

Visual-phonetic cues in the phonology of Toulousain French

Keith Johnson (UC Berkeley)
Laurel MacKenzie (UC Berkeley, UPENN)

One common historical development in languages with distinctively nasalized vowels is the excrescence of coda dorsal nasals on nasalized vowels. This is a property of the variety of French spoken in Toulouse. We will present data showing that the appearance of dorsal nasal is indeed a perceptual cue of the Toulousain dialect, though it is less common than popular accounts might suggest. Then in two experiments we will consider why the cross-linguistically unmarked place for this post-nasality nasal is dorsal. The experiments compare Ohala's (1975) acoustic explanation - namely that dorsal nasals, having no antiformants, are acoustically more similar to vowels than are labial or coronal nasals - with an explanation of our own based on visual correlates of distinctive features. The "visual correlates" explanation holds that if the perceiver detects a nasal coda consonant but does not see the lips or tongue tip produce a stop closure, then the visually "unmarked" place of articulation must be dorsal. The experiments contrast place of articulation judgments given to tokens ending in nasalized vowels by French- and American English- speaking participants. In one experiment we simply presented monosyllabic CV nonwords with nasalized vowel. Here the evidence argues for a spelling bias for French speakers and no change for audio versus audio-video presentation. Thus, experiment 1 supports Ohala's account. In the second experiment we obscured the last portion of CVN (N = /m/, /n/, or /ng/) and CV~ syllables with white noise. This experiment was designed to force listeners to assume the existence of a final consonant and to rely on visual cues to determine the place of articulation. The paper will present the results of this experiment and conclude with a discussion of phonetic modality and featural markedness.

Acknowledgments:

This research was supported by NSF grant #9817243.

Many thanks to the volunteers in California and France who participated in the study, and also to John Ohala for comments and inspiration.

1. Themes: Phonetic correlates of distinctive features.

Jakobson, Fant & Halle, 1952 is paradigm case.

Articulatory - lip closure, articulated with the front of the tongue, with vocal fold vibration, etc.

Acoustic - with a prominent mid frequency spectral peak, with low frequency energy, etc. (Stevens, 2002 is epitome of this approach)

Perceptual - JFH = psychoacoustic. Not as much consideration of this perspective on features.

In this talk we will consider some phonetic and phonological aspects of visual correlates of distinctive features.

2. Themes: Phonology and perception

- markedness patterns - associating greater perceptual salience with “marked”

- sound change - seeing patterns of change (and the resulting synchronic phonological patterns) in terms of misperception.

Ohala, J. (1981) The listener as a source of sound change.

3. Phonological context for this talk:

Excrescent [ŋ]

Standard French nasalized vowels -> Toulouse Vŋ

This process is found in many languages and language families:
Howe (2004) cites cases in Romance, West Germanic, Bantu, Niger-Congo, Austronesian, Papuan, Totonacan, Sino-Tibetan, Japanese, Mongolian.

Phonological explanations of excrescent [ŋ]:

Howe (2004), van der Toore (2003), Rice (1996)

** [ŋ] and vowels share a feature **

e.g. the “dorsal” articulator, as in Sagey (1986), Halle (1995)
[contra Clements & Hume (1995)]

This is fine, but phonetically unsatisfying -

- the “dorsal” articulation in [ŋ] and vowels is very different
- the relationship is more abstract than a phonetician would like
- is “share a feature” an explanatory mechanism?

Two phonetic hypotheses about excrescent [ŋ].

1. Acoustic similarity hypothesis - Ohala (1975)

- a. [ŋ] has no acoustic antiformants and is therefore more vowel-like than other nasals.
- b. Mouth cavity during [m] and [n] add acoustic zeros - antiformants - to nasal spectrum.
- c. Therefore, if a nasalized vowel is misperceived as a nasal segment the place of the segment will be velar because of the acoustic similarity of [ŋ] and vowels.
- d. Problem with this theory: [ŋ] has antiformants due to nasal sinuses. Is it really all that more acoustically vocalic?

Two phonetic hypotheses about excrescent [ŋ] - continued

2. Visual similarity hypothesis -

- a. [ŋ] has no visible mouth closure and is therefore more vowel-like than other nasals.
- b. Mouth movement during [m] and [n] has visible closure.
- c. Therefore, if a nasalized vowel is misperceived as a nasal segment the place of the segment will be velar because of the visual similarity of [ŋ] and vowels.
- d. Problem with this theory: lack of data.

4. About visual speech perception.

a. Visual input influences language acquisition.

Mills (1987) blind children learn [labial] later than sighted children -

[b]/[d] and [b]/[g] were confused by blind
[d]/[g] were confused by both blind and sighted

The “markedness” of [labial] for these children is partially determined by the visibility of the lips.

