#3aSCb7

Use of stimulus mixing to synthesize a continuum of nasality in natural speech

Will Styler, Rebecca Scarborough, Georgia Zellou
University of Colorado Phonetics Lab

Our Goal

- We set out to re-synthesize a continuum of nasality using recorded, natural speech
- We chose an approach which focuses on the spectral aspects of nasality
- This allows the creation of natural-sounding stimuli to test the effect of differing degrees of nasality on perception.

What is vowel nasality?

- Vowel nasality is the result of velopharyngeal port opening during the production of a vowel
- Vowel nasality is contextual in English, arising due to the presence of nasal consonants to either side of the vowel, but contrastive in some other languages
- Vowel nasality has particular effects on the amplitudes of three peaks in the vowel spectrum, A1, P0 and P1, as described in Chen (1997)
  - A1 is the amplitude of the harmonic under F1, which decreases with nasality
  - P0 is the amplitude of the nasal peak between 250-400 Hz, usually the first or second harmonic.
  - P1 is another nasal peak, usually found near 950 Hz.

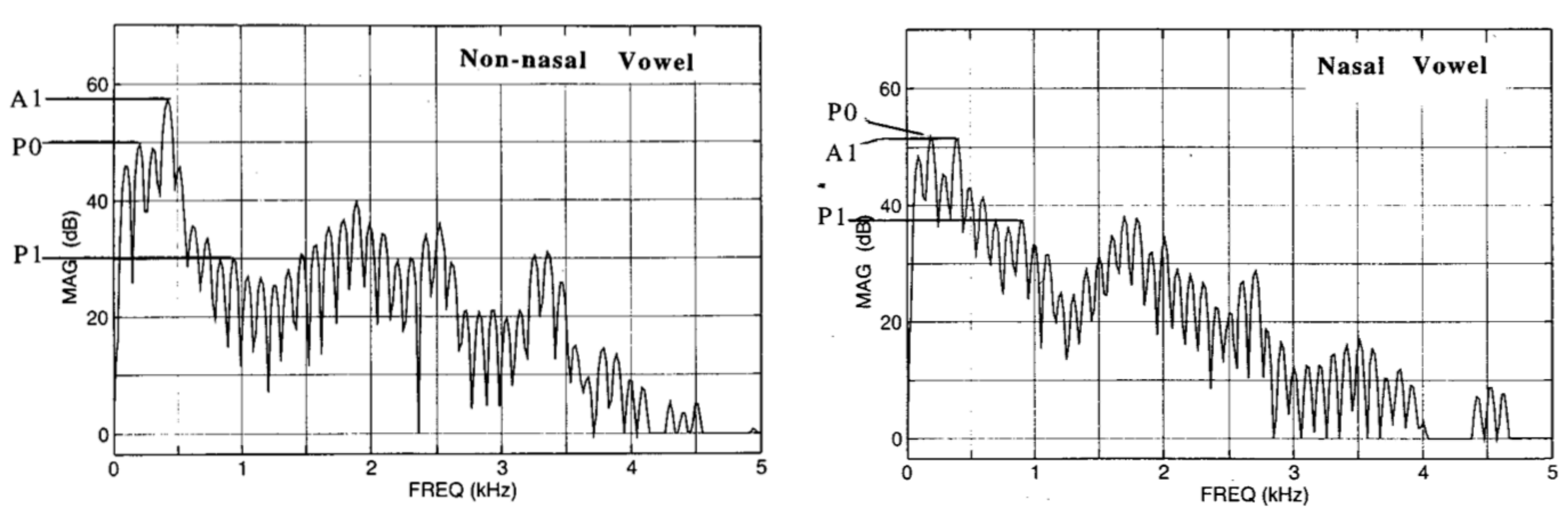


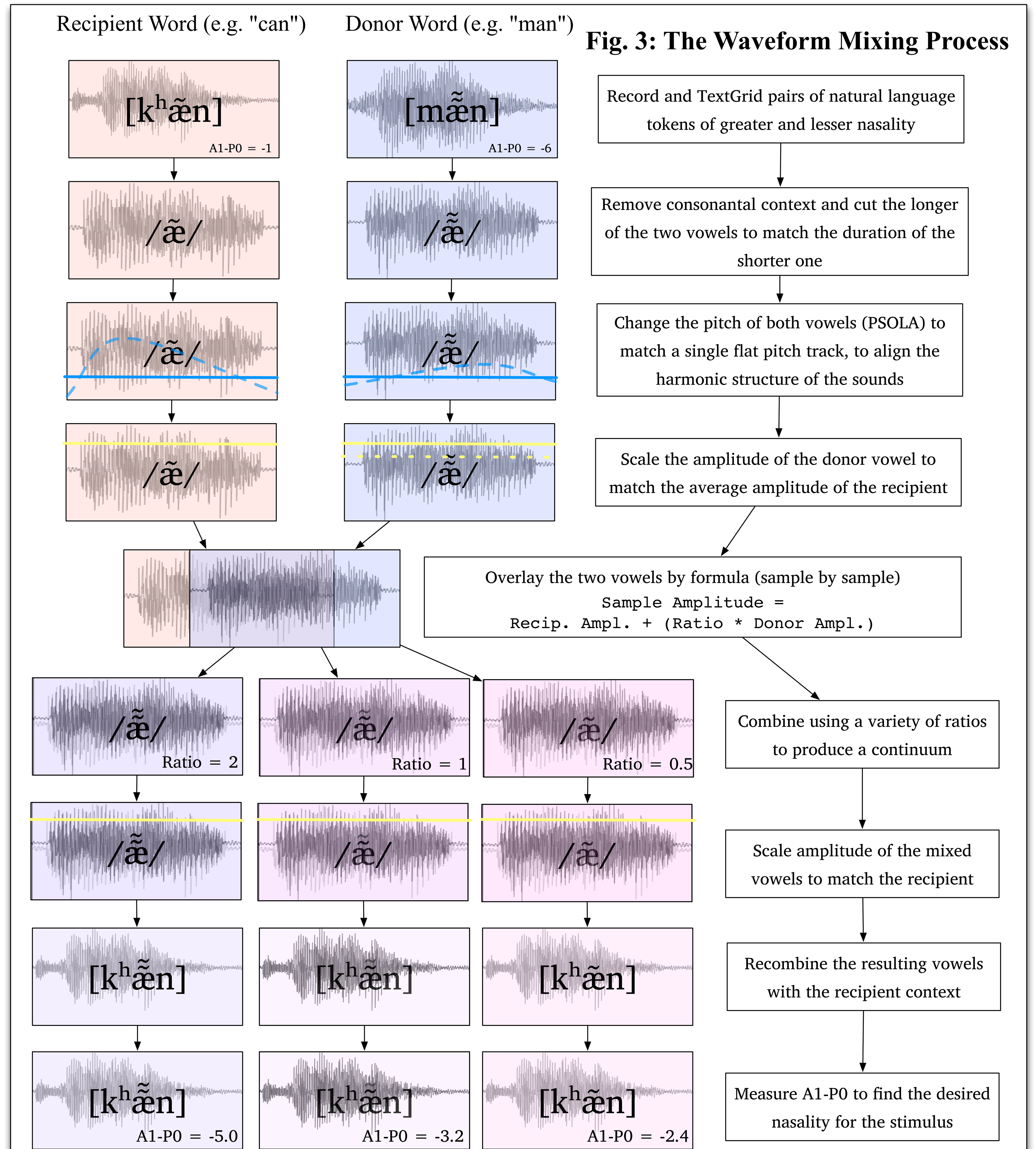
Fig. 1 & 2: Representative Spectra of Nasal and Non-nasal vowels showing A1, P0 and P1, and the changes in A1, P1 and P0 which occur when a vowel is nasalized (from Chen 1997, Fig. 2, pp. 2364)

- Chen (1997) compares the amplitudes of these spectral peaks to measure the nasality of vowels, examining A1-P0 and A1-P1
- Simply playing oral and nasal vowels simultaneously in varying proportions produces interference and poor modulation of nasality
- Instead, to manipulate vowel nasality, we adjust these spectral cues by adding pitch- and time-matched vowel waveforms

