Title

Imitation of coarticulatory vowel nasality across words and time.

Permalink

https://escholarship.org/uc/item/9wf6b540

Journal

Language, Cognition and Neuroscience, 32(6)

ISSN

2327-3798

Authors

Zellou, Georgia
Dahan, Delphine
Embick, David

Publication Date

2017

DOI

10.1080/23273798.2016.1275710

Peer reviewed

Published in final edited form as:

Lang Cogn Neurosci. 2017 ; 32(6): 776–791. doi:10.1080/23273798.2016.1275710.

Imitation of coarticulatory vowel nasality across words and time

Georgia Zellou,
University of California, Davis

Delphine Dahan,
University of Pennsylvania

David Embick
University of Pennsylvania

Abstract

We investigated phonetic imitation of coarticulatory vowel nasality using an adapted shadowing paradigm in which participants produced a printed word (target) after hearing a different word (prime). Two versions of primes with nasal codas were used: primes with a natural degree of vowel nasality and hypernasalized primes. The version of the prime participants heard varied, whether consistent with their past experience with nasality from the talker or inconsistent, and the duration of delay between prime and target. People spontaneously modify coarticulatory nasality to resemble that demonstrated in the prime they were exposed to. Furthermore, this imitation also reflects the degree of nasality demonstrated by overall experience with the speaker's vowels. The influence of past experience on imitation increases with increased delay between prime and target. Imitation of another speaker appears to involve tracking general articulatory properties about the speaker, and not solely what was specific to the most recent experience.

Keywords

Phonetic imitation; Coarticulation; Memory; Cross-word generalization

Imitation is a perceptually-guided action (Meltzoff & Moore, 1997). Vocal imitation involves the mapping of an auditory signal onto an articulatory program. The mapping is mediated by an internal encoding onto a representation of the auditory signal that gives rise to the production of an utterance that shares some phonetic properties with the signal. Some scholars have referred to this internal representation as the 'target of imitation'. Although the term 'imitation' can be used to describe a voluntary action, we will use it to describe the involuntary, and most likely unconscious, phenomenon in which the speech of a talker is altered to resemble that of a model talker. In its simplest form, imitation can be observed in a shadowing task when people repeat (or shadow) a spoken word immediately after hearing it. Under these circumstances, the target of imitation simply consists of the utterance being shadowed. However, imitation of a talker's pronunciation has been observed on words that had not been specifically heard before (e.g., Shockley, Sabadini, & Fowler, 2004; Nielsen,

---

Corresponding Author: Georgia Zellou, University of California, Davis, Linguistics Dept., 469 Kerr Hall, One Shields Avenue, Davis, CA 95616. gzellou@ucdavis.edu, Phone: +1-530-752-0966.

2011), as well as when there was a substantial delay (i.e., as long as several days) between exposure to an utterance and its subsequent imitation (e.g., Goldinger & Azuma, 2004). There is also evidence that the speech of adults can be altered to reflect phonetic characteristics of the dialect or language spoken in the environment these individuals are immersed in later in life (Sancier & Fowler, 1997; Evans & Iverson, 2007). Finally, the speech of interlocutors in conversational settings has revealed phonetic convergence (Pardo, 2006; Pardo, Jay, & Krauss, 2010). These examples of imitation must rely on a general representational target of the interlocutors' speech, one that has been built from specific utterances but that can abstract away from them. The present study examines how the target of imitation evolves with experience with a talker.

The question of what constitutes the target of imitation has been discussed in the literature but there is relatively little empirical work directed at systematically investigating its properties. In his seminal work, Goldinger (1998) used a shadowing technique in which people were asked to repeat a series of words (so called primes) presented to them through headphones. Imitation of the primes was assessed by testing an independent group of listeners in an AXB two-alternative forced choice task where X was the prime and A and B were recordings of the same word produced by the shadowing-task participant at baseline (before the shadowing task) and when shadowing the prime. Participants in the AXB task selected whether A or B was a better imitation of X. The rate at which the shadowed token was selected over the baseline one quantified the degree to which participants had imitated the prime.

Basing his predictions on an episodic model of memory storage and retrieval (Hintzman, 1984, 1988), Goldinger hypothesized that the perception of the spoken prime gives rise to the reactivation of a constellation of memory traces or episodes, each one activated in proportion to its similarity to the prime. These episodes, together with the memory trace associated with the prime itself, form an *echo*, or in our terms, a target of imitation, which is assumed to drive production. Thus, Goldinger contended, the influence of the prime on production depends on the number of other traces concurrently activated. In support of this claim, he showed that low-frequency prime words were better imitated than high-frequency ones: Because people have had relatively few encounters with a low-frequency word, the reasoning goes, the echo contains few past episodes with the word beside its exposure as prime, with the result that the prime exerts a disproportionate influence on its subsequent articulation. By comparison, imitation of a high-frequency prime is more limited because the echo driving imitation is populated by many memory traces beside that of the prime. This interpretation was further supported by the finding that increasing the number of exposures to a given prime word before shadowing increased the degree of imitation. This account rests on the assumption that which word the prime corresponds to is a major factor in determining which memory traces participate in the echo, i.e., what we have called the target of imitation.

Goldinger and Azuma (2004) corroborated this conclusion in a study using a variant of the shadowing methodology. Imitation resulting from exposure to a talker's word pronunciation was assessed not on participants' immediate shadowing but on a recording of their reading out loud the prime words a week after passive exposure to these primes. The design enabled the authors to examine participants' pronunciation of words before and after hearing the talker's pronunciation of those words, but also participants' pronunciation of words that they had not heard pronounced. Following Goldinger (1998)'s AXB procedure for assessing imitation, Goldinger and Azuma found that participants imitated only those specific words that they had heard pronounced by the talker. This finding, along with the influence of word frequency and number of repetitions with a given prime, suggests that the target for imitation—or, in Goldinger's terminology, the echo—is largely constrained by segmental characteristics of the word being produced; generalization across words is limited. However, other findings mitigate this conclusion.

First, and contra Goldinger and Azuma (2004), several studies have found evidence that the pronunciation of words can become similar to that of a model talker even when people did not hear the talker pronounce these specific words (e.g., Nielsen, 2011; Shockley et al., 2004), and such imitation can be just as robust as that observed on words that had been heard. Thus, people can generalize what they learn about the talker's pronunciation from the words they have heard to different words. A possible explanation for the discrepancy between these studies and the Goldinger and Azuma results, discussed by Shockley et al. themselves, concerns the homogeneity of the set of words used: Shockley et al. restricted themselves to bisyllabic words starting with a voiceless stop consonant. Homogeneity may facilitate the generalization of the pronunciation displayed by those instances to words that are phonetically similar but that have not been specifically experienced.

This possibility is corroborated by results from Nielsen (2011), who exposed participants to words characterized by an artificially extended aspiration period in the production of an initial /p/. She quantified imitation by measuring participants' pronunciation of words with an initial voiceless stop consonant before and after exposure to those stimuli. Some of the words also started with /p/ but were not part of the set people heard pronounced; yet others started with a /k/ and were new as well. She observed that aspiration duration was increased not only on those words that participants had heard pronounced with extended aspiration, but also on novel /p/-initial words and, in addition, on /k/-initial words, the latter effect being significantly smaller than the former. Thus, it appears that imitation of a pronunciation feature experienced on some words can generalize to similar words, evidently in a way that is proportional to the degree of similarity. This generalization must result from the extraction of articulatory features from exposure to specific word instances.

The influence of more abstract factors in the composition of the target of imitation is further supported by evidence of delayed imitation, as reported by Goldinger and Azuma (2004). There, imitation was established on participants' pronunciation of words a week after exposure to auditory recordings of those words from a model talker. The authors attributed this long-lasting effect to the role played by the context in which spoken words had been initially experienced in determining the composition of the echo driving imitation: Because people were asked to produce words in the same context as the context of their prior auditory exposure to those words, the memory episodes of this exposure, albeit temporally distant, were highly active and therefore played a large role in the echo. Thus, contextual similarity, in addition to segmental content, may be an important factor in determining the composition of the echo.

The composition of the echo—or, in more general terms, of the target that people recruit when imitating a prime—is thus only partially constrained by the prime's segmental make-up. The context and (as demonstrated by cross-word generalization) the talker's identity may also play a role. From these observations, it appears that an internal representation of the speech of a given talker emerges as people accumulate experience with that talker. This, in turn, raises the possibility that this representation captures a sort of aggregate computed over all utterances associated with a given talker, a form of 'central tendency' in the talker's pronunciation of speech. The present study tested this hypothesis by examining how accumulated experience with a talker affects the composition of the target of imitation.

To this end, we varied the order in which specific utterances, all from the same talker, were heard. These utterances, recordings of monosyllabic words acting as primes, varied in the degree of coarticulatory nasality present on the vowel before a nasal coda consonant. The vowel either (1) displayed a natural (albeit greater than all the participants) degree of nasality in anticipation of the upcoming nasal consonant, or (2) had been manipulated to be hyper-nasalized. (Details on hyper-nasalized vowels are given in Experiment 1's materials section.) A trial consisted of the auditory presentation of one of these primes, followed by the visual presentation of a word, the target, which participants were instructed to read out loud. The target was never the same word as the prime but, like the prime, contained a vowel before a nasal consonant. By assessing the degree of nasality in target productions, we asked whether the model driving the production of the target only reflects the degree of nasality of its immediately preceding prime (i.e., greater nasality produced after hyper-nasalized primes than after naturally nasalized ones) or also the cumulative set of utterances heard and their averaged overall degree of nasality.

An important factor to consider when examining the influence of an auditory prime on the production of the following target concerns the delay between prime and target. Goldinger (1998) observed that imitation was greatly reduced when a 3-s delay between the prime and participants' shadowed production was introduced. This, Goldinger argued, was evidence that the content of the echo changes as time elapses because the recruitment and activation of memory traces from long-term memory is a process that evolves over time, thereby changing the composition of the echo. As time passes, more episodes become active, resulting in a decrease in the proportional weight of recent episodes (and that of the prime, in particular) in the composition of the echo. If, as we propose, the echo (or what we have been calling 'target of imitation') is influenced by episodes of past utterances from the same talker as the prime, people's target-word pronunciation is expected to reflect accumulated experience with the talker more, and the immediately preceding prime less, as the delay between prime and target increases.

Experiment 1

The present investigation focused on the imitation of coarticulatory nasality in vowels. The presence of a nasal consonant within a syllable tends to cause spectral changes in the adjacent vowel. These changes are caused by the lowering of the velum, a gesture produced in anticipation of or carryover from the adjacent nasal stop consonant. Coarticulatory vowel nasality results from the nasal coupling that the velum lowering creates during the articulation of the vowel. A useful acoustic correlate of nasal coupling in vowels, developed by Chen (1997), consists of the difference between the amplitude of the first formant peak (denoted A1) and the amplitude of the low-frequency nasal pole FN (denoted P0). This amplitude difference decreases as nasal coupling increases. Using this measure as a proxy for degree of coarticulatory vowel nasality in English, past research has shown that coarticulatory vowel nasality varies in magnitude across speakers and within speakers, and across linguistic and pragmatic contexts (Scarborough, 2013; Podesva et al., 2013; Pycha, 2016; Scarborough & Zellou, 2013; Zellou & Tamminga, 2014; Zellou & Scarborough, 2015). Furthermore, Zellou, Scarborough, and Nielsen (2016) demonstrated that people tend to unconsciously imitate vowel nasality. In their study, participants were asked to repeat monosyllabic words with a nasal coda consonant with a vowel characterized by a high level of nasality (i.e., hyper-nasalized). Compared to their pronunciation of those words' vowels at baseline, i.e., before hearing the hyper-nasalized recordings, people produced the vowels with increased nasal coupling. Thus, people can spontaneously alter the dynamics of their velum-lowering gesture to imitate shadowed utterances.