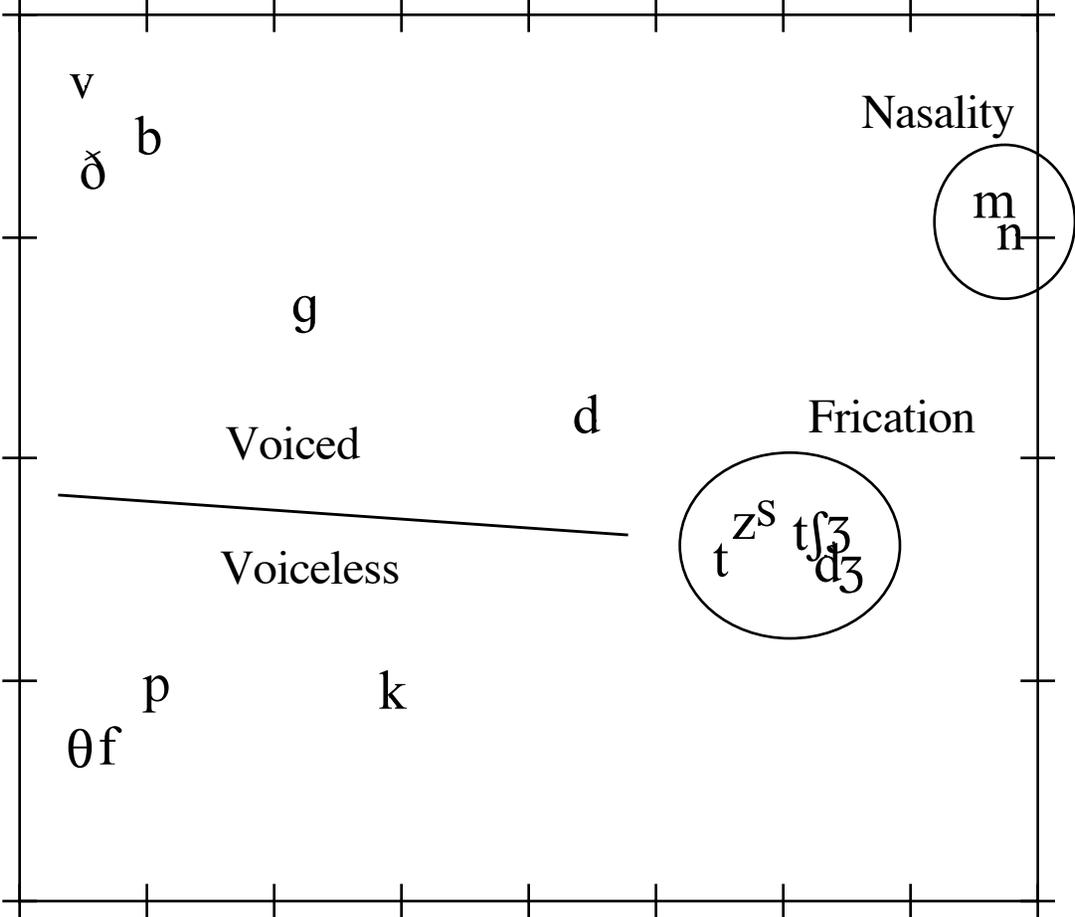
4. About visual speech perception.

b. Visual phonetic similarity is substantially different from auditory phonetic similarity.

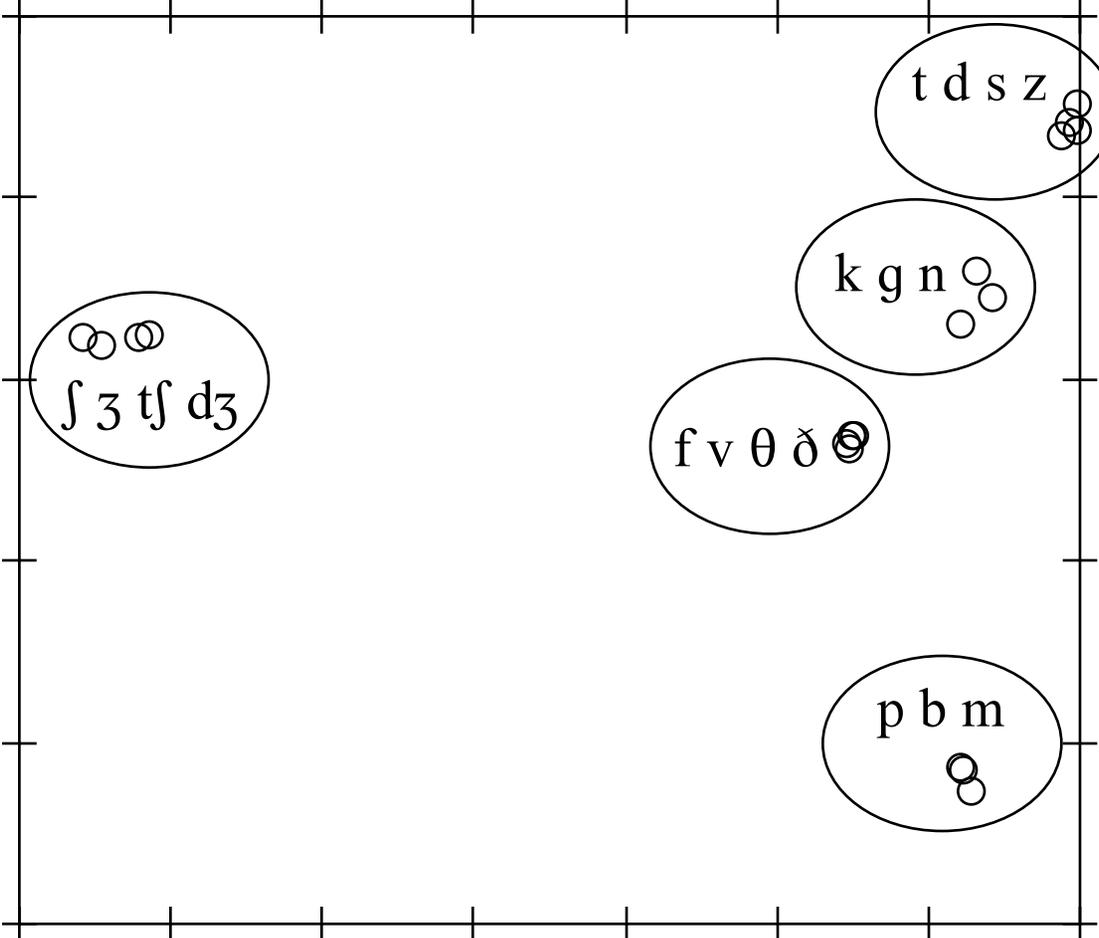
Grant, K.W., and Walden, B.E. (1996). Evaluating the articulation index for auditory-visual consonant recognition, *J. Acoust. Soc. Am.* 100, 2415-2424.

12 participants, 18 consonants in [a__a] context.