Creating Stimuli using this procedure

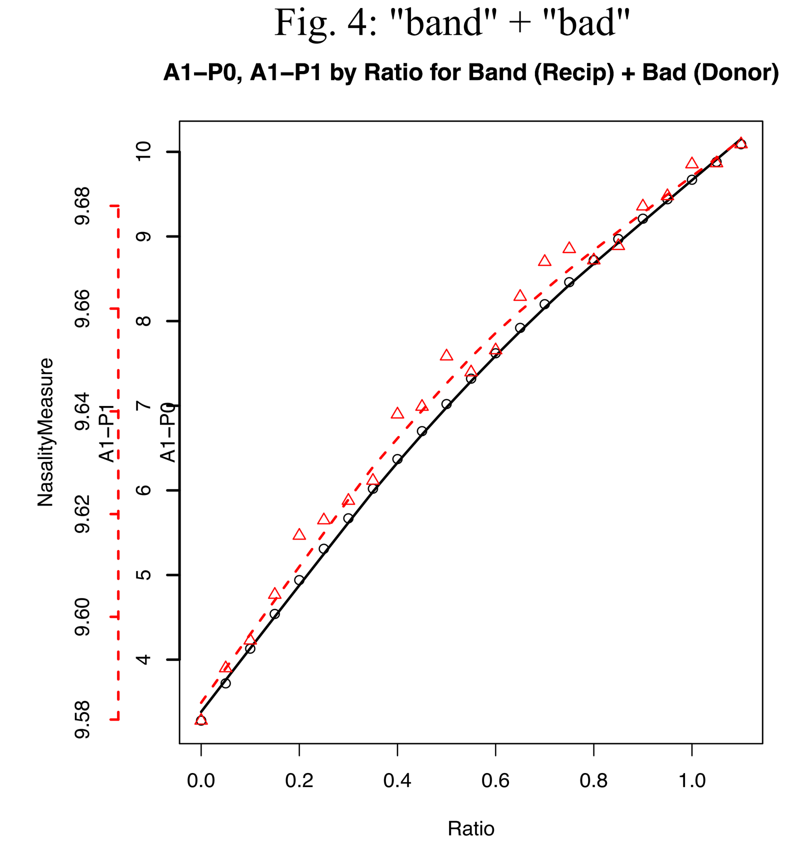
- Same vowel nasal/oral word pairs are recorded by a single speaker
- Word and vowel boundaries are annotated as Praat Textgrids
- These pairs are input into a Praat script, which generates a continuum of tokens with different ratios of nasal and non-nasal vowels using the process described in Figure 3
- Different step sizes between items in the continuum can be specified for more or less fine-grained
- The nasality of each resulting token from the continuum is then measured using A1-P0 and A1-P1 measurements
- Tokens are then selected from the continuum, based on the nasality measurements conducted above, to provide the desired amounts or differences in nasality

Fig. 3: The Waveform Mixing Process



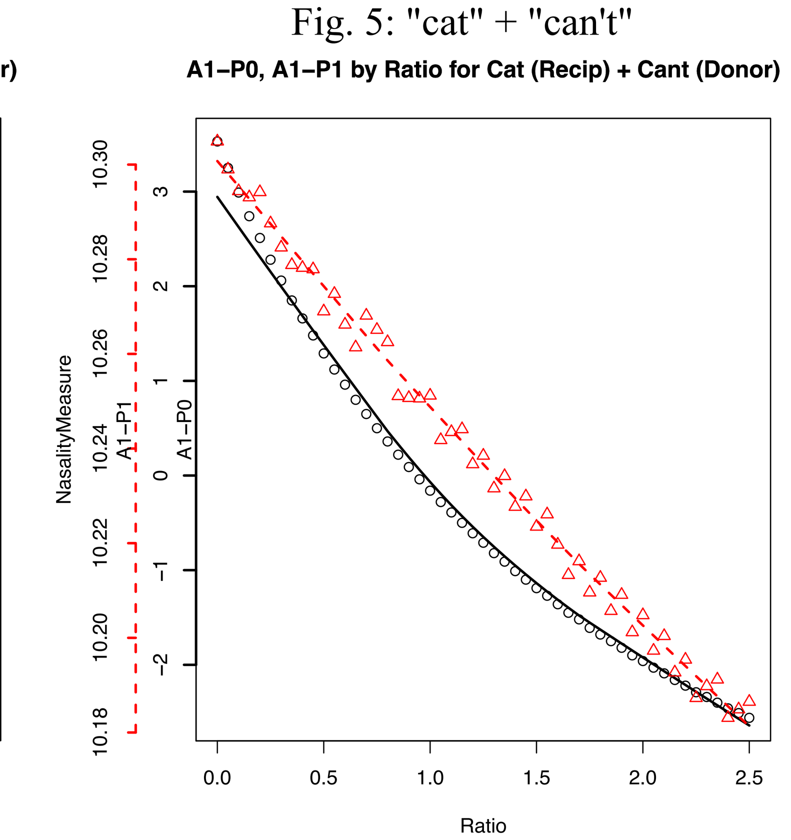
Decreasing Stimulus Nasality

Fig. 4: "band" + "bad"



Increasing Stimulus Nasality

Fig. 5: "cat" + "can't"



