Consonant Confusability: An MEG Study
Author(s): Valentine Hacquard and Mary Ann Walter
Proceedings of the Twenty-Ninth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Phonetic Sources of Phonological Patterns: Synchronic and Diachronic Explanations (2003) pp. 535-544

Consonant Confusability: An MEG Study*

VALENTINE HACQUARD and MARY ANN WALTER
Massachusetts Institute of Technology

0. Introduction

Numerous behavioral experiments have investigated consonant confusability by evaluating error rates in identification of phonemes masked with noise. These studies have found that the pair of nasals, /m/ and /n/, is more confusable than the corresponding oral pair /b/ and /d/ (Miller & Nicely 1955, Wang & Bilger 1973). The members of the two pairs differ in precisely the same way in terms of abstract phonological representations – both involve a change from labial to coronal place of articulation. Thus the difference in confusability between the two pairs of consonants has been explained in acoustic terms. The nasal pair is acoustically more similar for at least two reasons. First, there is a perseveration or coarticulation effect of the nasal on the following vowel. Second, there is a reduction of acoustic distance (mainly with respect to the second formant) between nasalized vowel transitions compared to oral vowel transitions (Wright 1986). Due to this acoustic similarity, the nasal pair is more confusable and, we assume, perceived to be more similar.

In this study, we examine subjects' perception of similarity for the same group of contrasts – /ma/ and /na/ versus /ba/ and /da/ – but with a much finer-grained time-course and level of discrimination than a behavioral judgment. We use magnetoencephalography (MEG) brain-imaging to evaluate variation in two early auditorily-evoked neural components – the M100 response and the mismatch field response (MMF). MEG records the magnetic field generated by electrical activity in the brain, and is able to provide a millisecond-by-millisecond picture of brain activity. Our goal is to shed light on the earliest available stage of perception with respect to the relative roles of abstract phonological features and acoustic similarity. If the latter proves relevant at this stage, then brain-imaging data can be used to test proposals for similarity metrics, and answer questions about the basic units of contrast and their relation to each other.

---

* We would like to thank Alec Marantz, Diana Sonnenreich, Donca Steriade, Karen Froud, Linnaea Stockall, and Pranav Anand. We are extremely indebted to Ken Stevens for help synthesizing the stimuli.

1. Two models of similarity

1.1. An acoustic-based model

Much current work in phonology motivates phonological processes by appealing to perceptual, acoustic factors. Similarity along acoustic dimensions has been argued to underlie phonological processes such as metathesis, assimilation, and phonotactic constraints. Hume (1998) argues that confusability considerations motivate metathesis processes, so that a segment is positioned in such a way as to maximally enhance its perceptual salience. Steriade (1999) proposes that distance along a perceptually-based similarity metric (the P-map) correlates with assimilation processes. In the context of the confusability studies reported above, the findings of Mohanan (1993) and Jun (1995) that nasals are relatively more susceptible to place assimilation than stops generally is unsurprising.

These confusability and assimilation facts, as well as studies of offline similarity judgments (Hura et al. 1992), support a model of similarity based on acoustic, perceptual factors. Such a model is dependent on phonetic context, but independent of the phonological inventory. According to an acoustic-based model, nasal confusability is explained in perceptual terms: although nasality itself is highly salient, persevering nasality alters the F2 transition into the following vowel, which is an important cue for place of articulation.

1.2. A model based on natural classes (Frisch 1996)

Frisch et al. (1997) and Frisch (1996) propose an alternative model of similarity, which differs from the acoustic-based model in that it is independent of phonetic context, but dependent on the phonological inventory and its structure. This similarity metric is computed according to natural classes, as illustrated in (1):

(1) $$\text{Similarity} = \frac{\text{shared natural classes}}{\text{shared + unshared natural classes}}$$

The similarity measure in (1) computed for English consonants gives the values in (2), where the ranking of similarity contradicts that of the acoustic-based model: nasals are equally (if a .01 difference is insignificant) or less similar to each other than orals.

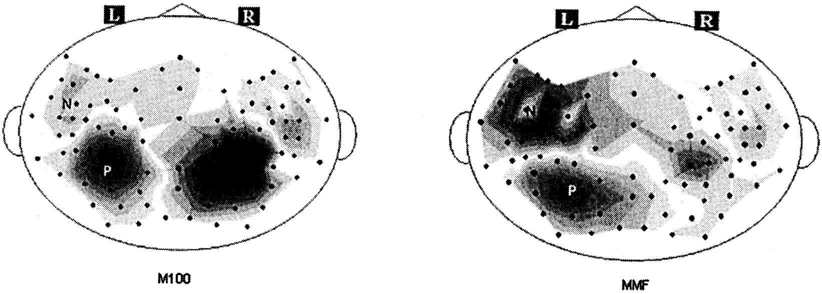
(2) b/d (= .29) > m/n (= .28), where 1 is identity (maximal similarity).