In the experimental trials in this study, people heard a prime, then saw the orthographic form of another word and produced it. We assessed the degree to which people's pronunciation of vowels in target words changes when produced after hearing different words as primes. Evidence for such nasality imitation would demonstrate people's ability to apply articulatory features extracted from a given context to a new context.

Experiment 1 was designed to assess whether this imitation, although observed on a word different from the one heard in the prime, is limited to the characteristics of coarticulatory nasality exemplified by the immediately preceding prime, or if it can also reveal the influence of a more general target of imitation, one characterized by the overall experience with the model talker accumulated over the course of the experiment. In order to test for prime specificity and overall experience on imitation, we divided trials into blocks and varied which kind of prime was presented in each block. For half of the participants, only naturally-nasalized primes were presented in Block 1; for the other half, only hyper-nasalized primes were heard. In Block 2, the prime types were switched. (A block of filler trials, in which primes and targets were monosyllabic words ending with an oral coda consonant, was inserted between the two test blocks.) With this design, the degree of nasality on each prime in Block 1 is consistent with what participants have heard from the talker (i.e., they have heard only naturally-nasalized or only hyper-nasalized primes); in Block 2, however, nasality exemplified on each prime is different from that experienced before, in Block 1. As discussed in the introduction, if imitation produced on a target word reflects the influence of the immediately preceding prime only, target vowels should be characterized by greater nasality (and therefore, greater change from baseline) after hyper-nasalized primes than after natural primes, with no effect of the block in which the two types of primes were presented. If, instead, imitation is also affected by participants' cumulative experience with the model's utterances, imitation after hyper-nasalized primes in Block 2 should be less than that observed in Block 1 because the target of imitation would incorporate the naturally-nasalized primes from Block 1. Conversely, imitation after naturally nasalized primes in Block 2 should be greater than that observed in Block 1 because of participants' past experience with hyper-nasalized vowels from the same talker in Block 1.

Two versions of each prime were created. One consisted of a natural recording of nasal-final words (e.g., 'bound') produced by a male talker who, based on acoustic assessment of his productions, produces vowels with a high degree of coarticulatory nasality. The other type of primes consisted of recordings of the same words from the same talker but with an artificially greater degree of nasality in the vowels. Following the procedure adopted in Zellou et al. (2016), we created primes with 'hyper-nasalized' vowels by extracting the vowel from the recording of the original word (e.g., 'bound') and the recording of the word or nonword that results from changing the initial consonant of the original word from oral to nasal (e.g., 'mound') and by mixing these two vowels. (More details on the procedure are given later.) The resulting vowel was reinserted to replace the original vowel in the recording of the prime word. The presence of a nasal consonant on either side of the vowel in, e.g., 'mound', causes the velum to stay lowered for the entire duration of the vowel. The resulting vowel has a high degree of nasal coupling because the velum remains lowered over vowels flanked by nasal consonants (Cohn, 1990). If participants' pronunciation of a target word is affected by the specific properties of an immediately preceding prime, greater vowel nasality (compared to participants' baseline) should be observed after hyper-nasalized primes than after naturally nasalized ones.

Materials

Stimuli were tokens of 16 monosyllabic real English words with a vowel nasal-consonant sequence in its coda (see appendix). The words were selected to have non-high vowels because the nasal-pole peak is easier to identify and distinguish from the F1 peak on non-high vowels than it is on high vowels. This, in turn, facilitates the extraction of the A1–P0 acoustic measure of nasality on those vowels. In addition, words were selected to be highly familiar and relatively frequent (with familiarity ratings of 6 or greater on the 7-point Hoosier Mental Lexicon scale and a frequency of 35 per million words, as estimated by Brysbaert and New, 2009). Because shadowing high-frequency words tend to result in less imitation than shadowing low-frequency words (e.g., Goldinger, 1998), any imitation effects reported here are likely to generalize to other, lower-frequency words. Finally, the 16 words were selected to differ minimally from many other English words by the addition, subtraction, or substitution of a single phoneme. As reported by Scarborough (2013), words in dense phonological neighborhoods such as these are pronounced with greater coarticulatory nasality than words in sparser neighborhoods. Thus, we expected to observe a high degree of nasality in our stimuli. The hyper-nasalized stimuli used in this study are a subset of the hyper-nasalized stimuli shown to exhibit nasal imitation in Zellou et al., (2016).

Recordings of the model talker, a 22 year-old male, for stimulus generation were made using an Earthworks M30 microphone in a sound-attenuated booth. In addition to the 16 test words described above, the model talker produced the 16 words or nonwords required for the creation of a hyper-nasalized version to replace each of the naturally nasalized vowels. Finally, a set of 20 monosyllabic words with an oral coda consonant was recorded as well, to serve as distractor primes in a block of filler trials inserted between Blocks 1 and 2.

Hyper-nasalized tokens of the nasal words were created by additively combining the waveform of a naturally nasalized vowel (from a CVN word) with the waveform of naturally hyper-nasal vowel (from the NVN counterpart of the CVN word). For example, for the prime word *den*, the extracted vowel was combined with a vowel extracted from the nonword *nen* (which would be naturally more nasalized) and then spliced back into the original context, resulting in a hyper-nasalized *den*. The vowels' waveforms were first adjusted to match in amplitude and duration by shortening the long one to match the duration of the other. Using the PSOLA algorithm in the speech-editing software Praat, the pitch contour of the two vowels was modified to remain constant throughout the vowel at the same F0 value as the CVN word. With identical F0, the harmonic structure of vowels align in the frequency dimension, which made the additive combination of the vowels' spectra for each sample possible and causing the relative amplitudes for the oral and nasal peaks to be modified. The resulting vowel was further modified to display the same intensity and pitch contour as the original target vowel, and spliced back into the original target-word context (see Styler, Scarborough & Zellou, 2011, for more details). This method ensures that the hyper-nasalized vowel differed from the naturally nasalized one in spectral properties only.

Impressionistically, the hyper-nasalized stimuli sound quite natural and difficult to distinguish from the unaltered, natural stimuli the model speaker produced. In order to provide support of this impression, we subjected eight UC Davis undergraduate students (none of these subjects participated in the Experiments reported below), recruited through the psychology subject pool, to a simple two-alternative force choice task. Participants sat in a sound attenuated booth and heard both versions of the natural and hyper-nasalized stimuli over headphones, one after the other. Orderings of the stimulus type (natural first, or hyper-nasalized first) was randomized. After hearing a stimuli pair, listeners were asked to indicate which version of the word sounded "doctored or artificially manipulated." On average, these participants were 42% accurate at selecting the hyper-nasalized version as the manipulated item, with individual listener means ranging from 31%−60% accuracy. This corroborates our impression that the hyper-nasalized stimuli fell squarely within the range of pronunciations deemed 'natural', at least indistinguishably so from their unaltered, natural counterparts.

In order to confirm that the vowels of the hyper-nasalized stimuli were more nasalized than those of the naturally occurring stimuli, we collected Chen (1997)'s acoustic measure of nasality, i.e., the difference between the amplitude of the low frequency nasal peak, P0 (found around 250 Hz) whose amplitude increases with increased nasality, and the amplitude of the first formant peak, A1, whose amplitude decreases with increased nasality. These spectral elements are illustrated in Figure 1, which displays an oral vowel and its nasalized counterpart as produced by the model speaker. Since A1 decreases and P0 increases as nasalization increases, a smaller (or even negative) A1–P0 value signifies greater degree of nasality.

Figure 2 presents values of A1–P0 measured at three different points within the vowel (i.e., early, midpoint, late) averaged across the 16 vowels of the naturally nasal primes, the 16 vowels of the hyper-nasalized primes, and 20 vowels from oral words from the model talker. The graph reveals lower values of A1–P0 on nasal words than on oral words, as well as lower values of A1–P0 on hyper-nasalized vowels than on naturally-nasalized ones, but reliably so in the later portion of the vowels only. This pattern suggests that the normally nasalized and hyper-nasalized vowels differ in terms of the magnitude of the velum-lowering gestures, and less so in terms of the temporal characteristics of the gesture. Past research has reported variability on the temporal dimension (i.e., the velum lowers earlier in the vowel) and on the spatial one (i.e., the velum lowering gesture is of a greater magnitude) with changes in speaking rate and stress (Krakow, 1994). Cohn (1990) reports that in American English vowels in NVC and CVN contexts show the same temporal extent of contextual nasalization (i.e., the same cline of nasalization, though reversed), but differ in terms of *magnitude* of the velum gesture, as indicated by nasal airflow measures. Our own analysis is consistent with this finding: The hyper-nasalized vowel appears to have been produced with a velum-lowering gesture of a greater magnitude that the gesture involved in the production of the naturally nasalized vowel.

Data analysis

Using the speech-editing software Praat, we marked the vowels in the words that participants produced at baseline and as targets following spoken primes. The onset and offset of the vowels were taken to be the points at which an abrupt increase or reduction in amplitude of the higher formant frequencies in the spectrogram was observed. An abrupt change in amplitude in the waveform, along with simplification of waveform cycles, was used to verify these measurements.

All A1–P0 measurements were made on the segmented vowels, at the midpoint of each vowel automatically, using a Praat script. To minimize the risk of misidentification of the oral and nasal peaks, we verified that the frequency value associated with P0 conformed to the expected value of the first or second harmonic, given the gender and/or pitch characteristics of the who produced the vowel. Indeed, P0 tends to correspond to the first harmonic in individuals with a relatively high fundamental frequency (usually women) while P0 tends to correspond to the second harmonic in individuals with a lower fundamental frequency (Klatt & Klatt, 1990). The frequencies of P0 and F1 were also verified to ensure that they were appropriate for a given vowel phoneme. No vowel measurements were excluded.

Figure 3 presents the mean A1–P0 values for each of our participants, measured at the vowels' midpoints and averaged across all 16 nasal words produced at baseline, i.e., before hearing the model talker. For comparison, the model talker's natural nasality value is also displayed. As apparent in the figure, the model talker's vowels display greater nasality (i.e., smaller A1–P0 value) than any of the participants. Thus, imitating the model talker's nasal vowels required all our participants to increase the degree of nasality in their own vowels.

In addition to measuring acoustic nasality on target vowels, imitation was assessed by calculating, for each participant and each nasal word, the change in A1–P0 value (in dB) between the vowel produced on the target word following a prime and the vowel produced on the same word at baseline, before any of the model talker's utterances was heard. For example, if the value of A1–P0 on the vowel of the word "band" was 5 dB at baseline and 4 dB (thus becoming more nasal) on the token produced following a prime word, the dependent measure for that experimental trial was 1 dB (=5–4). Note that while a lower raw

A1–P0 dB value indicates greater nasality, a positive change value indicates imitation because, as pointed out above, the model speaker's vowels were characterized by greater nasality and therefore contain smaller A1–P0 values than participants' vowels.

Participants and Procedure

Fourteen University of Pennsylvania undergraduates participated in Experiment 1. All participants in the current studies were native speakers of American English and received course credit for their participation. None reported any visual or hearing impairment.

