Perceptual space

- Identification experiments based on synthetic/edited speech can establish the perceptual significance of an acoustic property
- But they do not reveal the nature of perceptual representations
  - The dimensions of the perceptual space
- E.g. VOT is a cue to stop voicing, but is there a perceptual dimension that corresponds directly to VOT, or just something that correlates with VOT?
  - E.g. integrated intensity of aspiration noise (cf. Repp 1979)
- Multi-Dimensional Scaling is a method for investigating perceptual space directly.
  - But it does not reveal the mapping from acoustic signal to perceptual space.

Perceptual space - vowel quality

- The main dimensions of the perceptual space of vowel quality are related to the frequencies of the first two or three formants.
- Acoustic analysis shows that contrasting vowel qualities differ in formant frequencies.

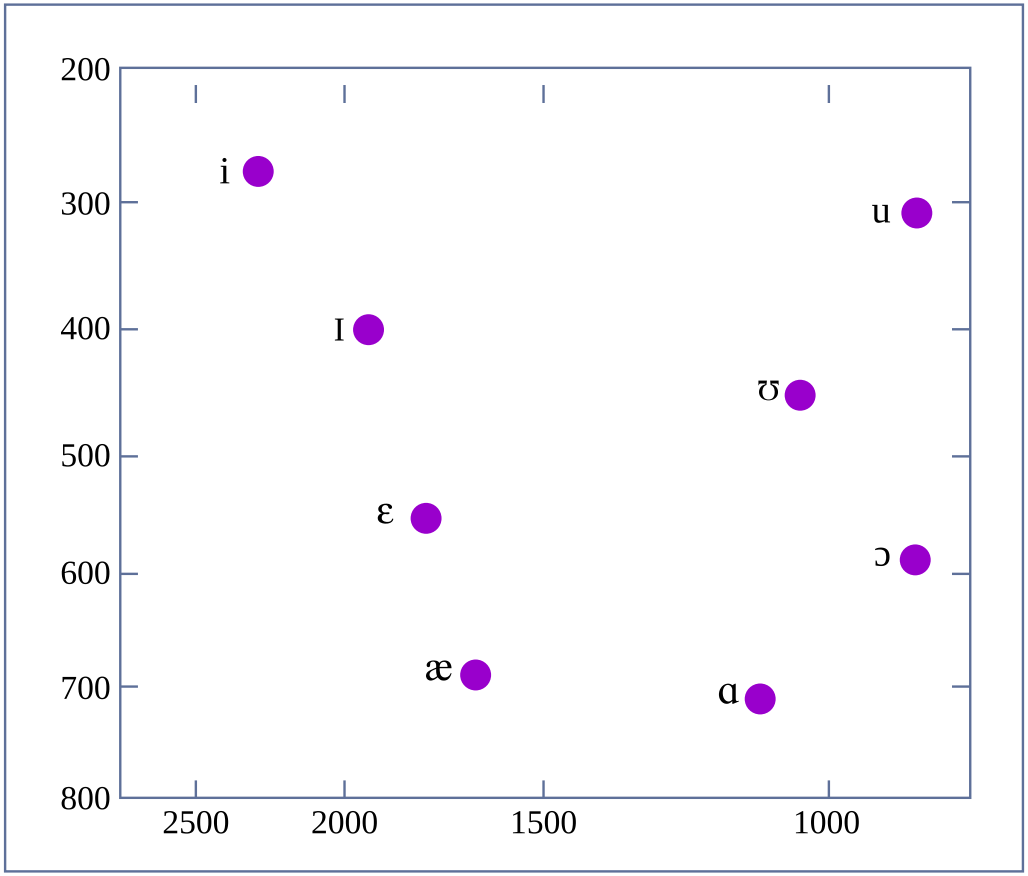


Image by MIT OCW.
Adapted from Peter Ladefoged. *A Course in Phonetics*. 5th ed. Berlin, Germany: Heinle, 2005. ISBN: 9781413006889. Available at: http://www.phonetics.ucla.edu/course/contents.html.