In our study, we will thus examine whether a difference in similarity, along any of the metrics discussed above, correlates with a difference in an auditory brain response. If it does, then brain data can be used to substantiate proposals for similarity metrics, such as the two proposed above, and better understand their internal organization.

2. Magnetoencaphalography (MEG)

MEG measures the magnetic fields generated by electrical activity in the brain, specifically by potentials in the apical dendrites of pyramidal cells in the cortex. The main component we look at here is the mismatch field response or MMF, an automatic, auditory brain response evoked by a deviant stimulus following a sequence of standards, peaking ~180-250 ms after the difference point between a standard and a deviant. We also look at the M100, an automatic auditorily-evoked response that peaks ~100 ms post-stimulus onset.

Figure 1 – Magnetic field distributions of the M100 and MMF in response to speech stimuli for one representative subject. The letter P indicates the positive field (emerging from the brain) and the letter N the negative field (entering the brain).



Previous MEG studies have identified several properties of the MMF. Sharma & Dorman (1999) and Phillips et al. (2000) show that the same VOT span crossing a phonemic category boundary evokes a far greater MMF than one that doesn't. Näätänen et al. (1997) show that a small acoustic difference crossing a phonemic category boundary evokes a far greater MMF than a large one that does not cross such a boundary. It appears, then, that phonological difference outweighs acoustic difference for the MMF response.

However, we would like to know whether similarity distinctions matter when category is kept constant. In our experiment, both the oral pair and the nasal pair involve crossing a phonological category (i.e., place of articulation from labial to coronal and vice-versa). If the gap within one pair differs from the gap of the other pair, then some other factor plays a role. We hypothesize that the factor is perceptual similarity.

3. Materials and methods

3.1. Participants

Seventeen subjects, all students and employees at MIT, all of whom gave informed consent, participated in the experiment. The sample included 9 males and 8 females. All subjects were right-handed and had no history of hearing or neurological disorders. Eleven were native speakers of English. Six were native speakers of languages which also have the relevant contrasts (Czech (2), Spanish, Brazilian Portuguese, Russian, and Persian). They were also fluent in English. One subject was later excluded due to technical problems that arose during the experiment.

3.2. Stimuli

Stimuli for the present experiment consisted of four CV syllables – /ba/, /da/, /ma/, /na/. The syllables were synthesized using the program KLSyn. The stimuli were correctly identified and discriminated by speakers of languages with the relevant contrasts, thus ensuring that they were good exemplars.

3.3. Procedure

Subjects lay supine in a magnetically shielded room while stimuli were presented binaurally over earphones. Evoked magnetic fields were recorded using the MIT/KIT whole-head biomagnetometer array, with 93 axial gradiometers.

Before beginning the experiment, subjects listened to a 1kHz tone presented 100 times, for the purpose of identifying and localizing the M100, and helping to select RMS sensors for the analysis (see below). We used a mismatch detection task (oddball paradigm) where a series of four identical precursor stimuli, separated by inter-stimulus intervals of 400 ms, was presented followed by a fifth stimulus, either identical to the previous four (control condition or *standard*), or different from it with respect to place of articulation of the initial consonant only (*deviant* condition). The stimuli were arranged into the eight conditions illustrated in Figure 2. The subjects heard each condition 30 times, in random order.

Trial presentation was randomized by Psyscope script. Subjects were instructed to press one button when the fifth item of each set was the same as the previous ones, and a second button if it was different. The trials were divided into six blocks of forty trials. Between blocks, the subject was given a break of self-determined duration.

Data were sampled at 1000 Hz, with acquisition between 1 and 200 Hz. The recording for each participant lasted approximately 25 minutes. The raw data was then subjected to a noise reduction routine to eliminate measured magnetic activity from external sources. Responses to stimuli were averaged by stimulus condition separately, in 700 ms windows keyed to the onset of the stimulus: 100 ms pre-, 600 ms post onset. The averaged signal was subjected to a low-pass filter below 30 Hz and adjusted to baseline using a 100 ms pre-stimulus interval.

