Psychophysical and MEG studies of the Timbre of Synthesized Approximations to Vowels

Julian Jenkins III[1,2], Jonathan Z. Simon[1,3], David Poeppel[4], William J. Idsardi[2,5]

[1]Department of Biology, University of Maryland, College Park; [2]Cognitive Neuroscience of Language Laboratory, University of Maryland, College Park; [3]Department of Electrical and Computer Engineering, University of Maryland, College Park; [4]Department of Psychology, NYU; [5]Department of Linguistics, University of Maryland, College Park

BACKGROUND

▪Auditory domain: **timbre** is responsible for object identification [1,2].

▪Allows for identification of both ecological and non-ecological auditory signals [3].

▪Speech signals: timbre responsible for processes such as speaker identity, gender and affective state [4].

▪Goal: link dimensions that contribute to the timbre of an object via parametrically controlled speech-like signals; measure how the physics of the signal relate to perceptual and physiological processing [5,6,7,8,9].

PSYCHOPHYSICAL DESIGN

1. AX discrimination task; participants randomly placed in one of two experimental pools
2. Subset of continuum signals – glottal excitation pattern, 4th removed, 2nd removed, square sources
3. Analyses: Proportion Correct (Vowel, Vowel Height, Vowel Position, Trial); GLMs: proportion correct (factors: Vowel, Vowel Height, Vowel Position, Trial)

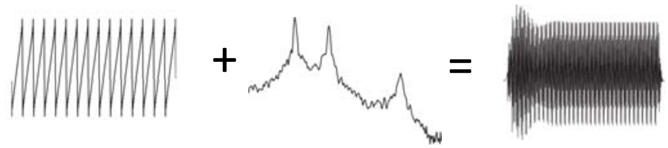
MEG DESIGN

1. Passive presentation; participants randomly placed in one of two experimental pools
2. Calculated Peak RMS and latency (M50, M100, P2m)
3. Analyses: Peak RMS and latency (M50 – vowel bandwidth; M100 – F0, transfer function structure; P2m – harmonics contained); Bootstrapped RMS: amplitude differences between condition; grand averages: response consistency

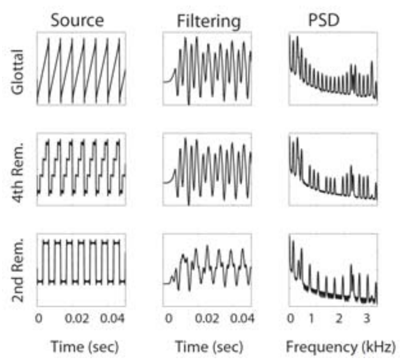
HYPOTHESES

1. Vowels equally discriminable
2. M50: elicited due to broadband nature
3. M100 latency and peak RMS: dependent on (i) F0 and transfer function and (ii) harmonics present
4. P2m: greater amplitude when more harmonics present

SIGNALS

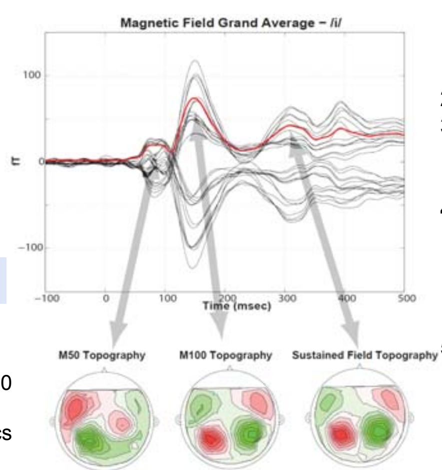


Each signal has its own timbre due to (i) source waveform harmonic structure and (ii) each vowel's transfer function (spectral envelope). Timbres within categories should be more similar than those across categories.



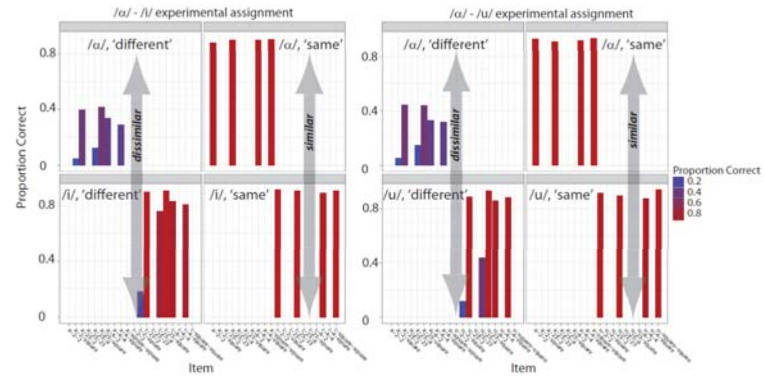
1. Ecological validity: source waveforms based on glottal excitation pattern (sawtooth wave comprised of 23 harmonics; (F0 = 150 Hz) via Fourier Synthesis
2. Filters (transfer functions) taken from formant values (F1,F2,F3) of male vocal tract measurements - /ɑ/, /i/, /u/
3. Create a timbral continuum by removing harmonics (every 6th, 5th, 4th, 3rd, 2nd)

MEG RESULTS



1. Major components in the evoked response : M50, M100, sustained field
2. **M50 very left-lateralized**
3. **M100 latency and RMS depend on F0 and transfer function structure**
4. **At least in preattentive state, evoked response not adequate to examine within-category timbral differences**
5. /ɑ/ - /i/: no statistically different peak RMS and latency (M100)
   /ɑ/ - /u/: /ɑ/ signals exhibited faster latency, larger amplitudes (M100)

PSYCHOPHYSICAL RESULTS



1. **Asymmetry in vowel proportion correct - /ɑ/ signals correctly discriminated much less than either /i/ or /u/– spectral peak averaging may 'blend' source spectrum**
2. **Asymmetry arises from signals in middle of the continuum**

DISCUSSION

• M100: latency and peak RMS determined by F0 and vowel structure [10,11]
• Evoked responses within vowel categories consistent

• Asymmetric processing of /ɑ/
Hypothesis 1: 'Averaging' of the first two formants – smears the source spectrum leading to discrimination impairment
Hypothesis 2 – signal dynamics (temporal evolution) differs based on transfer function; some filters produce more self-similar waveforms [12]

Future directions: (i) use vowels closer together in vowel space; (ii) solve wave equation [12]; (iii) examine induced response

REFERENCES

1. Binder et al. 2004. Nat Neurosci. 7(3): 295-301
2. Zatorre et al. 2004. J Neurosci. 24(14): 3637-42
3. Diesch and Luce 2000. J Cogn Neurosci. 12(5): 878-93
4. Fellowes et al. 1997. Percept Psychophys. 59(6): 839-49
5. Tiitinen et al. 2004. Neurol Clin Neurophysiol. 2004:103
6. Eulitz et al. 1995. J Neurosci. 15(4): 2748-55
7. Diesch and Luce 1997. Psychophysiology. 34(5): 501-10
8. Vihla et al. 2000. PNAS. 97(19): 10590-4
9. Caclin et al. 2005. J Acoust Soc Am. 118(1): 471-82
10. Poeppel et al. 1997. Neurosci Lett. 221(2-3): 145-8
11. Roberts et al. 2000. J Clin Neurophysiol. 17(2): 114-29
12. Balzano 1986. Music Perception. 3(3): 297-314

SfN 2011 172.13/KK8 contact: Julian Jenkins III (julianj3rd@gmail.com)

This work was supported by National Institute on Deafness and Other Communication Disorders of the National Institutes of Health: 2R01DC05660 to DP, JZS, and WJI and Training Grant DC-00046 support to JJIII