Nonspeech Sounds Prime Acoustically Similar Words

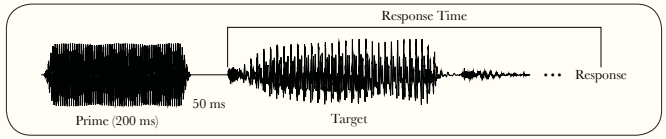
Andrew Wallace and Sheila Blumstein (email: andrew_wallace@brown.edu)
Dept. of Cognitive and Linguistic Sciences, Brown University

Introduction

Recent experiments have shown that the phoneme boundary shifts which compensate for coarticulation of speech segments can also be caused by nonspeech sounds (e.g. Lotto and Kluender 1998; Holt, Lotto, and Kluender 2000). This finding suggests that a general auditory mechanism for enhancing spectral contrast plays a role in speech perception. If this is the case, it is possible that other aspects of speech perception make use of auditory processes that are not specific to speech.

For example, the perception of vowels might depend on a general auditory mechanism for analyzing spectral patterns, which would extract information about formant structures from the acoustic signal. If so, it is hypothesized that identification of speech stimuli will be primed by tone complexes matched to their formant frequencies. The present experiments test this hypothesis.

Experiment 1



Subjects

Subjects were 21 native speakers of North American English with no history of speech or language disorders.

Stimuli

Targets were naturally produced tokens of the words "deed", "dad", and "dude". Target amplitude was normalized to produce a perceptual loudness of 20 sones, as predicted by the ISO532B standard.

Targets

	"deed"	"dad"	"dude"
Mean F0	120 Hz	107 Hz	113 Hz
mid-plateau F1	336 Hz	860 Hz	407 Hz
mid-plateau F2	2366 Hz	1540 Hz	1513 Hz
mid-plateau F3	3045 Hz	2472 Hz	2324 Hz
Voicing Duration	267 ms	307 ms	263 ms
Intensity	71.9 dB SPL	70.8 dB SPL	72.7 dB SPL
Loudness	20 sones	20 sones	20 sones

Primes were complexes of tones matched to the mid-plateau frequencies of the first two formants of each target. An additional neutral prime consisted of a pure tone between the highest F1 and the lowest F2. The different components of each prime were equally loud (by ISO226), and the total loudness of each prime was 20 sones (by ISO532B). Primes were 200 ms in duration, including 15 ms cosine-squared onset and offset ramps.

Primes

	prime(deed)	prime(dad)	prime(dude)	neutral
Frequencies	336 Hz 2366 Hz	860 Hz 1540 Hz	407 Hz 1513 Hz	1141 Hz
Duration	200 ms	200 ms	200 ms	200 ms
Intensity	76.7 dB SPL	76.4 dB SPL	76.2 dB SPL	79.5 dB SPL
Loudness	20 sones	20 sones	20 sones	20 sones

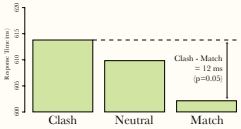
Procedure

Each trial consisted of a prime and a target separated by a 50 ms interval. Subjects responded by pressing a button to identify the target, choosing from three buttons labeled "deed", "dad", and "dude". Response times were measured from the onset of the target.

Combining the 3 targets and 4 primes made 12 unique trials. Each subject repeated each trial 40 times in random order.

Trials were classified as:
"Match" e.g. prime(deed)-deed
"Clash" e.g. prime(dad)-deed
"Neutral" e.g. neutral prime-deed

Results



Results show significant priming in the Match condition relative to the Clash condition.

Experiment 2

Goals

Experiment 2 sought to replicate the findings of Experiment 1, while developing a simpler and more controlled design that would allow for a parametric investigation of this effect in future studies.

Subjects

Subjects were 10 native speakers of North American English with no history of speech or language disorders.

Stimuli

Targets were 2 vowel sounds, [i] and [ɑ], produced by source-filter synthesis from a naturally produced [i] using Praat (Boersma and Weenink, 2005). The first three formants were set manually to constant values which produced the desired vowel quality. Targets were normalized to a loudness of 20 sones (ISO532B).