Figure 2 – Conditions (8x30=240 trials)

1) ba	(400 ms)	ba	(400 ms)	ba	(400 ms)	ba	(400 ms)	da	deviant
2) da	(400 ms)	da	(400 ms)	da	(400 ms)	da	(400 ms)	da	standard
3) da	(400 ms)	da	(400 ms)	da	(400 ms)	da	(400 ms)	ba	deviant
4) ba	(400 ms)	ba	(400 ms)	ba	(400 ms)	ba	(400 ms)	ba	standard
5) ma	(400 ms)	ma	(400 ms)	ma	(400 ms)	ma	(400 ms)	na	deviant
6) na	(400 ms)	na	(400 ms)	na	(400 ms)	na	(400 ms)	na	standard
7) na	(400 ms)	na	(400 ms)	na	(400 ms)	na	(400 ms)	ma	deviant
8) ma	(400 ms)	ma	(400 ms)	ma	(400 ms)	ma	(400 ms)	ma	standard

3.4. Data analysis

Both button-press responses and reaction times were recorded. Reaction times were calculated from the onset of the fifth stimulus. Incorrect trials and RTs deviating over 2.5 SD from the mean for that particular participant were excluded from the behavioral analysis.

In the analysis of the MEG data, averaged signals were first visually inspected to identify dipolar field distributions that appeared consistently across experimental conditions and across participants. Such distributions were identified in two time windows: the M100 window (150-170 ms) and the MMF window (225-275 ms). As shown in Figure 1, the MMF, being evoked by speech stimuli, was strongly left-lateralized (an MMF response to non-speech stimuli is more evenly distributed across both hemispheres). The amplitudes and latencies of these components were recorded by calculating the root mean square (RMS) field strength from the sensors that covered the field pattern of the MMF/M100 in the left hemisphere. The sensors used for the RMS analysis were selected for each subject by creating a grand average of all 8 conditions and choosing those sensors that showed the clearest dipolar distribution and held constant across conditions within a subject. The number of sensors used varied between 4 and 6. We then compared the RMSs of the selected sensors for the response to the two oral deviant conditions to those of the corresponding standards (e.g., the *da* response following a series of *ba* to the *da* response following a series of *da*), and likewise for the nasal conditions.

4. Results

4.1. Behavioral responses

Analyses of variance (ANOVA) [(Condition[1]) X (Manner[2])] were performed for reaction times and accuracy. Deviants overall received significantly more errors

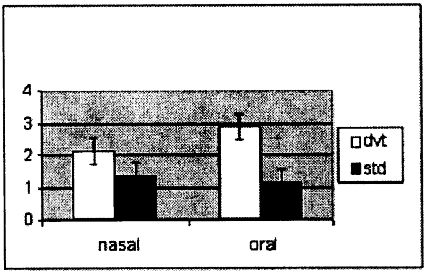
---

[1] Standard or deviant

[2] Oral vs. nasal

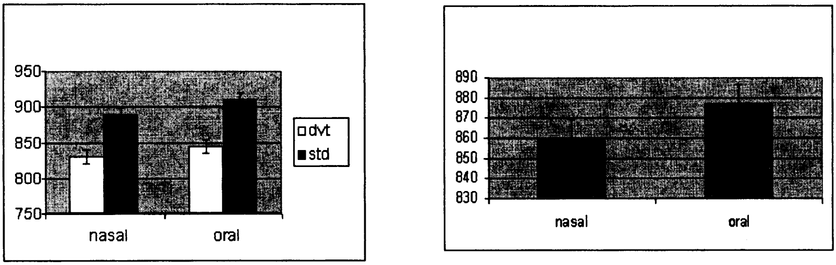
than standards (p=.0009) (Figure 3). No effect for manner was observed (p=.3538). Therefore, the results of previous behavioral studies, that nasals are more confusable than orals, were not replicated. There are several reasons why this might have been the case. First of all, the stimuli were synthesized so as to be clear and unambiguous tokens of the target syllables, to maximize the desired brain responses. No filtering was done, nor anything else to adversely affect the signal-to-noise ratio, as is typically done in behavioral confusability studies. Therefore, it is unsurprising for error rates to be unrevealing. Second, the relatively small number of trials (30 per condition) may have been a factor.

Figure 3 – Error rates. Y-axis: number of errors.