Before the main part of the study began, participants were instructed to read 16 nasal words (and 8 filler oral words), presented orthographically in the center of a computer screen, one at a time in random order. Responses were recorded digitally at a 44-kHz sampling rate. Once the recording was completed, the main part of the study began. Each trial consisted of the auditory presentation of a prime, followed by the visual presentation of the target word, with a 0.5 second delay between the end of auditory prime and the presentation of the printed word. Participants were asked to read out loud the word presented on the computer screen. The visual presentation of each word was always preceded by an auditory prime, presented over headphones. Participants were instructed simply to read the printed word.

Participants were presented with two separate blocks containing only the hyper-nasalized primes or the naturally nasalized primes, with an intervening block of filler oral words. Although the experimental software controlled block and trial organization, there was no pause between blocks. Each target word was presented four times within each block, with a different prime and target pairing on each repetition, with the constraint that prime and target never be the same word. Experimental block ordering, with either the hyper nasality block (7 participants) or the natural nasality block (7 participants) first, was counterbalanced across two groups of participants. One random order for all trials was created for each of the two list versions of the experiment.

There were a total of 128 experimental nasal trials collected from each participant (16 words × 4 repetitions × 2 stimuli types).

Results

Figure 4 presents the mean coarticulatory nasality at baseline and test conditions, as well as mean computed change in coarticulatory nasality (relative to baseline) produced by participants in their target-word productions in Blocks 1 and 2 as a function of which type of primes they heard in each block. In Block 1, participants who heard hyper-nasalized primes increased nasality of the target-word vowels more than those people who heard naturally nasalized primes. In fact, because the A1–P0 change associated with target vowels following naturally-nasalized primes is close to 0, people do not appear to have modified their vowel nasality in response to the talker's vowels. In Block 2, the change in nasality in the target-word vowels was quite different from that observed in Block 1. Indeed, participants who had heard naturally nasalized vowels in Block 1's primes produced less nasality after hyper-nasalized vowels than the group who heard these vowels in Block 1. Conversely, participants who had heard hyper-nasalized vowel primes in Block 1 produced greater nasality following the naturally nasalized vowel primes in Block 2 than the group who heard these naturally nasalized vowel primes in Block 1. This pattern is precisely what is predicted if people's nasality imitation of a prime in Block 2 is affected both by the prime's degree of nasality and that of the primes heard in Block 1.

To establish that the pattern observed is statistically robust, we modeled the amount of change in A1–P0 (in dB) from baseline to target production in a linear mixed effects model using the *lme4* package in *R*. For significance testing of critical fixed effects and interactions, we adopted a model-comparison approach. This approach consists of first constructing a model that captures the factors that we believe may affect the dependent variable (i.e., the change score in A1–P0 between baseline and at test on each trial) but are not central to our hypotheses. This fit of the so-called 'base' model to the data, i.e., the amount of the variance accounted for, is then used as a benchmark to which each augmented model is compared. Augmented models are models that include all of the base model's predictors but also a variable or factor whose contribution we wish to assess. By comparing a measure of the variance accounted for by a given augmented model to that of the base model (after correcting for the difference in degrees of freedom between the two), one can estimate whether the inclusion of the predictor in question in the augmented model results in a significantly improved fit to the data. The difference in goodness of fit between the base model and the augmented model corresponds to −2 times the change in log likelihood, which follows a χ^2 distribution with degrees of freedom equal to the number of parameter(s) that were added to the simpler model.

We built our base model so as to capture the fact that the overall value of change scores may vary across individuals and target words by including participants and words in the random structure of the model. In addition, the base model included the following two (fixed) predictors: whether people had heard the target as prime on preceding trials, and the phonetic similarity between prime and target. As stated above, the target word consisted of a different word from the immediately preceding prime. However, the same set of words was used for both primes and targets across trials. Thus, it is possible that participants' increased vowel nasality on a target word does not in fact result from a broad generalization of articulatory dynamics, as assumed here, but from direct exposure to the talker's pronunciation of this word as prime in a prior trial. To assess the possibility of such influence, we compared nasality (and degree of imitation) on target vowels in those trials where the target word had been heard as prime before within that Block to those where no such exposure had taken place yet (i.e., target words that had not yet been presented as primes within that Block). If generalization across words is limited, imitation of nasality on target vowels should be greater when participants had been exposed to the target word as prime (and could recruit the memory trace in the echo driving target production) than when no such trace was available. Pre-exposure to target was coded as a categorical variable that represented whether or not the target word had been encountered as prime on any preceding trial within the block (sum-coded with two levels, 'no' [reference level] and 'yes').

In addition to pre-exposure, the base model included a predictor that captures the phonetic similarity between prime and target for each trial, estimated by the number of phonemes the two words have in common in the same position. For example, the prime 'band' and target 'den' were estimated to have one phoneme in common (i.e., the 'n' in coda position), and ‘sand’ and ‘band’, three. Similarity between prime and target was coded in terms of the number of overlapping phonemes and treated as a continuous (and centered) variable. We hypothesized that imitation may increase with increased similarity between prime and target. Finally, the random structure of the base model included random variations that we wished to include in our augmented models because the model-comparison approach adopted here requires that the base and augmented models be nested in order to be valid. Thus, the random structure of the base model and of all of the augmented models described below included random intercepts and slopes per participant for the main effect of Block, and random intercepts and slopes per item for the main effects of Group and Block. (See below for further details about these factors. Models with more complex random structure failed to converge.)

Table 1 lists the effect-size estimate, standard error, and t-value associated with each fixed-effect predictor for the base model. For simplicity, a *t* value greater than 1.96 is taken to indicate that the predictor makes a significant contribution to accounting for the observed values of nasality change (Baayen et al., 2008). The intercept of the model is significant, indicating that the mean change in vowel nasality between baseline and test is greater than 0. Neither the similarity between prime and target, as captured by the number of phonemes in common, nor whether people had heard the target word as prime early in the experiment made a significant contribution to accounting for the nasality change score. (The sign of the estimate associated with the main effect of pre-exposure indicates that imitation is smaller where targets had been heard as prime before, which goes contra to prediction.) Thus, there is no evidence that imitation was enhanced by word-specific exposure.

We then computed a series of augmented models by adding to the base model the main effects of the factors Block (first or second block for a given trial, sum-coded) and Group (a subject-level variable defined by the type of prime heard in Block 1, i.e., naturally nasalized or hyper-nasalized primes, sum-coded), separately, and compared each augmented model’s fit to the data to the base model’s fit. Table 2 presents the log-likelihood associated with the base model, along with that of each of the two augmented models that included the main effect of Block or Group. There was not a significant contribution of Group as a main effect, reflecting the fact that nasality change was not greater across produced nasality in both blocks for people who heard hyper-nasalized primes in Block 1 than those who heard naturally-nasalized primes in Block 1.

In order to test our main hypothesis, resting on the interaction between Block and Group, we created an augmented model that included all main effects and compared it to a one including the Bock and Group interaction. As indicated in Table 2, the inclusion of the interaction between Block and Group significantly improved the fit to the observed data. This confirms Figure 4, where nasality imitation produced after hyper-nasalized primes is reduced when measured in Block 2 (after presentation of naturally nasalized primes) compared to Block 1, and nasality imitation following naturally-nasalized primes was greater when tested in Block 2 (after prior presentation hyper-nasalized primes) compared to Block 1.

Finally, and for the sake of completeness, we tested whether the inclusion of the interaction between Block and Target Pre-exposure to the model with all main effects improved the fit to the data. This interaction was considered important to test because pre-exposure was established within each block, as opposed to across the entire experiment because, at the beginning of Block 2, every target had been heard as prime before, i.e., in Block 1; the interaction aims to capture the possibility that the effect of pre-exposure would be apt at capturing influence on imitation better in Block 1, where no pre-exposure corresponds to no pre-exposure both with the block and within the experiment, than in Block 2, where no pre-exposure corresponds to no pre-exposure within the block only. As indicated in Table 2, the interaction did not improve the fit.

Two important points emerge from Experiment 1‘s results. First, participants produced target words with more nasalized vowels after exposure to a talker who heavily nasalizes his vowels, and this imitation took place across words, revealing people’s ability to generalize articulatory features. Second, with limited (and homogeneous) experience with the talker’s degree of nasality in vowels (i.e., in Block 1), imitation tracked the degree of nasality in prime vowels, with greater nasality produced after hyper-nasalized primes than after naturally nasalized primes. However, this pattern changed significantly after more varied experience with the talker, suggesting that immediate imitation is driven by a target that incorporates more than just the specific properties of the prime. The target of imitation, we claim, includes both the prime and the memory traces of the model talker’s vowels. Thus, the target of imitation recruited following a hyper-nasalized prime in Block 2 is populated by past experience with the naturally nasalized vowels from Block 1’s primes, causing target vowels to be less nasalized than those produced in Block 1, where only hyper-nasalized vowels had been presented. Likewise, the target driving imitation after a naturally nasalized prime in Block 2 includes the hyper-nasalized vowels from Block 1’s primes, yielding target vowels with greater nasality than what was observed in Block 1, where only naturally nasalized vowels had been heard.

An alternative to this interpretation of the data, however, can be offered. The observed changes between Blocks 1 and 2 may in fact reflect a reduction of the influence of the primes on the production of target-word vowels over time. The priming effect would not disappear entirely but its magnitude would be reduced, yielded the apparent reduction in the size of the Group effect in Block 2, compared to Block 1. Note that this account is more descriptive than explanatory because it does not readily explain why, with reduced influence of primes over time, people produced more nasality than their baseline following naturally-nasalized primes in Block 2 (and more than the degree of nasality produced after the same primes in Block 1). It remains the case that Experiment 1‘s design does not allow us to know how imitation changes over time independently of a change in nasality magnitude across the blocks.

We thus conducted a control experiment in which two new groups of participants were tested on the same material under the same conditions. The only difference with Experiment 1‘s design consisted of presenting the same primes in Block 2 as those in Block 1. Any changes between Blocks would reflect an effect of adaptation or repetition in the course of the study.

Experiment 2

Method

Participants were 14 University of California at Davis undergraduate students. The procedure and stimuli were identical to that of Experiment 1 except for the following: The prime stimuli presented on Block 2 remained the same as those presented on Block 1. Half of the participants (N=7) heard only naturally-nasalized primes, the other half (N=7), only hyper-nasalized primes.

Results

Figure 5 presents the mean coarticulatory nasality at baseline and test conditions and mean computed change in nasality (relative to baseline) produced by participants exposed to either naturally nasalized or hyper-nasalized primes in both the first and second experimental blocks. In Block 1, participants demonstrate similar responses to naturally nasalized and hyper-nasalized primes as in Experiment 1: participants produced greater nasality after hyper-nasalized primes than after naturally nasalized primes. In Block 2, both groups increased nasality in target-word vowels after hearing the same primes as in Block 1. Thus, there is no evidence that people imitated the primes less in the course of the study. In fact, the data suggest that people’s imitation grows stronger across Blocks. This finding, in turn, gives some support to the claim that, as people accumulate experiences with a given talker, any imitation following an utterance from that talker reflects an aggregate of those experiences.