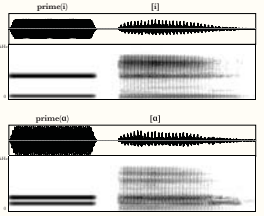
Targets

	[i]	[ɑ]
Mean F0	126 Hz	126 Hz
F1	362 Hz	772 Hz
F2	2191 Hz	1308 Hz
F3	3100 Hz	2700 Hz
Voicing Duration	280 ms	280 ms
Intensity	71.4 dB SPL	71.4 dB SPL
Loudness	20 sones	20 sones

Primes were nonspeech tone complexes, made to the specifications reported in Experiment 1, but matched to the targets of Experiment 2. The loudness of each prime was increased to 25 sones (ISO532B), in the hope that this would enhance the priming effect.

Primes

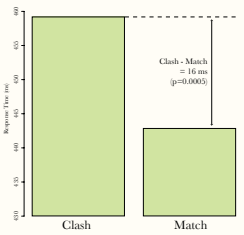
	prime(i)	prime(ɑ)
Frequencies	362 Hz 2191 Hz	772 Hz 1308 Hz
Duration	200 ms	200 ms
Intensity	79.1 dB SPL	82.5 dB SPL
Loudness	25 sones	25 sones



Procedure

Primes and targets were presented and response times measured as described for Experiment 1. Combining the 2 targets and 2 primes made 4 unique trials (2 Match and 2 Clash), each of which was repeated 250 times in random order by each subject. Subjects responded by pressing a button labeled [i] or [ɑ].

Results



Discussion

Experiments 1 and 2 confirm the hypothesis that spectrally matched tone complexes prime vowels.

Experiment 3

Goals

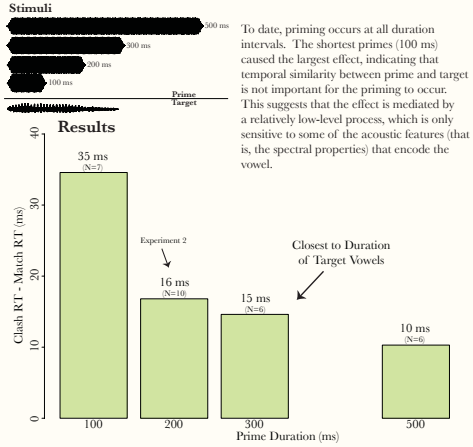
Experiment 3 (in progress) extends Experiment 2 by measuring the effect of variation in the duration of the nonspeech primes on the magnitude of priming. This will indicate whether the effect depends on temporal similarity between prime and target, in addition to spectral similarity.

If temporal similarity is important, the effect might be greatest when the duration of the prime is equal to the duration of the target. If only spectral similarity is important, then priming might be greatest for longer or shorter primes, or equal across all prime durations.

Methods

Stimuli and procedure were identical to experiment 3, except that the duration of the primes was varied between subjects. New data were collected for 100 ms, 300 ms, and 500 ms primes, while data from Experiment 2 for 200 ms primes are included for comparison.

The duration of the vowel target was 280 ms.



Discussion

To date, priming occurs at all duration intervals. The shortest primes (100 ms) caused the largest effect, indicating that temporal similarity between prime and target is not important for the priming to occur. This suggests that the effect is mediated by a relatively low-level process, which is only sensitive to some of the acoustic features (that is, the spectral properties) that encode the vowel.

Conclusions

These results confirm the hypothesis that spectrally matched nonspeech sounds prime the identification of vowels. This occurred even in a phonetic categorization task, supporting the idea that speech perception makes use of a general auditory process for spectral analysis to extract the formant structures of vowels.

Preliminary data suggest that this effect is strongest when primes are very short, rather than when they are matched to the duration of the target. Thus, it appears that it is mediated by a process that is sensitive to spectral, but not temporal, acoustic features.

References

Paul Boersma & David Weenink (2005). Praat: doing phonetics by computer (Version 4.4.07) [Computer program]. Retrieved February 2, 2006, from http://www.praat.org/

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

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

ISO532B. MATLAB implementation by Claire Churchill. http://www.acoustics.salford.ac.uk/research/air/cox/sound_quality/assess.htm

ISO226. MATLAB implementation by Jeff Tackett. http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=7028&objectType=FILE

Acknowledgements

Many thanks to Laurie Heller, who helped preparing stimuli.

This work was funded by NIH Grant DC006220 to Brown University.