Reaction time to deviants overall was significantly faster than to standards (p≤0001). Reaction time to nasals was significantly faster than to orals (p=.0197) (Figure 4). No [(Condition) X (Manner)] effect was found. Some subjects reported that they took longer to respond to 'same' trials because they were unsure whether or not the stimulus set was complete, which could obscure reaction time patterns.

Figure 4 – Reaction times. Y-axis: time, in ms.



4.2. Electrophysiological results

4.2.1. M100 response

This response proved very difficult to locate when looking at individual subject responses. An ANOVA [(Condition) X (Manner)] was performed for amplitude and latency. No effect was found for [(Condition) X (Manner)] for either latency or amplitude, but there was a main effect for manner: overall, nasals have significantly greater amplitude than orals in the M100 time window ($p \leq .0001$).

4.2.2. MMF response

An ANOVA [(Condition) X (Manner)] was performed for amplitude and latency. No effect for latency was found. We found a main effect for amplitude for condition: deviants overall have significantly greater amplitude in the MMF time window than standards ($p \leq .0001$), as expected. An effect for manner was also found: orals overall have a significantly greater amplitude in the MMF window than nasals ($p \leq .0001$) (Figure 5). Finally, the ANOVA yielded an effect for [(Condition) X (Manner)]: as illustrated in Figure 6, the MMF/baseline gap was significantly greater for oral consonant pairs than for nasals ($p=.0399$).

Figure 5 – MMF: The following charts show a single subject's averaged response to oral deviants and standards in the top chart, and to nasal deviants and standards in the bottom chart. The MMF window (225 ms – 275 ms) shows an MMF for both orals and nasals (Y-axis: magnetic field strength, in arbitrary units proportional to femtoTesla; X-axis: time, in ms). The bar graph on the right shows the differences in peak amplitude between standards and deviants for orals and nasals.

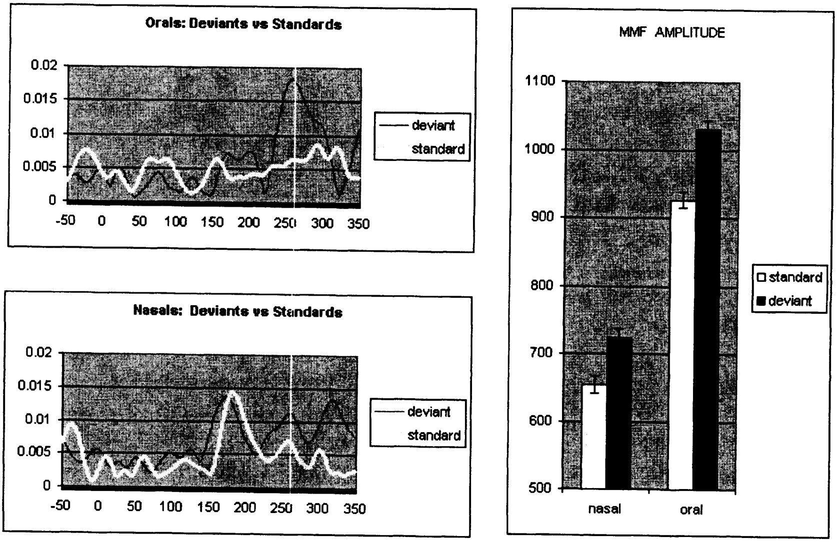
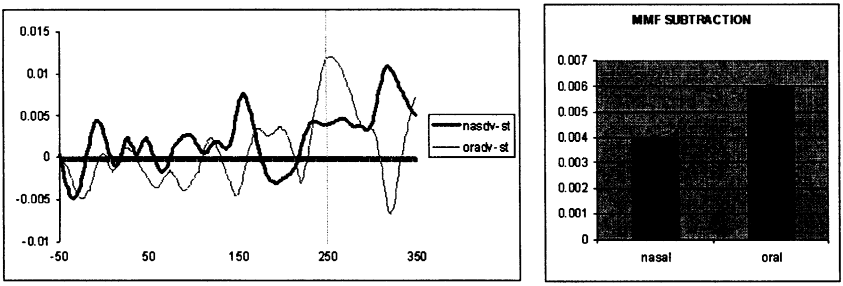


Figure 6 – MMF comparison: The following chart shows a single subject's MMF waveline (subtraction of the deviant minus standard) for nasals (thick line) and for orals (thin line). The oral waveline is higher in the MMF time window (Y-axis: magnetic field strength, in arbitrary units proportional to femtoTesla; X-axis: time, in ms). The bar graph illustrates the same subject's MMF subtraction at peak amplitude in the MMF window for nasals and orals.