Perceptual space - Audition only



Perceptual space - vision only



For consonants:

Auditory dimensions of similarity: Voicing, manner, nasality

Visual dimensions of similarity: Place

Note: Dorsal and Coronal stop place are distinct
- except [n]

5. Testing the visual similarity hypothesis for excrescent [ŋ]

a. Excrescent [ŋ] in Toulouse

Do French listeners associate the [ŋ] with Midi French?

b. Experiment 1 -

masked final nasal with or without visual information

American English listeners

c. Experiment 2

Unmasked nasalized vowels

American English and French listeners

5a. Excrescent [ŋ] in Toulouse

Generally it is reported that French nasalized vowels are pronounced [ŋ] in Toulouse. Does the presence of [ŋ] mark a speaker as Toulousian?

- i. One 25 year old male speaker from Toulouse -
described as having a strong accent
- ii. Audio clips from a 20 minute conversation -
half with examples of [ŋ] and half without [ŋ].
- iii. One group of listeners (n=10) heard [ŋ] clips,
another group (n=6) heard the no-[ŋ] clips
- iv. Presented via an experiment web site,
listeners recruited via e-mail to contacts in Toulouse.

Excrescent [ŋ] is relatively rare in this “strongly accented” talker’s speech.

Preliminary observations from a 20 minute recording

Of the hundreds of underlying nasalized vowels in the corpus, the majority of them did not show an excrescent [ŋ].

The excrescent [ŋ] was produced:

- at the end of a phrase (n=4) or utterance (n=4)
- before a vowel (n=8) [often the pause word "euh", n=6]
- part of the lexical item "enfin"/"fin" (n=11)

This speaker is judged to have a stronger Toulousian accent when listeners hear audio clips with [ŋ].

| | strong | weak | none | ?? |
|-------------|--------|------|------|----|
| with [ŋ] | 30 | 60 | 0 | 10 |
| without [ŋ] | 0 | 83 | 17 | 0 |

5. Testing the visual similarity hypothesis for excrescent [ŋ]

b. Experiment 1 - Masked final consonant.

Subjects. American English listeners (n=18)

Materials. Three sets of words

final nasal contrast [m n ŋ], or nasalized vowel

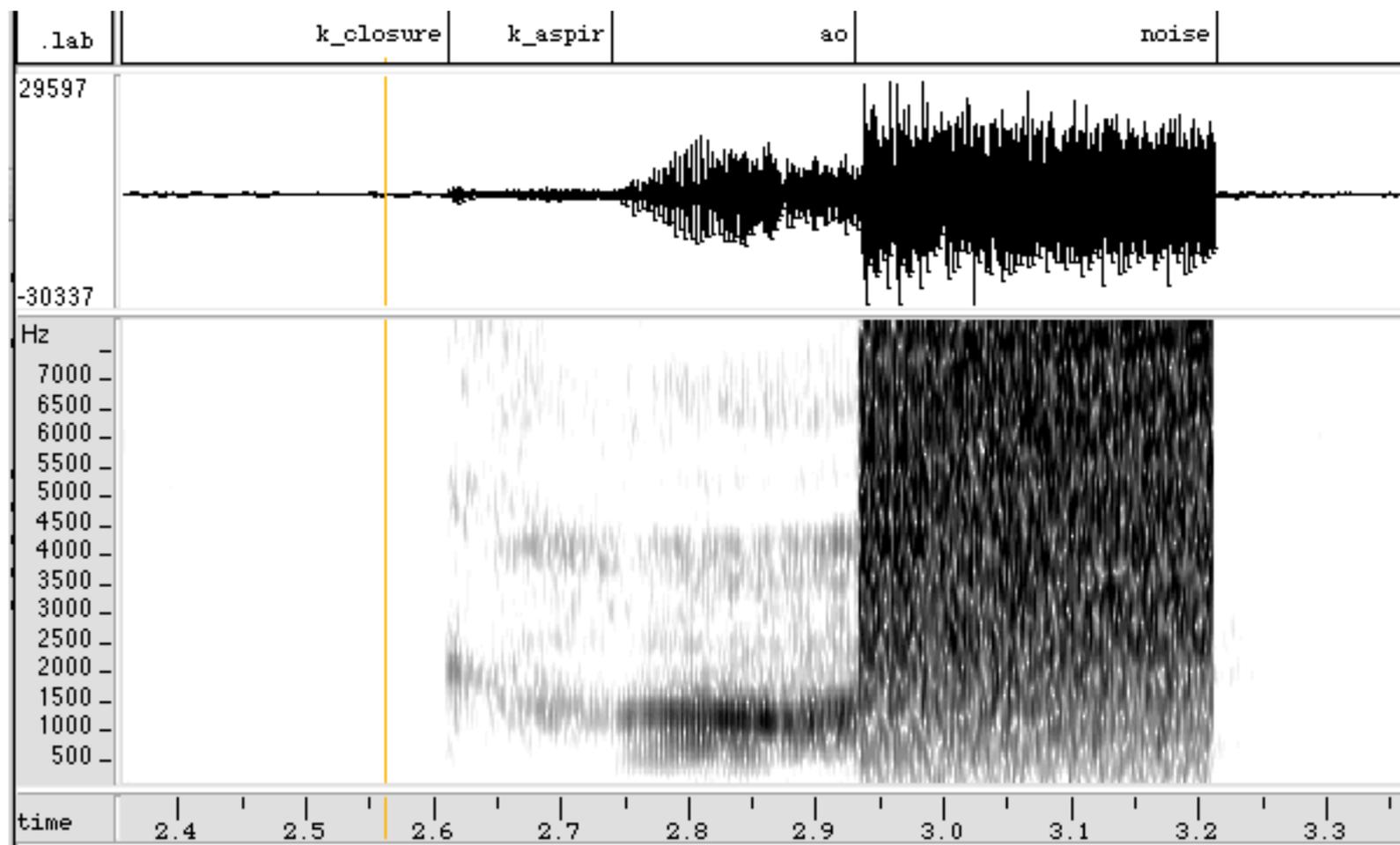
three vowel environments - [ʌ ɔ e^I]