Perceptual space - vowel quality

- Perceived vowel quality is affected by formant frequencies.
- A wide range of vowel qualities can be synthesized with two formants (Delattre, Liberman, Cooper, and Gerstman 1952).
- Multi-Dimensional Scaling (MDS) analyses of vowel confusions and similarity judgements yield spaces in which the most significant dimensions correspond to F1 and F2.

Multi-Dimensional Scaling

- Techniques for constructing perceptual spaces from confusion or similarity data (Shepard 1957, 1972).

Multi-Dimensional Scaling

- Input: a confusion matrix
- e.g. Peterson & Barney 1952

perceived

intended		i	j
	i	p_{ii}	p_{ij}
	j	p_{ji}	p_{jj}

perceived

intended

	i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɹ
i	10267	4	6			3				
ɪ	6	9549	694	2	1	1				26
ɛ		257	9014	949	1	3			2	51
æ		1	300	9919	2	2			15	39
ɑ		1		19	8936	1013	69		228	7
ɔ			1	2	590	9534	71	5	62	14
ʊ			1	1	16	51	9924	96	171	19
u			1		2		78	10196		2
ʌ		1	1	8	540	127	103		9476	21
ɹ			23	6	2	3			2	10243

Multi-Dimensional Scaling

- Input: a confusion matrix
- e.g. Peterson & Barney 1952
- convert to probabilities

perceived

intended		i	j
	i	p_{ii}	p_{ij}
	j	p_{ji}	p_{jj}

perceived

intended

	i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɹ
i	0.9987	0.0004	0.0006	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000
ɪ	0.0006	0.9290	0.0675	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0025
ɛ	0.0000	0.0250	0.8771	0.0923	0.0001	0.0003	0.0000	0.0000	0.0002	0.0050
æ	0.0000	0.0001	0.0292	0.9651	0.0002	0.0002	0.0000	0.0000	0.0015	0.0038
ɑ	0.0000	0.0001	0.0000	0.0018	0.8699	0.0986	0.0067	0.0000	0.0222	0.0007
ɔ	0.0000	0.0000	0.0001	0.0002	0.0574	0.9275	0.0069	0.0005	0.0060	0.0014
ʊ	0.0000	0.0000	0.0001	0.0001	0.0016	0.0050	0.9655	0.0093	0.0166	0.0018
u	0.0000	0.0000	0.0001	0.0000	0.0002	0.0000	0.0076	0.9919	0.0000	0.0002
ʌ	0.0000	0.0001	0.0001	0.0008	0.0525	0.0124	0.0100	0.0000	0.9221	0.0020
ɹ	0.0000	0.0000	0.0022	0.0006	0.0002	0.0003	0.0000	0.0000	0.0002	0.9965

Multi-Dimensional Scaling

- Input: a confusion matrix
- e.g. Peterson & Barney 1952
- convert to probabilities
- distances are symmetrical ($d_{ij} = d_{ji}$) by definition.
- Confusion matrices are usually not symmetrical.
  - Explanation is disputed. One possible source is bias.
- Convert confusion probabilities to a symmetrical measure of similarity, s_{ij}.
  - $s_{ii} = 1$

perceived

		i	j
intended	i	p_{ii}	p_{ij}
	j	p_{ji}	p_{jj}

$$S_{ij} = \frac{p_{ij} + p_{ji}}{p_{ii} + p_{jj}}$$

Multi-Dimensional Scaling

- Symmetrical similarity matrix
- Similarity is related to distance in psychological space by an exponential decay function, where D_{ij} is the perceptual distance between *i* and *j*:

$$S_{ij} = ae^{-bD_{ij}} + c$$

  - based on observation, derivation attempted in Shepard 19??.

	i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɹ
i	1.0000	0.0005	0.0003	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000
ɪ		1.0000	0.0512	0.0002	0.0001	0.0001	0.0000	0.0000	0.0001	0.0013
ɛ			1.0000	0.0660	0.0001	0.0002	0.0001	0.0001	0.0002	0.0038
æ				1.0000	0.0011	0.0002	0.0001	0.0000	0.0012	0.0022
ɑ					1.0000	0.0868	0.0045	0.0001	0.0417	0.0005
ɔ						1.0000	0.0063	0.0003	0.0099	0.0009
ʊ							1.0000	0.0086	0.0141	0.0009
u								1.0000	0.0000	0.0001
ʌ									1.0000	0.0012
ɹ										1.0000