In order to establish that prime-nasality imitation in Experiment 1 and in Experiment 2 was different whether the primes remained the same between Blocks 1 and 2 or changed, we fitted the data from both experiments to a baseline linear mixed effects model, which we gradually expanded via model-comparisons to assess the contribution of specific factors to the models’ fit. Our goal was to evaluate the effects of Group (defined by the type of prime heard in Block 1, i.e., naturally nasalized or hyper-nasalized primes, sum-coded), Block (first or second, sum-coded), and Experiment Type (Alternating [base level] and Repeating, sum-coded) in accounting for the amount of change in A1–P0 (in dB) from baseline to target production. First, following the same procedure as that reported above for just Experiment 1, we generated a base model that included Target pre-exposure and Prime-Target Similarity (detail about these variables are described in Results for Experiment 1). Table 3 lists the effect-size estimate, standard error, and t-value associated with each fixed-effect predictor for the base model. The intercept of the model is positive, indicating that participants overall did increase their produced nasality across both experiments. Again, neither similarity between prime and target and whether people had heard the target word as prime early in the experiment significantly improve the model’s fit to the data.

We then computed a series of augmented models by adding to the base model the main effects of the factors Group, Block, and Experiment Type separately. Each augmented model model’s fit to the data was compared to the base model’s fit. A model with all main effects was used as the comparison for models with each of the two-way interactions. Finally, a model with all main effects and all two-way interactions was compared to a model augmented with the three-way interaction between Group, Block, and Experiment Type. The random structure of those models was identical to the random effects of the original models used to assess the data from Experiment 1.

Table 4 provides the log-likelihood associated with the base model, along with that of each of the three augmented models that included the main effect of Group, Block, or Experiment Type. The addition of the effect of Group improved the model's fit, as found in Experiment 1: Participants increased the nasality of target-word vowels more after hyper-nasalized primes than after naturally-nasalized primes.

We then created an augmented model that included all main effects and compared it to three separate ones including the two-way interactions between Group and Block, Block and Experiment Type, and Group and Experiment Type. A significant two-way interaction between Block and Experiment significantly improved the model's fit: There was an increase in change of nasality from Block 2 to Block 1 in the repeating experiment that was not observed in the alternating stimuli experiment. Finally, to test whether there was an interaction between Group, Block and Experiment Type, we created an augmented model that included all of these two-way interactions and compared it to one that included a three-way interaction. Critically, the inclusion of the three-way interaction of Group, Block, and Experiment Type also significantly improved the model's fit: Compared to what was observed in Block 1, nasality imitation in Block 2 increased for the two groups who heard the same primes on both blocks; for those groups who heard different primes on Block 2, nasality increased or decreased depending on which primes were heard in Block 1.

Our claim regarding the recruitment of a target of imitation that consists of an aggregate of the talker's utterances was tested in Experiment 3. There, we manipulated the temporal delay between the presentation of the prime and the pronunciation of the target word. As reviewed earlier, the composition of the target of imitation/echo is believed to evolve over time as more memory traces or past episodes with the talker are recruited and the proportional contribution of the immediately preceding prime decreases. If the changes in vowel nasality from Block 1 to Block 2 observed in Experiment 1 are attributable to the inclusion of the overall experience with the model talker, we reasoned, expanding the delay between prime and target should enhance this influence: In Block 2 (where the influence of the aggregate target of imitation can be distinguished from that of the prime), nasality after hyper-nasalized prime should decrease more, and nasality after naturally-nasalized prime, increase more with longer delay.

Experiment 3

Given Experiment 1's suggestion that people respond to a prime by recruiting their overall experience with the talker's vowels, we sought converging evidence for such recruitment by examining how vowel nasality on target words changes with greater delay between prime and target. We hypothesized that the local influence of the prime would be reduced, and that of past experiences with the talker's nasal vowels increased, with greater temporal distance between prime and target. To this end, we replicated Experiment 1 but varied the delay between the presentation of the auditory prime and that of the printed target word. The effect of delay on imitation of each group should be modest in Block 1 (for which the prime's vowel and previously experienced vowels have the same degree of nasality) but substantial in Block 2, where we should see a stronger effect of accumulated experience with the model talker's vowels as delay between target and prime increases.

Method

Fifteen University of Pennsylvania undergraduates participated in Experiment 3. The procedure and stimuli were identical to that of Experiment 1 except for the following: In each block, each of the 16 nasal target words was presented three times, each repetition associated with a different delay between the prime and the target (0.5 second, 1.5 seconds, and 5.5 seconds). Which type of prime preceded the target was varied across blocks, and the delay with which each prime was presented was varied across trials within a block. Experimental block ordering, with either hyper-nasalized (N=8) or naturally-nasalized (N=7) block first, was counterbalanced across two groups of participants. One random order for all trials was created for each of the two list versions of the experiment. Experiment 3 consisted of a total of 96 experimental trials (16 words × 3 delay levels × 2 stimuli types).

Results

Figure 6 presents the mean coarticulatory nasality at baseline and test productions, as well as mean change in A1–P0 value at test (compared to their value at baseline) in words produced by participants who heard either naturally-nasalized or hyper-nasalized primes in Block 1, separately for each block, as a function of the delay between prime and target. At short (0.5 sec) delay (Fig 6a. and b.), people produced vowels with greater nasality (compared to baseline) after hyper-nasalized primes than after naturally nasalized primes in Block 1 but, as observed in Experiment 1, the influence of the prime was tempered in Block 2: People who had heard the hyper-nasalized primes in Block 1 produced more nasality after naturally nasalized primes than the people who had heard them on Block 1; conversely, people who heard the naturally nasalized primes in Block 1 produced less nasality after hyper-nasalized primes than people who had heard hyper-nasalized primes in Block 1. With a slightly longer delay between primes and targets (1.5 sec, Fig 6c. and d.), a very similar pattern emerged. However, with a substantially longer delay (5.5 sec, Fig 6e. and f.), the influence of group on nasality imitation in Block 2 was significantly weakened or even gone: Participants who had heard hyper-nasalized primes in Block 1 produced only a little less nasality after naturally nasalized primes in Block 2; more strikingly perhaps, people who had heard naturally nasalized primes in Block 1 produced even less nasality after hyper-nasalized primes in Block 2.

As before, we conducted model comparisons to assess the contribution of the effects of Group, Block, and Delay in accounting for the amount of change in A1–P0 (in dB) from baseline to target production. Again, we first generated a base model that included Target pre-exposure and Prime-Target Similarity. Table 5 presents a summary of the base model. The intercept failed to reach significance, indicating that overall and all conditions averaged, participants' target vowels were not more nasalized than their baseline vowels. Neither similarity between prime and target and whether people had heard the target word as prime early in the experiment made a significant contribution to fitting the data.

Again, model comparisons assessed the significance of critical factors and interactions. The main effects of Group, Block, and Delay were added separately to the base model, and we compared each augmented model's fit to the data to the base model's fit. The random structure of every model fitted to these data included random intercepts and slopes per participant for each of the predictors Block and Delay, as well as random intercepts per item. (Models with more complete random structures failed to converge.)

Next, a model including all main effects was generated and this model was used to assess the contribution of each two-way interaction. Finally, our critical hypothesis in this Experiment was that in the second block, we should see a stronger effect of accumulated experience with the model talker's vowels as the delay between prime and target increases. In order to assess this critical prediction, a model including all main effects and all two-way interactions was compared to the same model augmented with the critical three-way interaction between Group, Block, and Delay. Table 6 reports the outcome of these model comparisons. Group significantly improved the model's fit, as well as the interaction between Group and Delay. Crucially, inclusion of the three-way interaction of Group, Block, and Delay significantly improved the model's fit: Confirming what we observe in Figures 6e. and f., after the longest delay, nasality increased for those who heard naturally nasalized primes and decreased for those who heard hyper-nasalized primes in Block 2, relative to productions after shorter delays, signaling that there is a more robust effect of past experiences with the model talker's speech as delay between prime and target increases.

These results extend Experiment 1's findings in important ways: Nasality change was more pronounced after hearing hyper-nasalized primes than after natural primes in Block 1 (when experience with the talker's vowels was homogeneous), irrespective of the delay between prime and target. However, the prime's specific influence on target vowel production diminished to reveal the influence of the overall experience with the talker's vowels in Block 2, and this influence became even more pronounced as the delay between prime and target increased.

The opposite effect of delay on nasality produced after each kind of primes is significant: Delay did not cause participants' nasality production to return to baseline, as it has often been claimed (e.g., Goldinger, 1998). If it did, the amount of nasality would have decreased after both prime types. Instead, nasality increased after naturally-nasalized primes. Thus, we contend, the echo/target driving imitation incorporates talker-specific episodes. Their inclusion in the target of imitation is evident when the delay between prime and target is short, and their influence increases as the delay is extended.

General Discussion—As reviewed in the introduction, there is good evidence that people imitate the pronunciation of a word they are asked to repeat. Whether people can alter their pronunciation of a word they haven't heard to resemble that of a model talker, however, is unclear. The absence of cross-word generalization has been viewed as a consequence of the way memory traces and episodes are recruited and contribute to word production (Goldinger, 1998): The segmental make-up of the word to be uttered is viewed as the central constraint on the composition of the target of imitation that drives pronunciation. Yet, naturalistic cases of imitation—such as adults' adoption of aspects of the pronunciation of their ambient language or dialect (e.g., Sancier & Fowler, 1997)–suggest that generalization can take place. Here, we focused on the dynamics and amplitude of the velum-lowering gesture that accompanies the production of a nasal consonant, and asked if people can extract this feature from specific utterances and integrate it into their own production of different utterances.

To this end, we used an altered version of the shadowing task, one in which the target word people produce is a different word from the prime they hear immediately prior. We examined whether people altered their pronunciation of the target word's vowel immediately after hearing the pronunciation of a different word with a high degree of coarticulatory vowel nasality. Two versions of the nasal primes were presented, a naturally-nasalized version from a talker who produces substantial coarticulatory nasality, and one which we generated to have an even greater degree of nasality ('hyper-nasalized'). We varied which prime people heard, whether people's experience with the model speaker's utterances was limited to that kind of prime or also included the other kind as well, and how much time elapsed between the prime presentation and the target production.

We found that people can spontaneously modify coarticulatory vowel nasality to resemble that demonstrated in the speech they are exposed to. This imitation is directly affected by the characteristics of the prime just heard: Participants increased nasality in their target vowels after hearing hyper-nasalized primes more than they did after hearing naturally-nasalized primes. Thus, people's pronunciation is under the influence of some recent experience. However, we also observed that the influence of the prime was modulated by the nature of past experience with the talker's nasalized vowels: Nasality following a naturally-nasalized prime was greater when people had also heard hyper-nasalized vowels, compared to that observed when only naturally-nasalized vowels had been heard. Conversely, nasality following a hyper-nasalized prime was reduced when people had also heard naturally-nasalized vowels, compared to when they had only heard hyper-nasalized vowels. Finally, the influence of that past experience grew even larger, and that of the local prime, smaller, with longer delay between the prime and target production. Taken together, these findings support a view of imitation driven by a dynamically assembled representational target. This target of imitation is influenced by both properties of the immediately heard prime and by experience with a talker's utterances, the latter having a stronger influence with increased delay.

Evidence for the involvement of the overall experience with the talker's utterances in the target of imitation even when imitation immediately follows the presentation of a prime is consistent with the effect of lexical frequency on immediate shadowing reported by Goldinger (1998). Perhaps because of the complexity of mapping an auditory signal onto articulatory program, in addition to the other layers of complexity such as social factors that are involved in a given context (cf. Pardo, et al., 2010; Babel, 2012), imitation may always rely on a rich and complex target. The present finding demonstrates that talker identity can be one of the dimensions that influence the composition of the target of imitation.

The effect of prime-target delay on nasality imitation observed here is also particularly noteworthy. Past literature has assumed, explicitly or implicitly, that the fading of the specific influence of word instances over time results in a return to people's baseline articulation. What we observed here is quite different. People's delayed imitation of a model talker appears to be based on a target of imitation tracking what is most general about the speaker, and less what was specific to the particular instance they just heard. In our view, this is because the effect of accumulated experience with the talker becomes more robust as delay increases, causing the proportional contribution of a specific prime on the target of imitation to decrease.