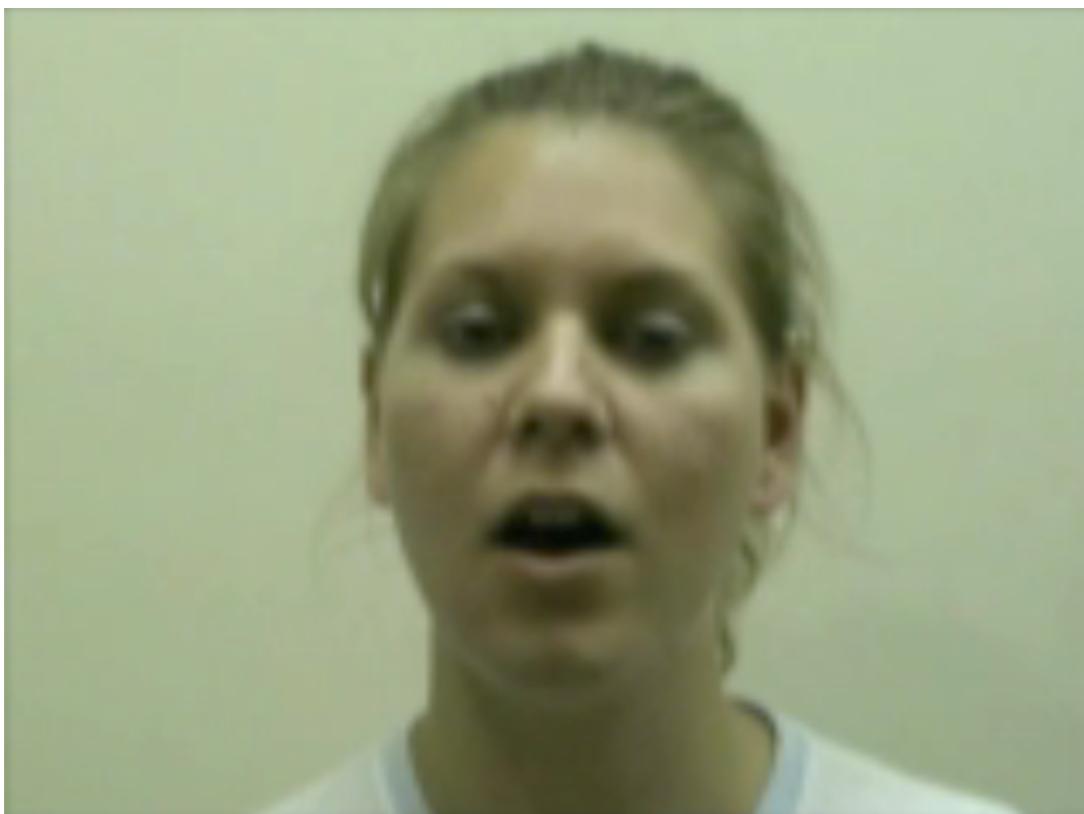
Speaker: Phonetically trained, native speaker of English, French L2 speaker.

[ʌ] dumb, done, dung, [dʌ̃]
rum, run, rung, [rʌ̃]
sum, sun, sung, [sʌ̃]

[ɔ] calm, con, kong, [kɔ̃]
pom, pawn, pong, [pɔ̃]
rom, ron, wrong, [rɔ̃]

[e^I] fame, feign, fang, [fɛ̃^I]
dame, dane, dang, [dɛ̃^I]
same, sane, sang, [sɛ̃^I]





5. Testing the visual similarity hypothesis for excrescent [ŋ]

b. Experiment 1 - Masked final consonant.

Task. Identify the final consonant in each audio token and then in a second block of trials in each AV token . (i.e. a within-subjects design).

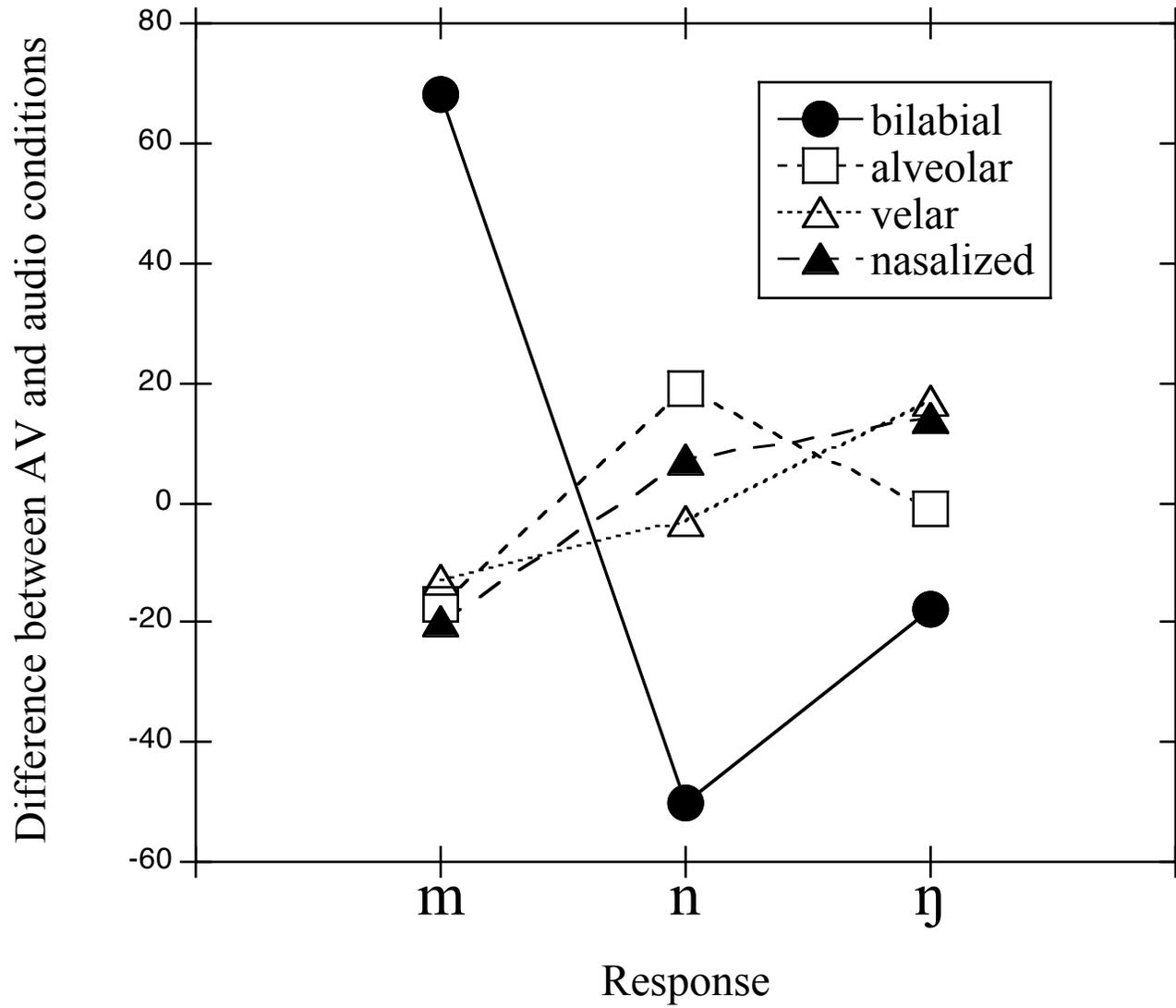
14/18 subjects also participated in experiment 2 (3 trials) at the conclusion of experiment 1.