5. Discussion

The predictions made by the different models of similarity discussed earlier were as follows:

- According to a perceptual similarity framework, the MMF-baseline gap (i.e., deviant minus standard) should be larger for oral consonant pairs than for nasals.
- According to a natural-class-based one such as the one presented in Frisch (1996), the gaps should be equivalent for oral and nasal pairs, or the nasal gap should be larger.
- If abstract phonological features were the only relevant factor in perceptibility at this stage, the gaps should be equivalent.

We found that the oddball paradigm evoked a stronger MMF response for orals than nasals, and thus that orals are perceived as more different from one another than nasals at this stage, the earliest available. It appears then that phonological category is not the only relevant factor in perception at this stage: acoustic similarity also plays a role. Finally, it is a perceptual similarity metric that appears to be operating at this time period, rather than a natural-class one, or feature-counting one.

It remains unclear why the M100 response was so weak for most of the subjects. However, previous studies have also reported difficulties in localizing it (Bruening et al. 2001).

6. Conclusion and future research

Our results suggest that the MMF can be used to test proposals about degree of perceptual distance in acoustic-based similarity frameworks. Our experiment showed that, once we control for phonological category, perceptual similarity matters. In future work, we would like to explore the possibility that a Frisch-type of similarity also plays a role, which is here being disguised by the stronger role of perceptual similarity, just as perceptual similarity is generally disguised by phonological category. Our next experiment will isolate Frisch-style similarity as a variable, by controlling acoustics and testing the same phonological contrast with speakers of languages that have the contrast, but whose inventories differ in other ways.

References

Bruening B., E. Flagg & V. Lin 2001 "An MEG Study of Tone Processing Asymmetries in English versus Mandarin Speakers" *Proceedings of the Northeast Linguistic Society Conference*, 31

Frisch S. 1996, "Similarity and frequency in phonology", unpublished PhD. Dissertation, Northwestern University

Frisch S., M. Broe and J. Pierrehumbert 1997 "Similarity and Phonotactics in Arabic" *Rutgers Optimality Archive 223*

Hume E. 1998 "The role of perceptibility in consonant/consonant metathesis." *Proceedings of the West Coast Conference of Formal Linguistics,* 17, 293-307

Hura S., B. Lindblom and R. Diehl 1992 "On the role of perception in shaping phonological assimilation rules," *Language and Speech* 35 (2), 59-72

Jun, J. 1995 "Place assimilation as the result of conflicting perceptual and articulatory constraints" *Proceedings of the West Coast Conference of Formal Linguistics,* 14, 221-237

Miller G. A. & P. E. Nicely 1955 "An analysis of perceptual confusions among some English consonants." *Journal of the Acoustical Society of America, 27,* 338-352

Mohanan, K. P. 1993 "Fields of attraction in phonology" in John Goldsmith (ed.) *The Last Phonological Rule*, University of Chicago Press

Näätänen R., A. Lehtoskoskl, M. Lennes, M. Cheour, M. Huotilainen, A. Ilvonen, M. Vainio, P. Alku, R. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen & K. Alho 1997 "Language-specific phoneme representations revealed by electric and magnetic brain responses." *Nature,* 385, 432-434

Phillips C., T. Pellathy, A. Marantz, E. Yellin, K. Wexler, D. Poeppel, M. McGinnis & T. Roberts 2000 "Auditory cortex accesses phonological categories: an MEG mismatch study" *Journal of Cognitive Neuroscience,* 12

Sharma A. & M. F. Dorman 1999 "Cortical auditory evoked potential correlates of categorical perception of voice-onset time." *Journal of the Acoustical Society of America,* 106, 1078-1083

Steriade D. 1999 "P-map effects in assimilation" in Elizabeth Hume and Keith Johnson (eds.) *The role of speech perception phenomena in phonology,* Academic Press

Wang M. D. & R. C. Bilger 1973 "Consonant confusions in noise: a study of perceptual features." *Journal of the Acoustical Society of America,* 54, 1248-1266

Wright, J. T. 1986 "The behavior of nasalized vowels in the perceptual vowel space" in J. J. Jaeger and J. J. Ohala (eds.) *Experimental Phonology*, Orlando: Academic Press, 45-67

Valentine Hacquard
Mary Ann Walter
Department of Linguistics and Philosophy
77 Massachusetts Ave, Bldg E39-245
Massachusetts Institute of Technology
Cambridge, MA 02139

hacquard@mit.edu
walterma@mit.edu