Voices are highly salient signals to people. Indeed, people pay attention to and encode the many aspects of the voice they hear, even in situations where these aspects are seemingly irrelevant to the task they are engaged in, as when hearing isolated words produced by a disembodied voice through headphones in a research lab (e.g., Palmeri, Goldinger, & Pisoni, 1993, Goldinger, 1996). Furthermore, people often imitate the voice of their interlocutor spontaneously. This propensity leads to some convergence in manner of speaking between people engaged in a conversation (Capella, 1981; Giles, Coupland, & Coupland, 1991). Such imitation undoubtedly plays a role in marking a form of social affiliation as people match their linguistic systems to each other. Imitation that has been observed previously across specific words and experiences with a given talker must rely on an abstract representation of the talker's speech (e.g., Pardo, 2006; Nielsen, 2011). The present study examined how this representation emerges and evolves with experience with a talker across words and time. People do indeed generalize coarticulatory properties of a heard word and apply them onto similar words. But the target of imitation includes more than just the specific properties of the heard word; it incorporates overall experience with the talker's utterances, and the influence of that experience plays a greater role in explaining people's imitation as time passes and experience with that talker accumulates. These results add to a growing body of work on the perceptual, cognitive, and social constraints on imitation.

Acknowledgements

We thank Sam Beer for help with the stimulus recordings. We are grateful for the detailed comments provided by three reviewers and by editor Bob McMurray.

Appendix:: List of Nasal words used as primes and targets in the study.

band	den
bound	grain
found	clam
pant	lime
tent	rhyme
lend	rum
rung	pun
dime	sand

References

Baayen RH, Davidson DJ, & Bates DM (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of memory and language, 59(4), 390–412. doi:10.1016/j.jml.2007.12.005

Brysbaert M, & New B (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior research methods, 41(4), 977–990. doi:10.3758/BRM.41.4.977 [PubMed: 19897807]

Cappella JN (1981). Mutual influence in expressive behavior: Adult–adult and infant–adult dyadic interaction. Psychological Bulletin, 89(1), 101. doi:10.1037/0033-2909.89.1.101 [PubMed: 7232607]

Chen MY (1997). Acoustic correlates of English and French nasalized vowels. The Journal of the Acoustical Society of America, 102(4), 2360–2370. doi:10.1121/1.419620 [PubMed: 9348695]

Cohn AC (1990). Phonetic and phonological rules of nasalization. Phonetics Laboratory, Department of Linguistics, UCLA.

Evans BG, & Iverson P (2007). Plasticity in vowel perception and production: A study of accent change in young adults. The Journal of the Acoustical Society of America, 121(6), 3814–3826. doi:10.1121/1.2722209 [PubMed: 17552729]

Giles H, Coupland J, & Coupland N (Eds.). (1991). Contexts of accommodation: Developments in applied sociolinguistics. Cambridge University Press.

Goldinger SD (1996). Words and voices: episodic traces in spoken word identification and recognition memory. Journal of experimental psychology: Learning, memory, and cognition, 22(5), 1166. doi:10.1037/0278-7393.22.5.1166

Goldinger SD (1998). Echoes of echoes? An episodic theory of lexical access. Psychological review, 105(2), 251. doi:10.1037/0033-295X.105.2.251 [PubMed: 9577239]

Goldinger SD, & Azuma T (2004). Episodic memory reflected in printed word naming. Psychonomic Bulletin & Review, 11(4), 716–722. doi:10.3758/BF03196625 [PubMed: 15581123]

Hintzman DL (1984). MINERVA 2: A simulation model of human memory. Behavior Research Methods, 16(2), 96–101. doi:10.3758/BF03202365

Hintzman DL (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological review, 95(4), 528. doi:10.1037/0033-295X.95.4.528

Klatt DH, & Klatt LC (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. the Journal of the Acoustical Society of America, 87(2), 820–857. doi:10.1121/1.398894 [PubMed: 2137837]

Krakow RA (1994). Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. Haskins Laboratory Status Report on Speech Research, 31–48.

Liberman AM, Cooper FS, Shankweiler DP, & Studdert-Kennedy M (1967). Perception of the speech code. Psychological review, 74(6), 431. doi:10.1037/h0020279 [PubMed: 4170865]

Meltzoff AN, & Moore MK (1997). Explaining facial imitation: A theoretical model. Early Development & Parenting, 6(3–4), 179. doi:10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R [PubMed: 24634574]

Namy LL, Nygaard LC, & Sauerteig D (2002). Gender differences in vocal accommodation: The role of perception. Journal of Language and Social Psychology, 21(4), 422–432.

Nielsen K (2011). Specificity and abstractness of VOT imitation. Journal of Phonetics, 39(2), 132–142. doi:10.1016/j.wocn.2010.12.007

Palmeri TJ, Goldinger SD, & Pisoni DB (1993). Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(2), 309. doi:10.1037/0278-7393.19.2.309

Pardo JS (2006). On phonetic convergence during conversational interaction. The Journal of the Acoustical Society of America, 119(4), 2382–2393. doi:10.1121/1.2178720 [PubMed: 16642851]

Pardo J (2013, 6). Reconciling diverse findings in studies of phonetic convergence. In Proceedings of Meetings on Acoustics (Vol. 19, No. 1, p. 060140). Acoustical Society of America. doi:10.1121/1.4798479

Pardo JS, Jay IC, & Krauss RM (2010). Conversational role influences speech imitation. Attention, Perception, & Psychophysics, 72(8), 2254–2264. doi:10.3758/BF03196699

Podesva RJ, Hilton K, Moon K, & Szakay A (2013). Nasality as non-aggressiveness: An articulatory sociophonetic study New Ways of Analyzing Variation (NWAV) 42, Pittsburgh, PA.

Pycha A (2016). Co-articulatory cues for communication: An investigation of five environments. Language and Speech, 59, 364–386. doi:10.1177/0023830915599183 [PubMed: 29924531]

Sancier ML, & Fowler CA (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. Journal of Phonetics, 25(4), 421–436. doi:10.1006/jpho.1997.0051

Scarborough R (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. Journal of Phonetics, 41(6), 491–508. doi:10.1016/j.wocn.2013.09.004

Scarborough R, & Zellou G (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. The Journal of the Acoustical Society of America, 134(5), 3793–3807. doi:10.1121/1.4824120 [PubMed: 24180789]

Shockley K, Sabadini L, & Fowler CA (2004). Imitation in shadowing words. Perception & Psychophysics, 66(3), 422–429. doi:10.3758/BF03194890 [PubMed: 15283067]

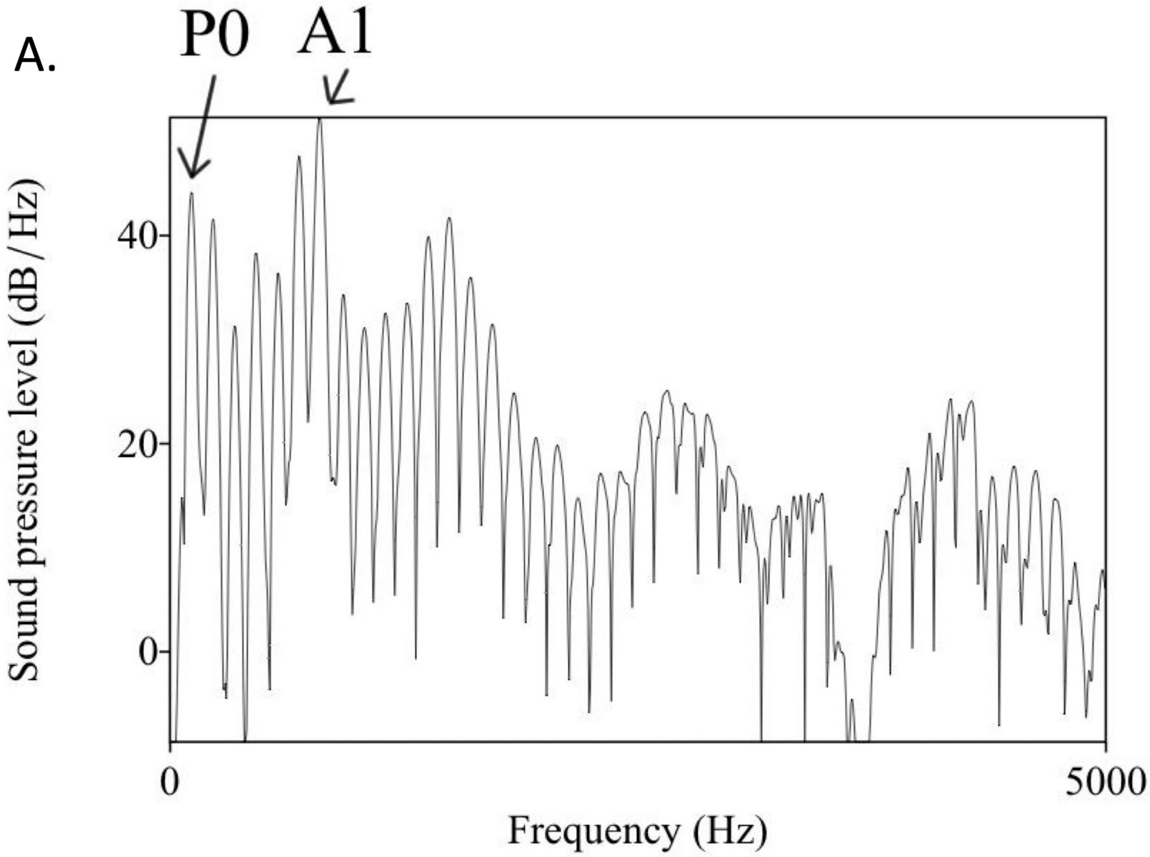
Styler WF, Scarborough R, & Zellou G (2011). Use of waveform mixing to synthesize a continuum of vowel nasality in natural speech. The Journal of the Acoustical Society of America, 130(4), 2443–2443. doi:10.1121/1.3654800

Wright JT (1986). The behavior of nasalized vowels in the perceptual vowel space. Experimental phonology, 45–67.