Overall results of experiment 1.

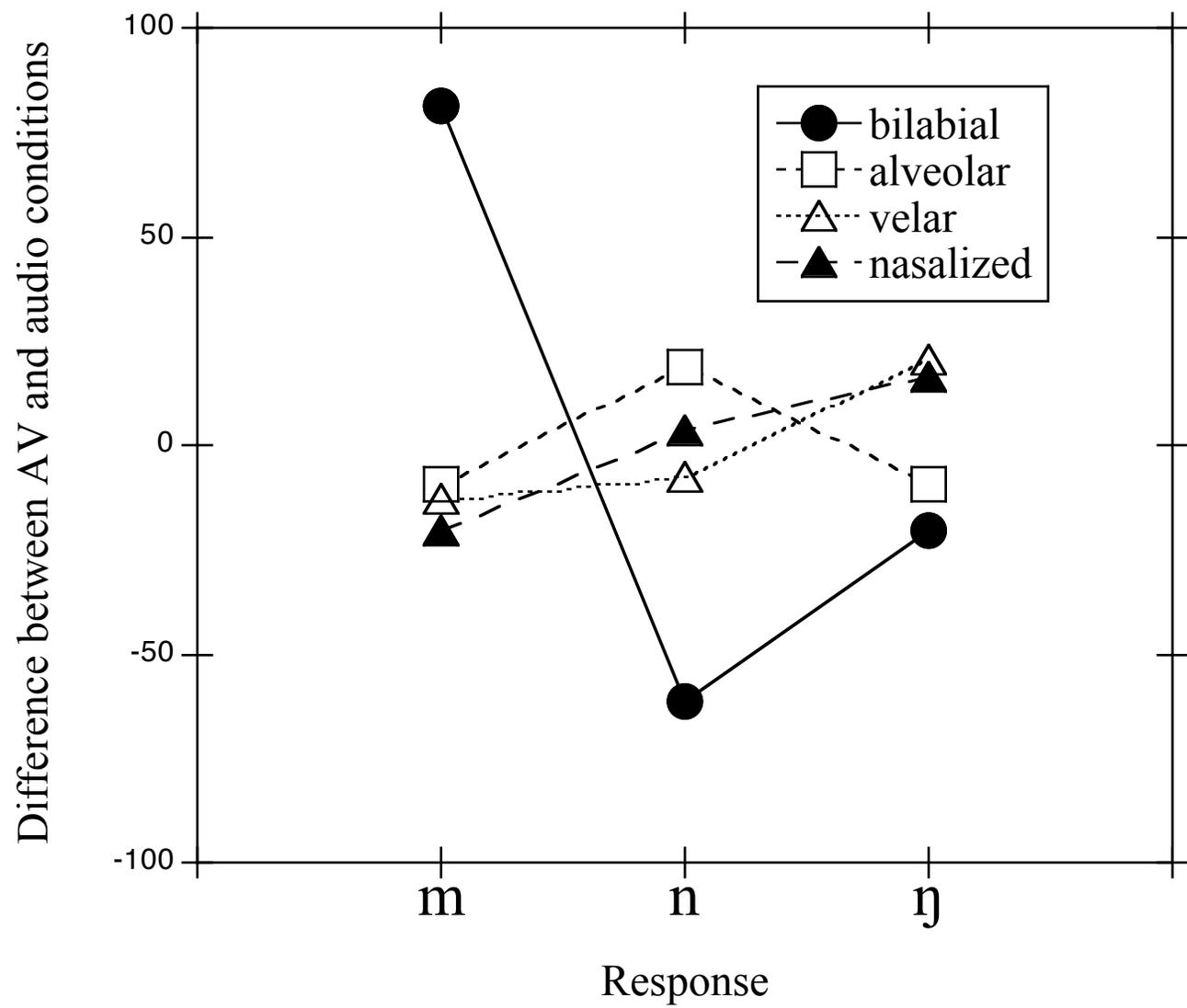
Percent responses for each condition

| | | [m] | | | [n] | |
|-------|-----------|-----------|-----------|-----|-----------|-----------|
| | “m” | “n” | “ng” | “m” | “n” | “ng” |
| Audio | 24 | <u>55</u> | 21 | 21 | <u>56</u> | 23 |
| AV | <u>92</u> | 5 | 3 | 4 | <u>75</u> | 22 |
| | | [ŋ] | | | [x̃] | |
| | “m” | “n” | “ng” | “m” | “n” | “ng” |
| Audio | 17 | 41 | <u>42</u> | 27 | 31 | <u>42</u> |
| AV | 4 | 38 | <u>59</u> | 7 | 38 | <u>56</u> |