Acoustic Evaluation of the Stimuli

- Our method successfully changes nasality as measured by A1-P0 and A1-P1 (see Fig. 4 & 5)
- F0 does not change by ratio, nor does overall amplitude of tokens
- Other voice characteristics, such as speaker identity and dialect-specific vowel characteristics, are unchanged
- Number of tokens in the continuum can be modified to allow more fine-grained nasality degree distinctions
- There is a ceiling/floor effect, as the donor and recipient represent the most and least nasal sounds which can be created in this process

Perceptual Evaluation of the Stimuli

- Stimuli created in this way were used in two perceptual experiments testing the effect of sub-phonemic nasality changes on lexical decision in words of varying neighborhood densities (Scarborough, Styler and Zellou 2011, in preparation)
- Fig. 6 & 7 show reaction times (RTs) for high neighborhood density words from these two experiments

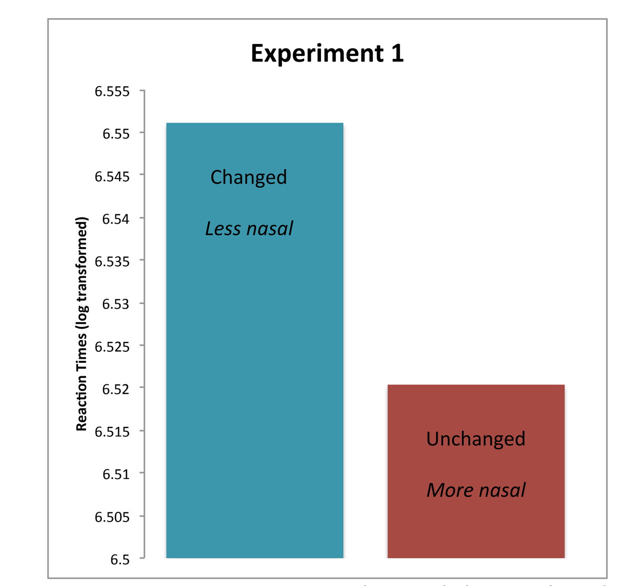


Fig. 6: Faster responses to unchanged, low ratio tokens (6.5203) than to modified (6.5510) tokens ($p < 0.001$)

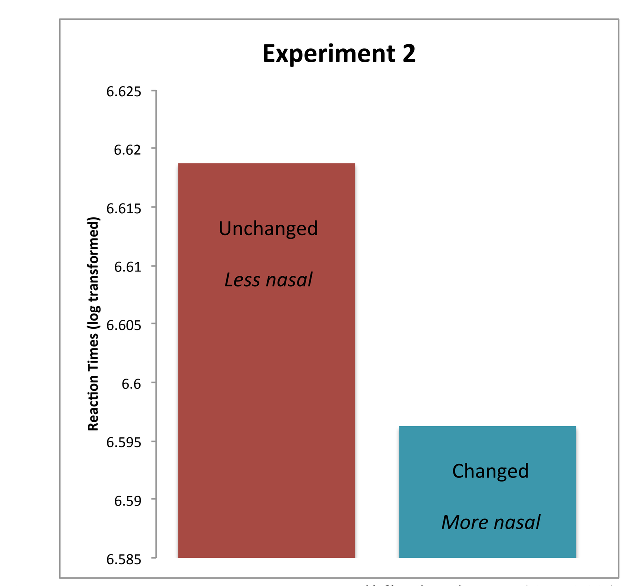


Fig. 7: Faster responses to modified tokens (6.5962) than to modified (6.6187) items ($p < 0.001$)

- In both cases, listener reaction times were clearly affected by our modification of the tokens, showing that these differences in nasality are perceptible enough to affect listener judgement
- Figure 7 shows that with some manipulations, listeners actually preferred (by reaction time) the modified, higher-ratio tokens to the unmodified, natural tokens, showing that this process does not reduce intelligibility in any significant way

Limitations of this approach

- This approach relies on the existence of natural tokens with more or less nasality than the desired continuum, thus, degrees of nasality greater than naturally produced are not possible
- This approach relies on the two tokens matching (nearly) for formant height and surrounding places of articulation. Speakers with pre-nasal formant changes may produce irregular stimuli
- The crucial pitch matching stage is limited by Praat's built-in pitch detection and modification
- Additional work must be performed to fully account for phase differences in the sounds

Additional Information

- You can find more information about the process, its implementation, or the experiments discussed above, along with sample continuua, at http://savethevowels.org/asa2011

References

- Chen, Marilyn Y. (1997) *Acoustic correlates of English and French nasalized vowels*, J. Acoust. Soc. Am. 102, 2360
- Scarborough, R. Styler, W. Zellou, G. (2011). *Nasal Coarticulation in Lexical Perception*. Proc. of the 17th ICPhS