Zellou G, Scarborough R, & Nielsen K (2016). Phonetic imitation of coarticulatory vowel nasalization. Journal of the Acoustical Society of America, 140, 3560–3575. doi:10.1121/1.4966232 [PubMed: 27908038]

Zellou G, & Scarborough R (2015). Lexically conditioned phonetic variation in motherese: age-of-acquisition and other word-specific factors in infant- and adult-directed speech. Laboratory Phonology, 6(3–4), 305–336. doi: 10.1515/lp-2015-0010

Zellou G, & Tamminga M (2014). Nasal coarticulation changes over time in Philadelphia English. Journal of Phonetics, 47, 18–35. doi:10.1016/j.wocn.2014.09.002


A.
P0
A1
Sound pressure level (dB/Hz)
40
20
0
0
5000
Frequency (Hz)

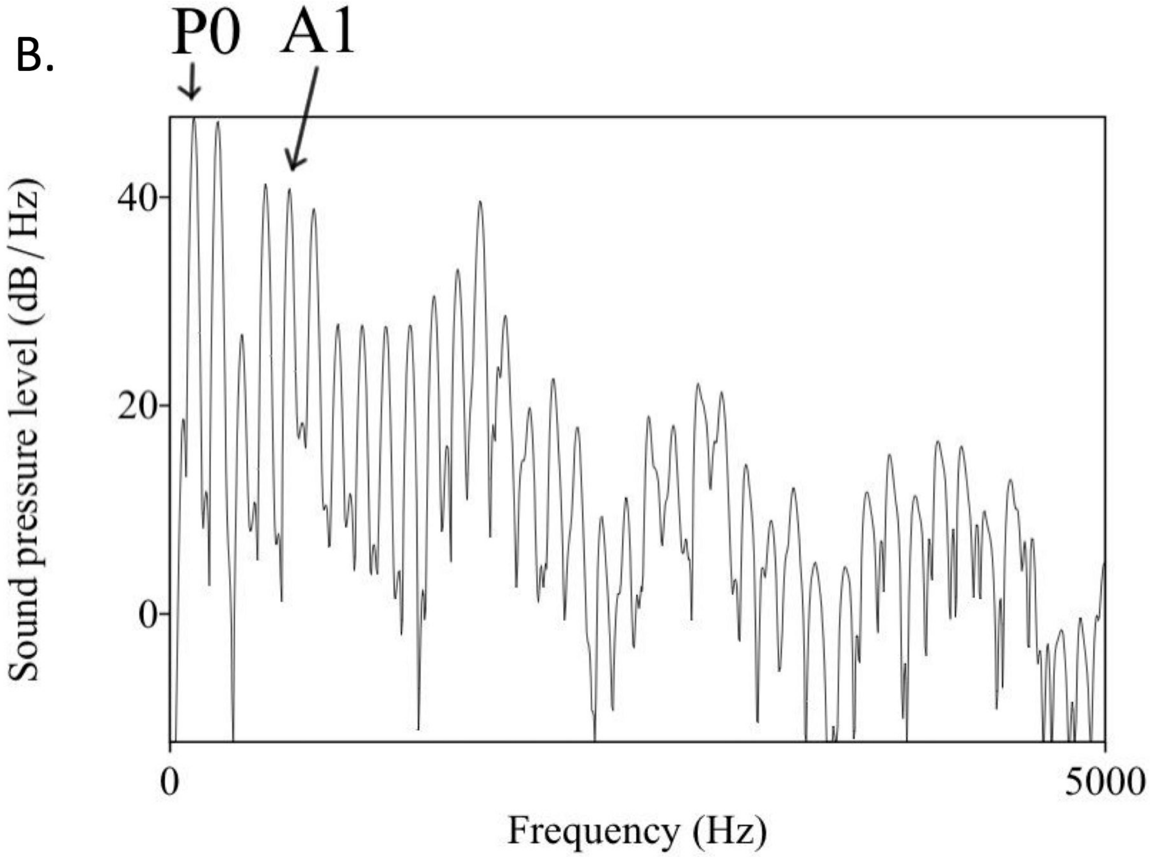


Figure 1:

Spectra from the midpoint of (A) an oral vowel (from the word “bad”) and (B) a nasalized vowel (from “band”) from the model speaker in this study.

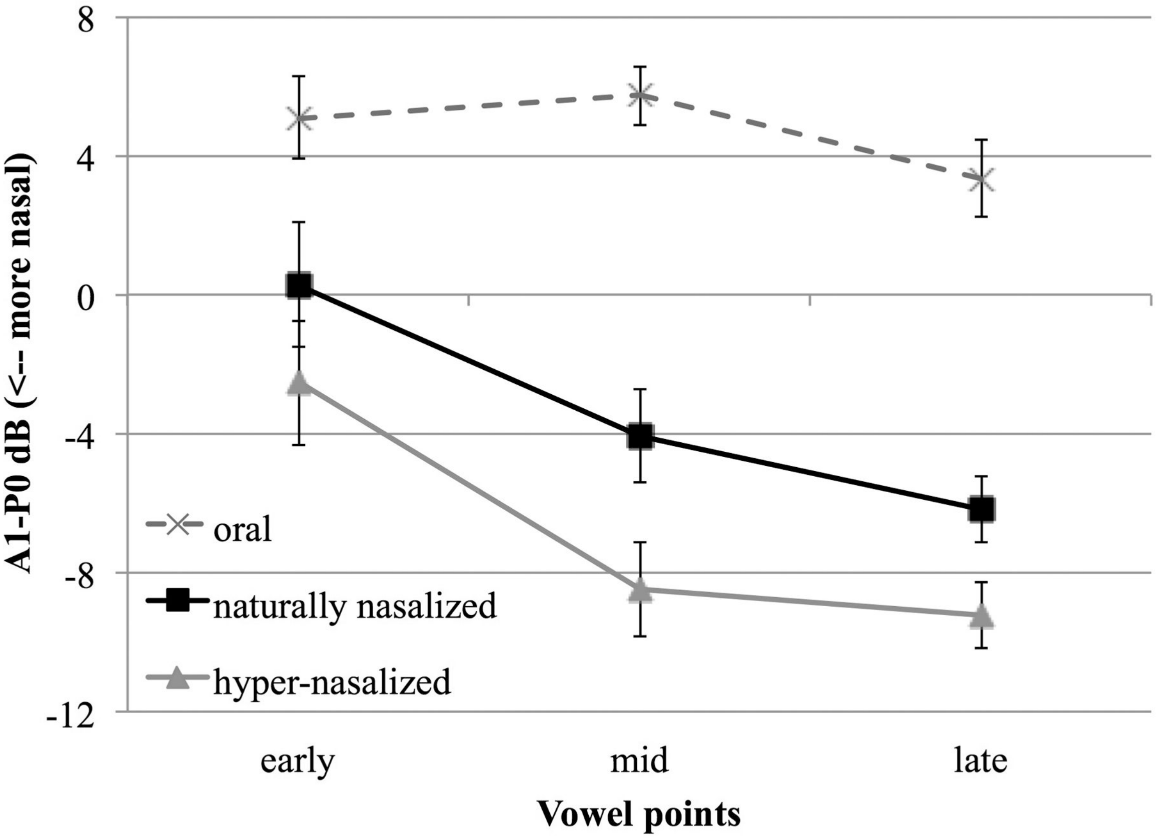


Figure 2:

Acoustic nasality (A1–P0, in dB) means and standard errors of the natural and hyper-nasalized stimuli used in the study, as well as oral vowels from this talker, taken at early, mid and late points in the vowel.

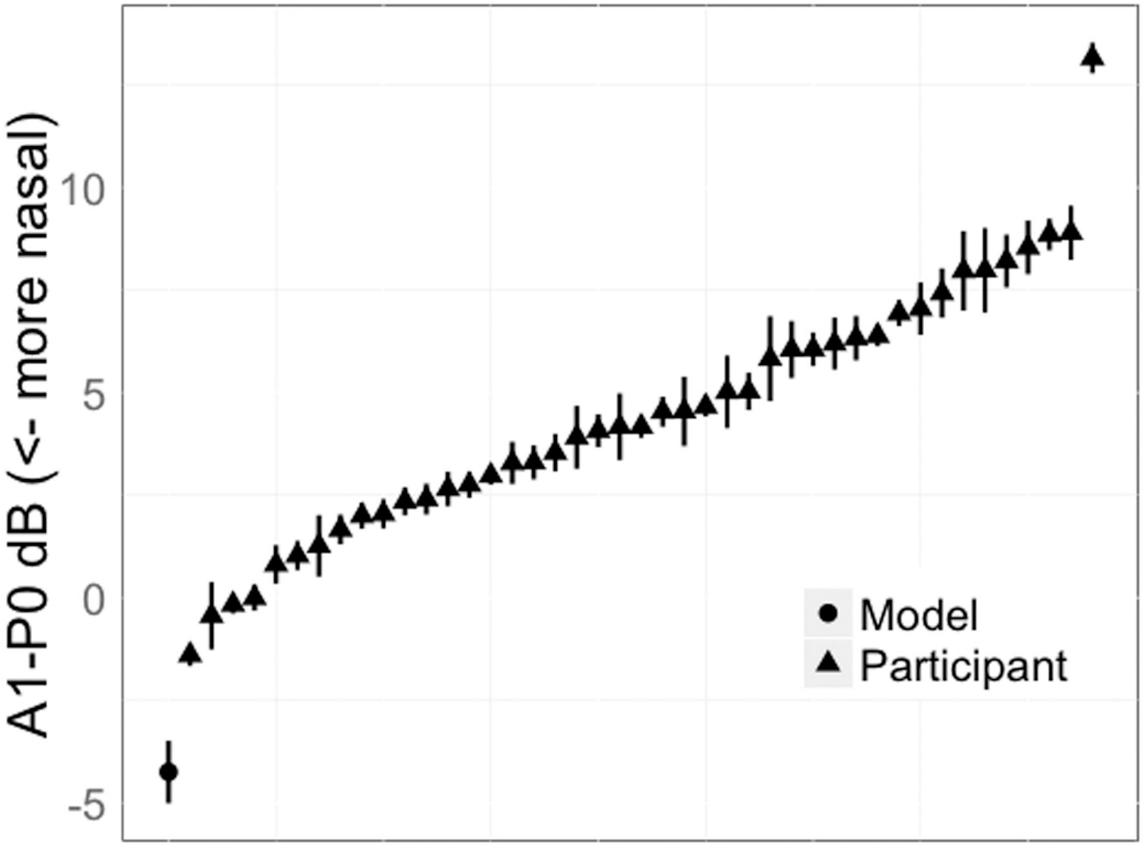


Figure 3:

A1–P0 value (in dB) measured at vowel's midpoint and averaged across 16 nasal words for the model talker (circle) and each of the 43 participants in the study at baseline (triangles), with standard errors.

