

Speech perception

Our computational model helps us derive an auditory representation in terms of the neural firing (spatial) patterns that change in time with the acoustic signal characteristics.

It is of interest to consider how the spatio-temporal auditory representation encodes the important acoustic cues (and correspondingly, the articulatory cues) of the different phones.

Vowel perception: Perception studies on synthetic vowels indicate that vowel quality is determined by pattern of spectral peaks. Thus formant regions should correspond to higher neural activity in the corresponding regions of the basilar membrane.

Natural speech: formant trajectory information is also used => temporal variation in the spatial pattern of neural activity.

Consonant perception: Voicing, Manner, Place
(see specgrams in Class-SP-1.one)

Fricatives: continuous turbulence => cued by absence of harmonic structure; place = spectral energy concentration

Nasals: spectral envelope with resonances and anti-resonances

Stops: voice onset time, spectrum of burst, formant transitions

The **further mechanisms of the brain to translate these into linguistic messages** are not understood that well.

What is known: Speech perception is different from non-linguistic audition. It takes place in the left brain.

The challenge in speech perception is that the same phone displays a significant amount of acoustic variability depending on the speaker, phonetic context,...

The linguistic approach to phone classification uses a set of distinctive features relating to: voicing, manner, place of articulation. "Motor theory" of speech perception => sounds are subsumed into phonemic equivalence classes.

Categorical perception: Listeners can only discriminate between sounds that they would identify as belonging to different categories => listeners do not perceive differences within phones of the same class.

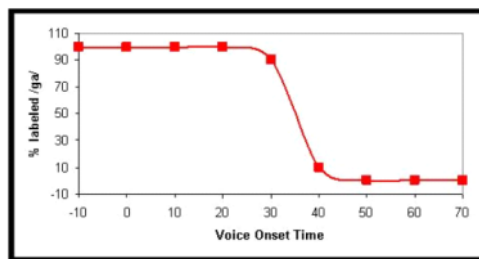


E.g. Differences in Voice Onset Time (VOT = vowel onset after burst):

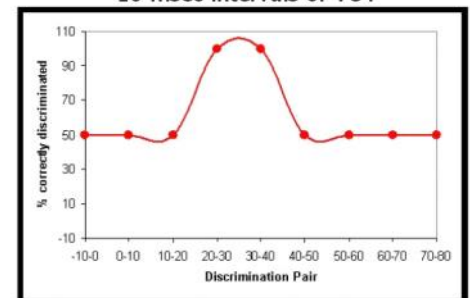
voiced<- low VOT; unvoiced<- high VOT

Was that "ga" or "ka"?

% labeled /ga/ in /ga/-/ka/ continuum



Results of discrimination task:
10 msec intervals of VOT



Vowel perception is "continuous".

Examples...