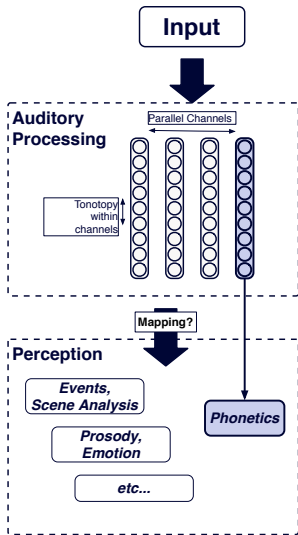


Auditory Representation of Vowel Quality

Andrew B. Wallace and Sheila E. Blumstein

Brown University, Dept. of Cognitive and Linguistic Sciences
contact: andrew_wallace@brown.edu



Introduction

•Speech perception is a hierarchical process. A sequence of transformations map from one level of analysis to the next (cf. McClelland & Elman 1986, Blumstein 1995).

•Parallel auditory processing: multiple representations have been observed in auditory cortex, differing in their spectral and temporal properties (cf. Sutter & Schreiner 1991, Poeppel 2003). They are presumed to map onto different higher-level processes, according to task demands.

Questions

Which of these representations maps onto phonetics? What acoustic information does it encode?

A Methodological Problem

•Speech stimuli don't isolate auditory stages.
•Simple nonspeech stimuli (tones, noise) don't isolate speech perception.

A Solution

Effects of nonspeech sounds on speech perception.
•Lotto and Kluender (1998).
•Holt, Lotto, and Kluender (2000).

Priming vowel perception with tones.

•Subjects identify vowel targets [i] and [a], resynthesized with steady-state formants:

Frequency	Bandwidth	
[i]	[a]	
362 Hz	772 Hz	82 Hz
2191 Hz	1308 Hz	39 Hz
3100 Hz	2960 Hz	261 Hz
3487 Hz	204 Hz	
3997 Hz	238 Hz	

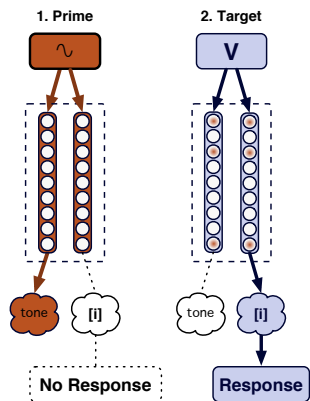
•Targets are preceded by nonspeech "primes", pure tones or tone complexes matched to the formant frequencies (e.g. F1 and F2) of the targets.

•Primes not heard as speech, based on debriefing of subjects.

•Subjects identify targets faster after matching prime than after mismatched prime. This priming effect is taken to reflect the degree to which the primes activate auditory representations specific to the matched target.

Wallace and Blumstein (2009, 2006):

Nonspeech priming methods suggest short window (\leq ca. 50 ms) for auditory analysis of vowels.



Experiment 1: Spectral Resolution

Critical bandwidth (e.g. Fletcher 1940, Zwicker 1961) is well studied, seems to reflect frequency tuning of cochlea and auditory nerve.

Frequency tuning is more variable in cortex:

- Tuning gradient perpendicular to tonotopic axis of A1 (Schreiner and Mendelson 1990).
- Even broader tunings in nonprimary auditory areas (Kowalski et al. 1995, Rauschecker and Tian 2004).
- "Ultrafine" tuning in humans (Bitterman et al. 2008).

Speech perception doesn't require high resolution.

- Shannon et al. (1995)
- Chistovich (see 1985; but cf. Klatt 1985, Assmann 1991).

Except at high signal-to-noise ratios.

- Dorman et al. (1998)

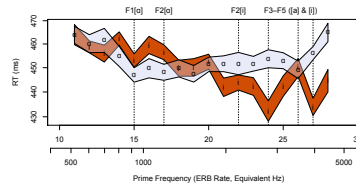
Methods

- Targets: [i] and [a].
- Primes: 18 pure tones
- 519–4426 Hz, evenly spaced in ERB_N rate scale (Glasberg and Moore 1990).

•Subjects: 20 native speakers of N. American English.

•25 repetitions per prime x target condition.

Results



Experiment 2: Formant Combinations

•Vowel quality is encoded by patterns of the first 2–3 formants, rather than by any single formant (cf. Peterson and Barney 1952, Delattre et al. 1952). As a result, information spread across distant regions of the spectrum must be integrated during perception.

•A1 cells respond to spectral combinations:

•Linear "multi-peaked" neurons (Sutter and Schreiner 1991): tuning curve has two distinct peaks. Combinations of tones at both peaks produce predictably greater response.