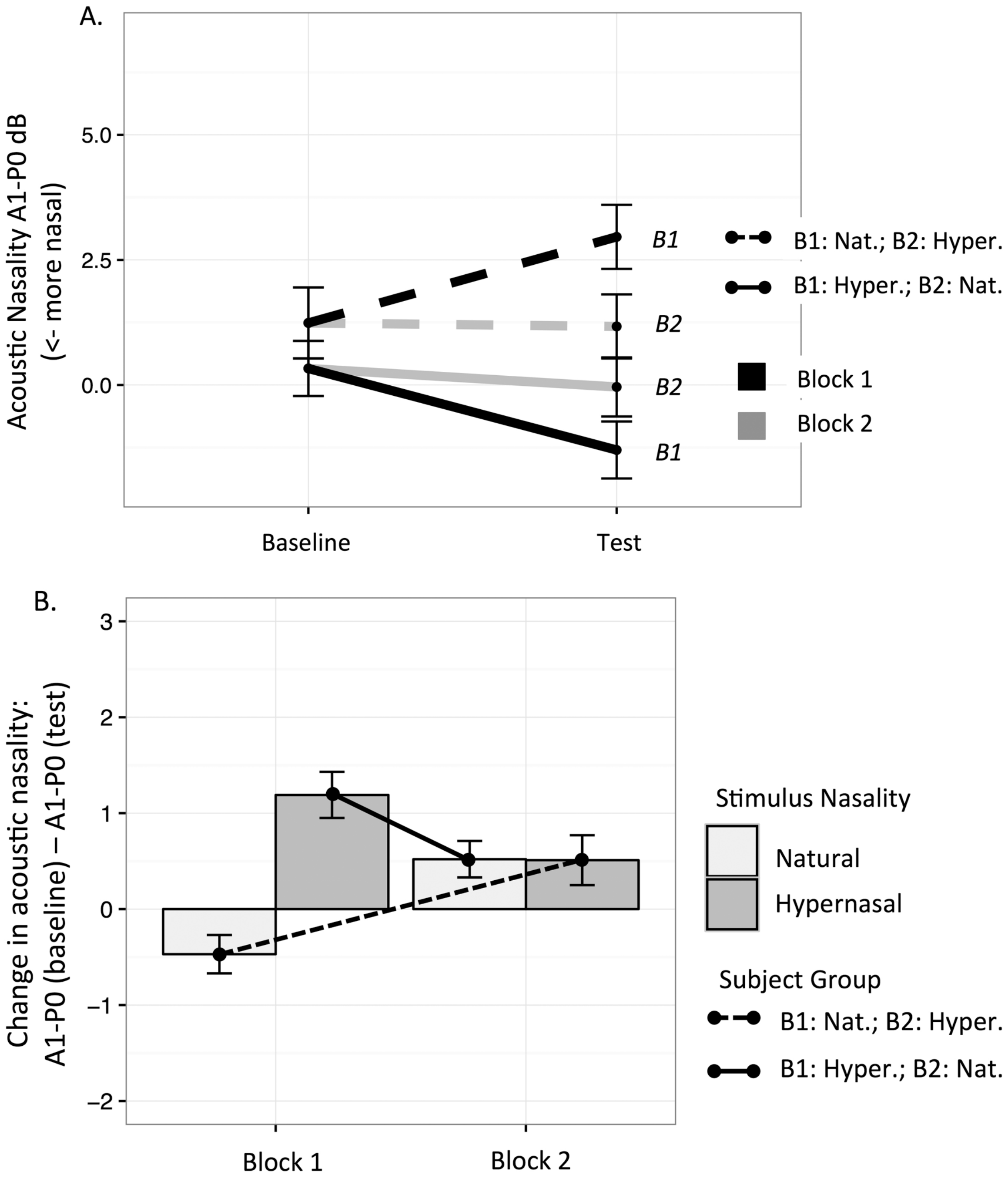


Figure 4:

Experiment 1: (A) Acoustic nasality (as A1–P0 dB) in baseline and test conditions and (B) Difference between the amplitude of A1–P0 (in dB) measured in the vowels of the target productions and that measured in the vowels of the same words at baseline as a function of participant group (group who heard naturally nasalized primes in Block 1 and hyper-nasalized primes in Block 2 or group who heard hyper-nasalized primes in Block 1 and naturally nasalized primes in Block 2) and the experimental block (1 or 2).

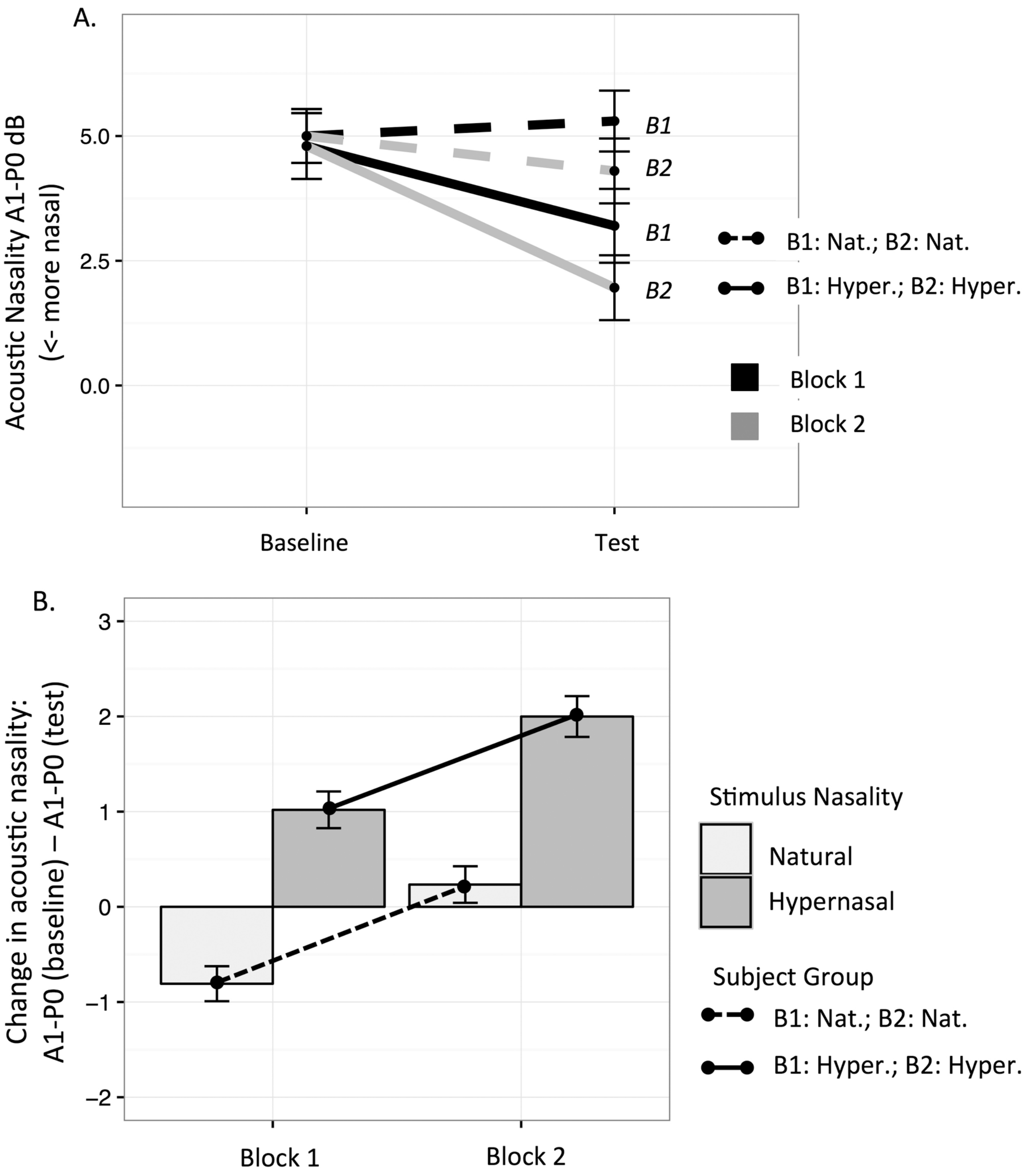
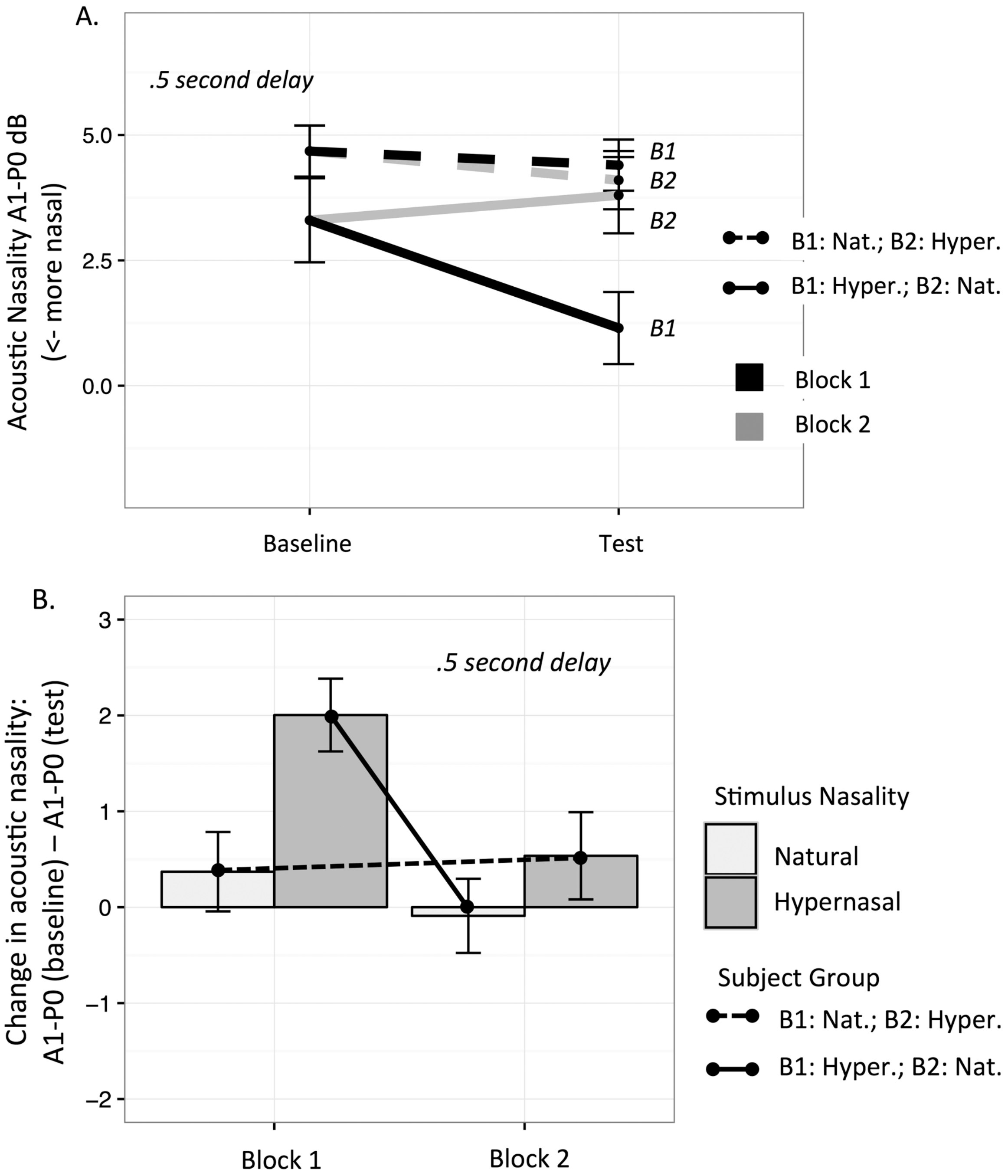
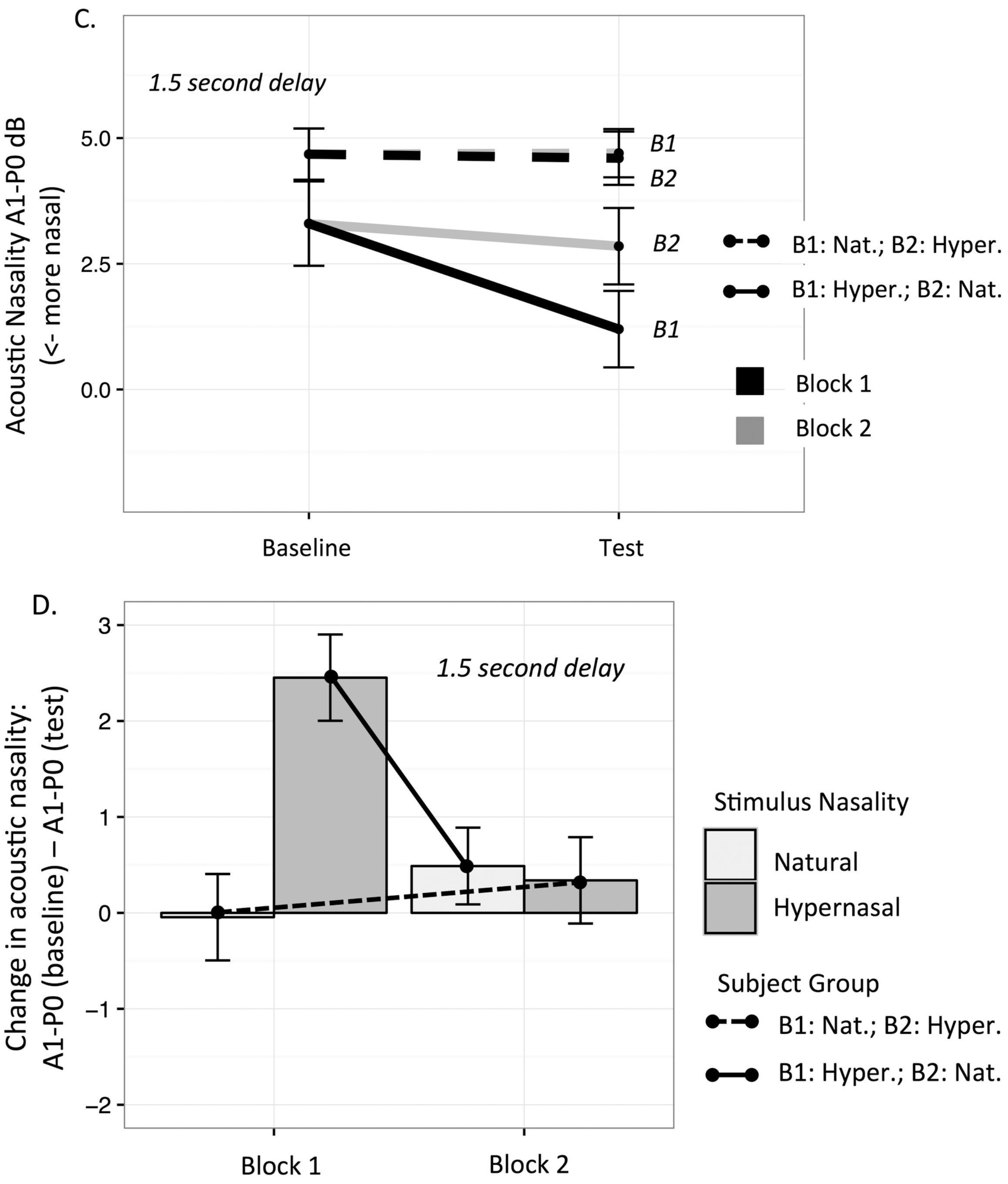