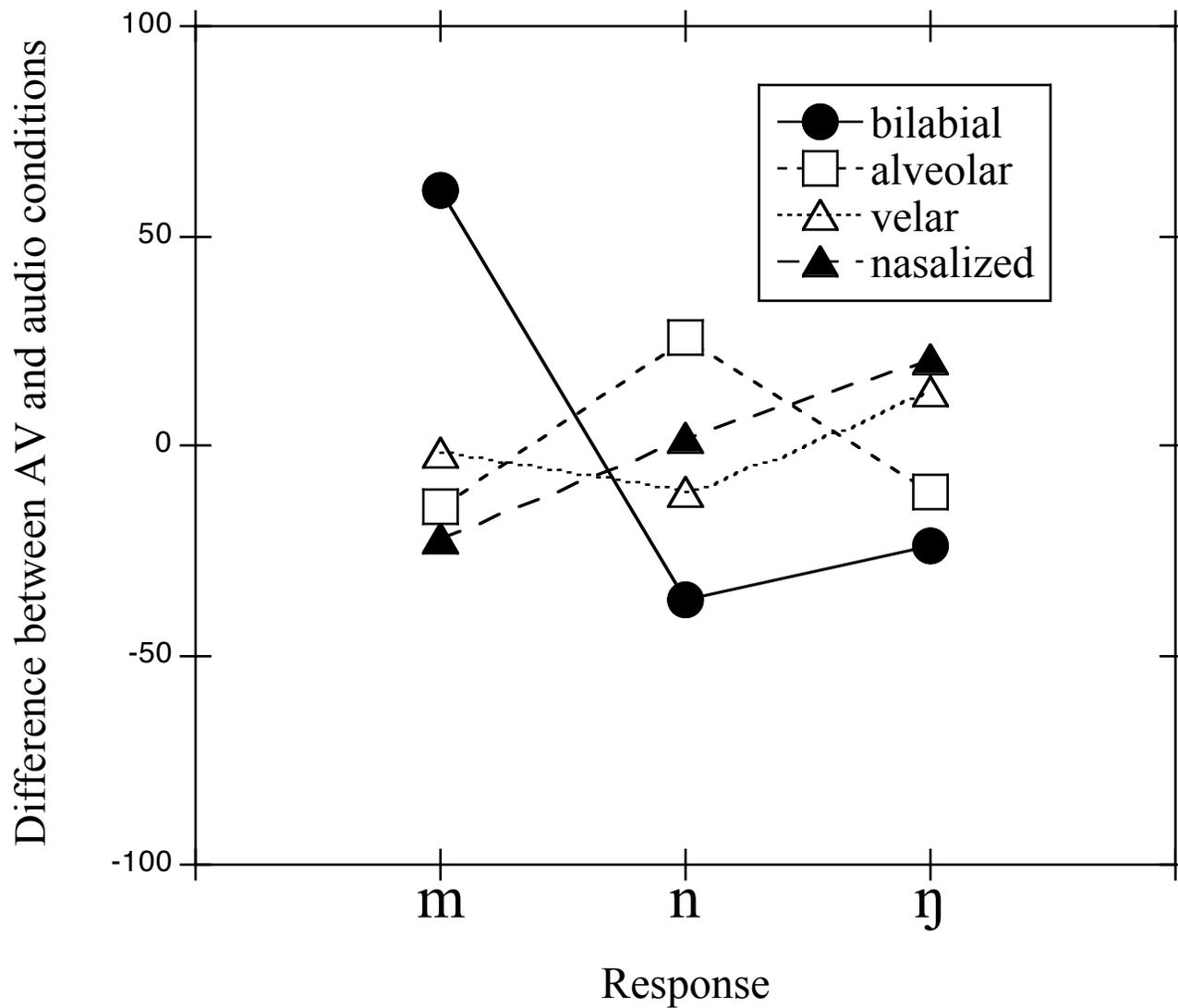
Overall Result



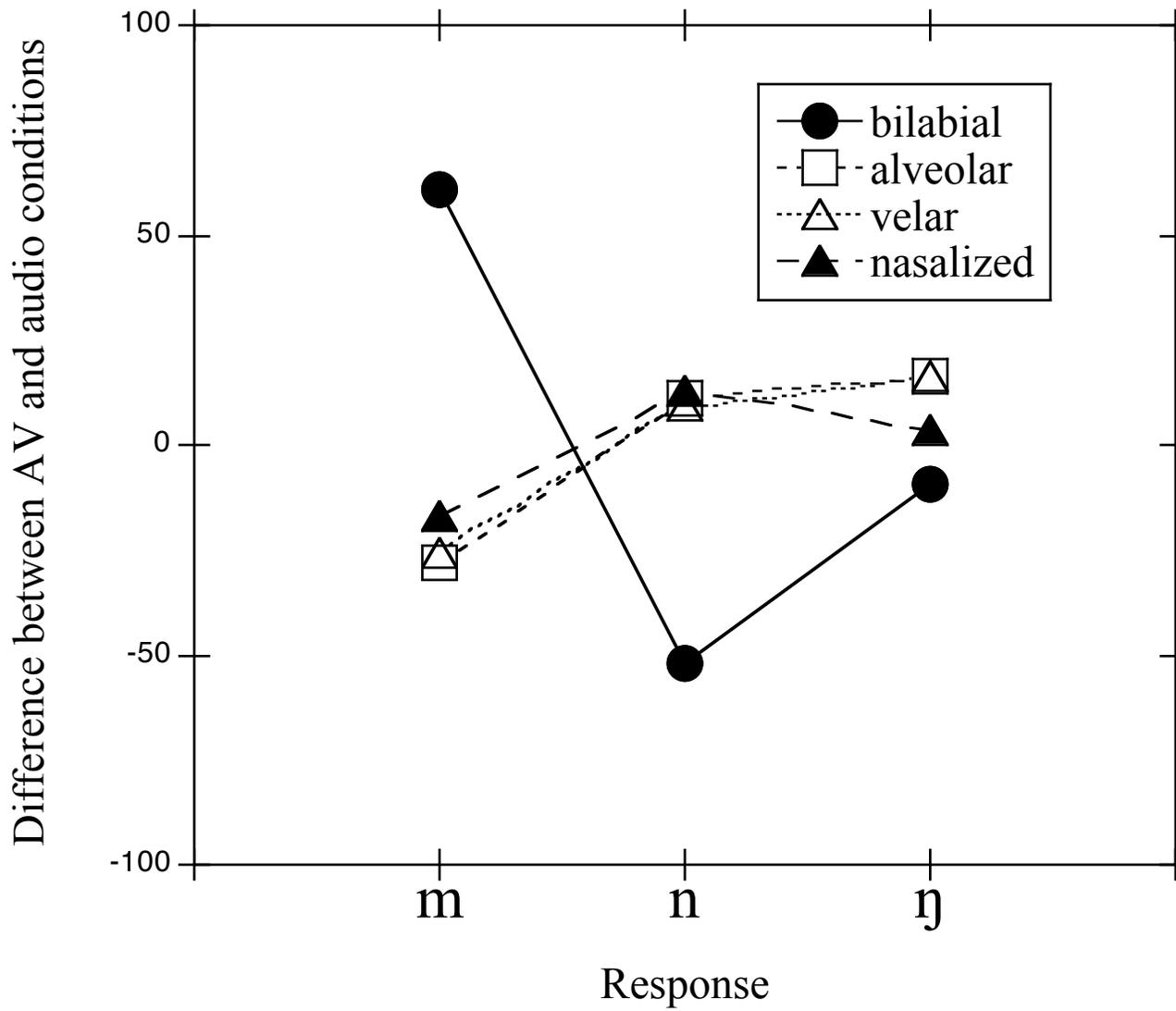
"sum", "sun", "sung" words



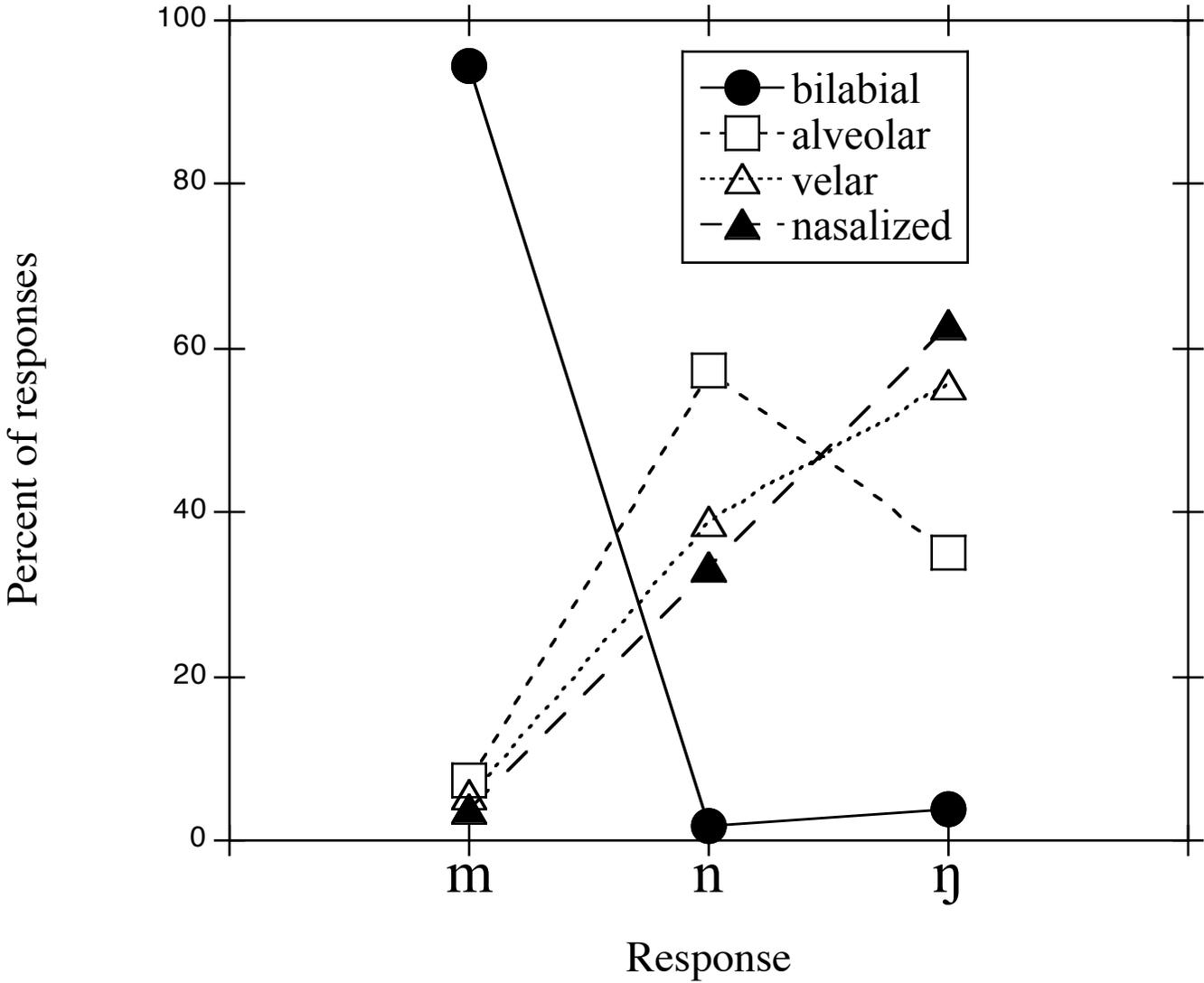
"rom", "ron", "wrong" words



"same", "sane", "sang" words



"same", "sane", "sang" words - AV trials



Some conclusions -

1. Demonstrated visual similarity of [ŋ] and [x̃].
2. When acoustic information is obscured, listeners will use visual information to identify [x̃] as [xŋ].
3. Auditory similarity also exists between [ŋ] and [x̃].
4. Combined effects of auditory and visual similarity may explain excrescent [ŋ].