•Nonlinear "combination sensitive" neurons (Suga et al. 1978): respond to combinations greater than predicted from responses to individual components. Much studied in bats, involved in highly specialized biosonar systems.

•Kadia and Wang (2003): Nonlinear combination sensitivity in primates (common marmoset). Subset of cells which show a single peak in response to pure tones are facilitated by off-peak areas in two-tone measurements.

Questions:

Are formant combinations integrated at auditory level? Is the integration mechanism linear or nonlinear?

Methods

Targets: [i] and [a].

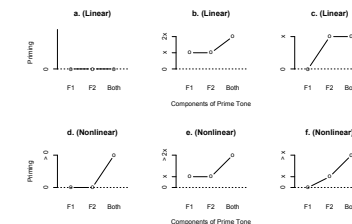
Primes: Tones/complexes matched to:

- F1 frequency of target only.
- F2 frequency of target only.
- F1 and F2 frequencies combined.

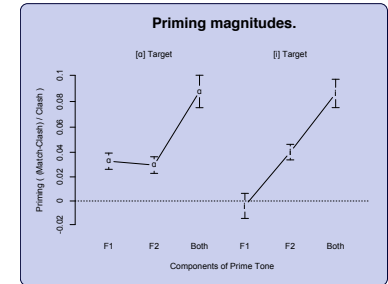
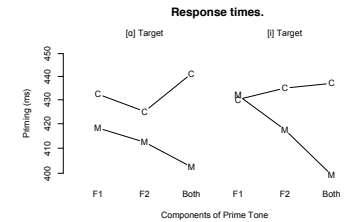
Subjects: 16 native speakers of North American English with normal hearing and no history of speech or language disorders.

Predictions

- Linear integration → additive priming.
- Nonlinear integration → super-additive priming.



Results



References

Assmann, P. (1991). The perception of back vowels: Centre of gravity hypothesis. *The Quarterly Journal of Experimental Psychology Section A*, 43(3), 453–468.

Bitterman, Y., Mukamel, R., Malach, R., Fried, I., & Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature*, 451, 197–201.

Blumstein, S. E. (1995). The neurobiology of the sound structure of language. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 815–859). Cambridge, MA: MIT Press.

Chistovich, L. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America*, 77, 786.

Delattre, P. C., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel colour, observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word*, 8, 195–210.

Dorman, M., Luizou, P., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. *Journal of the Acoustical Society of America*, 104(8), 3583–3595.

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, 12(1), 47–65.

Glasberg, B., & Moore, B. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1–2), 103–138.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, 108(2), 710–722.

Klatt, D. H. (1985). A shift in formant frequencies is not the same as a shift in the center of gravity of a multimodal energy concentration. *Journal of the Acoustical Society of America*, 77(5), 1711–1717.

Kadia, S. C., & Wang, X. (2003). Spectral integration in A1 of Awake Primates: Neurons With Single and Multi-peaked Tuning Characteristics. *Journal of Neurophysiology*, 89, 1603–1622.

Kowalski, N., Versnel, H., & Shamma, S. (1995). Comparison of responses in the anterior and primary auditory fields of the ferret cortex. *Journal of Neurophysiology*, 73(4), 1513–1523.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop component identification. *Perception and Psychophysics*, 60(4), 602–619.

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology*, 18, 1–88.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, 24(2), 175–184.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as asymmetric sampling in time. *Speech Communication*, 41, 249–255.

Rauschecker, J. P., & Tian, B. (2004). Processing of Band-Pass noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol*, 91(6), 2579–2599.

Schreiner, C. E., & Mendelson, J. (1990). Functional topography of cat primary auditory cortex: distribution of integrated excitation. *Journal of Neurophysiology*, 64(5), 1442–1459.

Shannon, R. V., Zang, F.-G., Kamath, V., Wygnanski, J., & Eklid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 300–304.

Suga, N., O'Neill, W. E., & Manabe, T. (1978). Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the mustache bat. *Science*, 204(4343), 778–791.

Sutter, M. L., & Schreiner, C. E. (1991). Physiology and topography of neurons with multiplexed tuning curves in cat primary auditory cortex. *Journal of Neurophysiology*, 65(5), 1207–1228.

Wallace, A. B., & Blumstein, S. E. (2006). Nonspeech sounds prime acoustically similar words. *Journal of the Acoustical Society of America*, 119(5), Pt 2, 3245.

Wallace, A. B., & Blumstein, S. E. (2009). Temporal integration in vowel perception. *Journal of the Acoustical Society of America*, 125(3), 1711–1717.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (frequency groups). *Journal of the Acoustical Society of America*, 33(2), 248.

Acknowledgements

This work was supported by NIH grants DC000314 and DC006220 to Brown University.