Figure 5:

Experiment 2: (A) Acoustic nasality (as A1–P0 dB) in baseline and test conditions and (B) Difference between the amplitude of A1–P0 (in dB) measured in the vowels of the target productions and that measured in the vowels of the same words at baseline as a function of participant group (group who heard only naturally nasalized primes or group who heard only hyper-nasalized primes) and the experimental block (1 or 2).


A.
.5 second delay
Acoustic Nasality A1-P0 dB
(<- more nasal)
5.0
2.5
0.0
B1
B2
B2
B1
Baseline
Test
B1: Nat.; B2: Hyper.
B1: Hyper.; B2: Nat.
Block 1
Block 2
B.
.5 second delay
Change in acoustic nasality:
A1-P0 (baseline) – A1-P0 (test)
3
2
1
0
−1
−2
Block 1
Block 2
Stimulus Nasality
Natural
Hypernasal
Subject Group
B1: Nat.; B2: Hyper.
B1: Hyper.; B2: Nat.


C.
1.5 second delay
Acoustic Nasality A1-P0 dB
(<- more nasal)
5.0
2.5
0.0
B1
B2
B2
B1
Baseline
Test
B1: Nat.; B2: Hyper.
B1: Hyper.; B2: Nat.
Block 1
Block 2
D.
1.5 second delay
Change in acoustic nasality:
A1-P0 (baseline) – A1-P0 (test)
3
2
1
0
−1
−2
Block 1
Block 2
Stimulus Nasality
Natural
Hypernasal
Subject Group
B1: Nat.; B2: Hyper.
B1: Hyper.; B2: Nat.

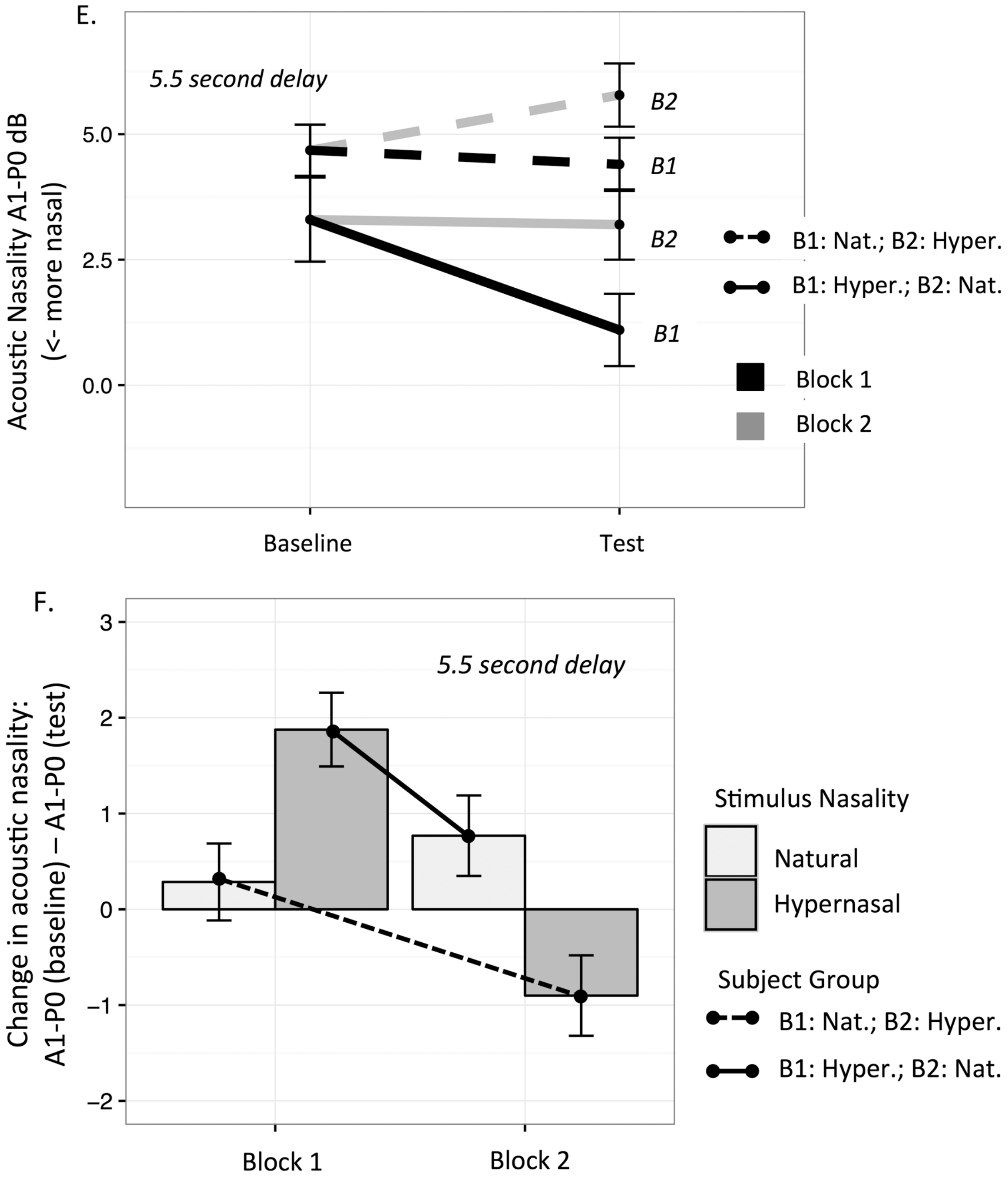


Figure 6:

Experiment 3: Acoustic nasality (as A1–P0 dB) in baseline and test conditions and difference between the amplitude of A1–P0 (in dB) measured in the vowels of the target productions and that measured in the vowels of the same words at baseline as a function of participant group (group who heard naturally nasalized primes in Block 1 and hyper-nasalized primes in Block 2 or group who heard hyper-nasalized primes in Block 1 and naturally nasalized primes in Block 2) and Block (1 or 2) for .5s delay between prime and target word presentation (A., A1–P0, and B., change), 1.5s delay (C., A1–P0, and D., change), and 5.5s delay (E., A1–P0, and F., change).

Table 1:

Experiment 1: Fixed effect estimate, standard error, and t-value, for each fixed effect of the base model (see text for details). (For each categorical variable, the level used as reference is indicated in parenthesis; the sign of the effect indicates how going from the reference level to the other level changes the dependent variable.)

	Est.	**Std. Error**	**t-value**
(Intercept)	.82	.48	2.1
Similarity	−0.001	0.17	−.01
Target Exposure (Yes)	−0.36	0.19	−1.9

Table 2:

Experiment 1: Log likelihood ratio, chi squared statistic, and p-value for each model comparison.

Model	**Log Likelihood**	$\chi^2(1)$	$Pr(>\chi^2)$
Base	−9209.1		
Main Effects			
Group	−9209	.27	n.s.
Block	−9209	.29	n.s.
Model with all main effects	−9208.9		
Two-way interactions			
Group:Block	−8876.6	664.5	p<.001
Target Pre-exposure:Block	−9214.4	0	n.s.

Table 3:

Experiments 1 & 2: Fixed effect estimate, standard error, and t-value, for each fixed effect of the base model (see text for details). (For each categorical variable, the baseline level is indicated in parenthesis.)

	Est.	**Std. Error**	**t-value**
(Intercept)	.61	0.36	1.9
Similarity	−.02	0.09	−.28
Target Exposure (Yes)	−0.05	0.11	−.5

Table 4:

Experiments 1 & 2: Log likelihood ratio, chi squared statistic, and p-value for each model comparison.

Model	**Log Likelihood**	**$\chi^2(1)$**	**Pr($>\chi^2$)**
Base	−13822		
Main Effects			
Group	−13817	8.1	p<.01
Block	−13820	3.7	p=.05
Experiment Type	−13821	.32	n.s.
Model with all main effects	−13815		
Two-way interactions			
Group:Block	−13815	.04	n.s.
Block:Experiment Type	−13814	3.1	p=.07
Group: Experiment Type	−13814	2.07	n.s.
Model with all two-way interactions	−13814		
Three-way interaction			
Group: Experiment Type: Block	−13425	777.9	p<.001

Table 5:

Experiment 3: Fixed effect estimate, standard error, and t-value, for each of the fixed effect of the base model (see text for details). (The baseline level of the categorical predictor is indicated in parentheses.)

	Est.	Std. Error	t-value
(Intercept)	0.64	0.54	1.19
Similarity	0.09	0.26	0.35
Target Exposure (Yes)	0.15	0.12	1.22

Table 6:

Experiment 3: Log likelihood ratio, chi squared statistics, and p-value for each model comparison.

Model	**Log Likelihood**	**$\chi^2(1)$**	**$Pr(>\chi^2)$**
Base	−7050		
Main Effects			
Group	−7045.8	6.7	p<.01
Block	−7048.6	2.6	n.s.
Delay	−7048.6	1.1	n.s.
Model with all main effects	−7044.8		
Two-way interactions			
Group:Block	−7053.9	1.9	n.s.
Group: Delay	−7050	9.6	p<.05
Block:Delay	−7053	.04	n.s.
Target pre-exposure: Block	−7053	.24	n.s.
Model with all simple effects and two-way interactions	−7050		
Three-way interaction			
Group:Block:Delay	−7040	18.2	p<.01