5. Testing the visual similarity hypothesis for excrescent [ŋ]

c. Experiment 2 - AV perception of nasalized vowels.

- * no masking noise
- * nasalized vowels only
- * three groups of participants:
 - American English - group one (web experiment)
 - 21 participants,
 - 14 in the audio condition, 7 in the AV condition
 - American English - group two
 - 14 who participated after responding in experiment 1,
 - 7 in the audio condition, 7 in the AV condition
 - French - (web experiment)
 - 23 participants
 - 10 in the audio condition, 13 in the AV condition.

▫ *Stimuli.* A native speaker of English (who has had experience with both French and phonetics) was videotaped speaking each of three CV syllables with a word-initial /h/ followed by a nasalized vowel: /hã/, /hõ/, /hẽ/.

▫ *Task.* Identify the (nonexistent) final nasal segment as “m”, “n” or “ng”.

- * Audio condition - heard sound files of these three productions
- * AV condition - saw movie files of these three productions

Results of experiment 2. Percentage of subjects who responded “m”, “n” or “ng” in each condition.

| <u>English listeners</u> - first group | “m” | “n” | “ng” |
|---|-----|------------------|------------------|
| Audio | 19 | 40 | 40 |
| AV | 10 | <u>71</u> | 19 |
| <hr/> | | | |
| <u>English listeners</u> - second group | | | |
| Audio | 14 | <u>57</u> | 29 |
| AV | 10 | 43 | <u>48</u> |
| <hr/> | | | |
| <u>French listeners</u> | | | |
| Audio | 20 | <u>77</u> | 3 |
| AV | 0 | <u>95</u> | 5 |

Experiment 2 conclusion

French based their responses on spelling conventions.

For the Americans, experience in experiment 1 (with masked consonants that could sometimes - e.g. [m] - be identified clearly in the movies) seems to have affected their use of visual information.

Audio presentation of nasalized vowels did not lead to [ŋ] responses, while in one condition - with sensitized listeners - AV presentation does favor hearing [ŋ] for nasalized vowels.

6. Overall Conclusion

Final [ŋ] is visually similar to nasalized vowels.

Acoustic similarities noted by Ohala (1975) seem to fail to predict listeners' behavior in experiment 2.

The visual similarity hypothesis for excrescent [ŋ] is thus supported.

However, the conditions in which visual similarity might be relevant in “the wild” may be limited:

- noisy conditions

- visual contrast with other possible articulations

- Auer, Jr., E. T. (2002) The influence of the lexicon on speechread word recognition: Contrasting segmental and lexical distinctiveness. *Psychonom. Bull. Rev.* 9(2), 341-347.
- Brancazio, L. (1998). 'Contributions of the lexicon to audiovisual speech perception, unpublished Ph.D., dissertation, University of Connecticut.
- Burnham, D. & Dodd, B. (2004) Auditory-visual speech integration by pre-linguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, 44, 209-220.
- Clements, G. N., and Hume, Elizabeth (1995) The internal organization of speech sounds. In *The Handbook of Phonological Theory*, ed. John A. Goldsmith, 245-306. Cambridge, MA, and Oxford, UK: Blackwell.
- Grant, K.W., and Walden, B.E. (1996). Evaluating the articulation index for auditory-visual consonant recognition. *J. Acoust. Soc. Am.* 100, 2415-2424. Source of auditory and a-v confusion matrices
- Grant, K. W., Walden, B. E., and Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *J. Acoust. Soc. Am.* 103, 2677-2690.
- Halle, Morris (1995) Feature geometry and feature spreading. *Linguistic Inquiry* 26:1-46.
- Hampson, M., Guenther, F. and Cohen, M. (1998) Visual influences on the perception of alveolar/velar place discrimination. *JASA* 104, 1854.
- Howe, Darin (2004) Vocalic Dorsality in Revised Articulator Theory. Ms, Univ. of Calgary.
- Lachs, L. and Pisoni, D.B. (2004) Specification of cross-modal source information in isolated kinematic displays of speech. *J. Acoust. Soc. Am.* 116, 507 (2004)
- Massaro, D. W. (1998). *Perceiving Talking Faces: From speech perception to a behavioral principle*. MIT, Cambridge, MA.
- McGurk, H., and MacDonald, J. (1976). "Hearing lips and seeing voices," *Nature* (London) 264, 746-748.
- Ohala, J. J. 1975. Phonetic explanations for nasal sound patterns. In: C. A. Ferguson, L. M. Hyman, & J. J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*. Stanford: Language Universals Project. 289 - 316.
- Ohala, J. J. 1981. The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Ling. Soc. 178 - 203.
- Rice, Keren (1996) Default variability: the coronal-velar relationship. *Natural Language and Linguistic Theory* 15. 493-543.
- Rosenblum, L. D. (1994) How special is audiovisual speech integration? *Curr. Psychol. Cognition* 13(1), 110-116.
- Sagey, Elizabeth. (1986) The representation of features and relations in nonlinear phonology, MIT: Doctoral dissertation.
- Schwartz, J.-L., Robert-Ribes, J., and Escudier, P. (1998). "Ten years after Summerfield: A taxonomy of models for audio-visual fusion in speech perception," in *Hearing by Eye II*, edited by R. Campbell, B. Dodd, and D. Burnham. Psychology, East Sussex, UK, pp. 85-108.
- van der Torre, Erik Jan. (2003) *Dutch Sonorants: the role of place of articulation in phonotactics*. Utrecht, the Netherlands: LOT.