Multi-Dimensional Scaling

- Symmetrical similarity matrix
- Similarity is related to distance in psychological space by an exponential decay function, where D_{ij} is the perceptual distance between *i* and *j*:

$$S_{ij} = ae^{-bD_{ij}} + c$$

  - based on observation, derivation attempted in Shepard 19??.
- MDS finds the best configuration of points for stimuli and the values of parameters *a, b, c* that provide the best fit to the similarity data (S_{ij}).
- Solutions are for a space with a specified number of dimensions.
  - Usually select dimensions based on how goodness of fit increases as number of dimensions is increased.

MDS analysis of vowel confusions

- Shepard (1972) presents a 3-dimensional MDS analysis of Peterson & Barney's vowel confusion data.

- MDS is based on confusions alone - can be difficult to relate derived space to physical stimulus dimensions.
- With vowel space, there are two nearly orthogonal dimensions that correlate well with F1, F2.
  - dimension that best correlates with F3 is not orthogonal - close to F2.
  - relation between dimensions of perceptual space and Hz formant frequencies are non-linear.

References

- Delattre, Pierre C., Alvin M. Liberman, Franklin S. Cooper, and Louis J. Gerstman (1952). An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word* 8, 195-210.
- Haggard, Mark P., Stephen Ambler, and Mo Callow (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America* 47, 613-17.
- Jun, Jongho (1995). *Perceptual and Articulatory Factors in Place Assimilation: An Optimality-Theoretic Approach*. PhD dissertation, UCLA.
- Kawasaki, Haruko (1982). An acoustical basis for universal constraints on sound sequences. Ph.D. dissertation, University of California, Berkeley.
- Kingston, John, and Randy L. Diehl (1995). Intermediate properties in the perception of distinctive feature values. Bruce Connell and Amalia Arvaniti (eds) *Papers in Laboratory Phonology IV* , Cambridge University Press, Cambridge.
- Lindblom, Björn, and Michael Studdert-Kennedy (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America* 42, 830-843.
- Lisker, Leigh (1957). Closure duration and the intervocalic voiced-voiceless distinctions in English. *Language* 33, 42-49.
- Lisker, Leigh (1975). Is it VOT or a first formant transition detector?. *Journal of the Acoustical Society of America* 57, 1547-51.

References

- Lisker, Leigh (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29.3-11.
- Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on the perception of [ ]-s] distinction: I. Temporal factors. *Perception & Psychophysics*, 28, 213-228.
- Massaro, Dominic W., and Michael M. Cohen (1983). Consonant/vowel ratios: An improbable cue in speech. *Perception and Psychophysics* 33, 501-5.
- Ohala, J.J. (1990) The phonetics and phonology of aspects of assimilation. M. Beckman and J. Kingston (eds) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. CUP, Cambridge.
- Plomp, Reinier (1975). Auditory analysis and timbre perception. Gunnar Fant and Michel Tatham (eds.) *Auditory Analysis and Perception of Speech*. Academic Press, New York, 7-22.
- Repp, Bruno (1979). Relative amplitude of aspiration noise as a cue for syllable-initial stop consonants. *Language and Speech* 22, 947-950.
- Shepard, Roger N. (1972). Psychological representation of speech sounds. Edward David and Peter Denes (eds.) *Human Communication: A Unified View*. McGraw-Hill, New York, 67-113.
- Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. Ms, UCLA.
- Stevens, Kenneth N., and Sheila E. Blumstein (1981). The search for invariant acoustic correlates of phonetic features. Peter D. Eimas and Joanne L. Miller (eds.) *Perspectives on the study of speech*. Lawrence Erlbaum, Hillsdale.
- Wright, R., Frisch, S., & Pisoni, D. B. (1999). Speech Perception. In J. G. Webster (Ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 20 (pp. 175-195). New York: John Wiley and Sons.