Journal of Phonetics 61 (2017) 13–29

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

Individual differences in the production of nasal coarticulation and perceptual compensation

Georgia Zellou

Linguistics Dept., University of California, Davis, 469 Kerr Hall, One Shields Avenue, Davis, CA 95616, United States

A R T I C L E I N F O

Article history:
Received 30 March 2016
Received in revised form
18 November 2016
Accepted 6 December 2016

Keywords:
Individual differences
Production–perception link
Perceptual compensation
Coarticulation

A B S T R A C T

The current study investigates correlations between individual differences in the production of nasal coarticulation and patterns of perceptual compensation in American English. A production study (Experiment 1) assessed participants' nasal coarticulation repertoires by eliciting productions of CVC, CVN and NVN words. Stimuli for two perception tasks were created by cross-splicing oral vowels (from C_C words), nasal vowels (from C_N words), and hypernasal vowels (from N_N words) into C_C, C_N, and N_N consonant contexts. Stimuli pairs were presented to listeners in a paired discrimination task (Experiment 2), where similarity of vowels was assessed, and a nasality ratings task (Experiment 3), where relative nasalization of vowels was judged. In the discrimination task, individual differences in produced nasal coarticulation predicted patterns of veridical acoustic perception. Individuals who produce less extensive anticipatory nasal coarticulation exhibit more veridical acoustic perception (indicating less compensation for coarticulation) than individuals who produce greater coarticulatory nasality. However, in the ratings task, listeners' produced nasal coarticulation did not predict perceptual patterns. Rather, more veridical perceptual response patterns were observed across participants in context-inappropriate coarticulatory conditions, i.e., for hypernasal vowels in C_N contexts (e.g. *bε̃̃n*) and nasal vowels in N_N contexts (e.g. *mε̃n*). The results of this study suggest a complex and multifaceted relationship between representations used to produce and perceive speech.

1. Introduction

Producing and perceiving speech are distinct processes: they employ separate parts of human anatomy and involve different mechanical actions. Yet, at the cognitive level, production and perception must utilize (at least some of) the same representational structures in order for communication to be successful. Every person is physiologically different and each individual acquires the phonetic patterns of their speech community using their own idiosyncratic articulatory characteristics. This process of negotiating idiosyncratic physiological constraints to achieve native acoustic-phonetic targets can possibly influence the mental representations shared with production and perception and, in turn, shape how each listener experiences incoming speech. Currently, we test this possibility by examining how idiosyncratic patterns of produced nasal coarticulation in individuals predict their patterns of perceiving nasal coarticulation on vowels in various contexts. We explore how these results speak to the nature of speech representations and the production-perception relationship.

Speaking produces context-dependent sounds, the result of which is a blended acoustic output. Perceptual compensation is the process of attributing the acoustic features of temporally overlapping gestures to the appropriate sound. Compensation for coarticulation allows the coalesced acoustic signal to be parsed into context-independent units. For instance, a nasalized vowel adjacent to a nasal consonant does not sound nasalized to a listener, indicating that the acoustic effect of coarticulation is attributed to the influence of the adjacent segment. Meanwhile, a nasalized vowel in isolation is heard as nasal since there is no context to which the nasalization can be attributed (Kawasaki, 1986). There are competing explanations for perceptual compensation, stemming from different theoretical views and empirical observations about the nature of speech perception and set up testable predictions about the relationship between the production and perception of speech.

One set of explanations relies on the assumption that the goal of speech perception is to retrieve the gestural sources of the acoustic signal. Motor Theory (MT) suggests that the listener parses speech to retrieve the "intended phonetic gestures of the speaker" (Liberman & Mattingly, 1985: 2). Compensation for coarticulation is viewed as evidence of this process. For example, listeners' phoneme categorizations along a continuum shift depending on the spectral properties of an adjacent

E-mail address: gzellou@ucdavis.edu

0095-4470/$ - see front matter © 2016 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2016.12.002

segment: more /da/ responses following an /r/ for stimuli along a d-g continuum is the result of attributing a lower F3 to the preceding rhotic (Mann, 1980). MT interprets such effects as reflecting the listener's ability to recover the articulatory sources of the heard acoustic signal using her innate knowledge about the acoustic consequences of vocal tract dynamics and gestural production. The MT stance is that this knowledge is not learned, rather a result of the specialized function of the linguistic system and biologically-based, which does not fully align with our assumption that individual differences in speech behavior are shaped through linguistic experience. A useful premise from MT, however, is that speech production and perception are closely linked, since the representations used to determine “the speaker's production is the distal object [i.e., the intended gesture] that that listener perceives; accordingly, speaking and listening are both regulated by the same structural constraints and the same grammar” (Liberman & Mattingly, 1985: 30).

MT does not make an explicit prediction that the gestural-grammatical system of an individual can vary. However, we extend and revise the notion of a production-perception link in an explicit and substantive way: If the goal of speech perception is to uncover underlying gestures, one way in which this might be achieved is by recruiting one's own idiosyncratic production targets. We suggest that the cognitive structures governing speaking and listening are learned, shaped by an individual's linguistic experience, and idiosyncratic in that they can vary across individuals (which is a departure from the original MT stance that the perception–production link is purely biologically based (Liberman & Mattingly, 1985)). Therefore, our *individual differences prediction* is that differences in degree of compensation (more or less) should be observed as a function of a listener's own production targets (reflecting representational structures for coarticulation). If listeners use their own phonetic knowledge of coarticulation and its acoustic consequences to parse the speech signal, we predict that individuals who produce greater coarticulatory vowel nasalization should likewise expect a larger overlapping velum gesture in an utterance and, subsequently, exhibit perceptual patterns indicating more perceptual compensation than individuals who produce less coarticulation.

Direct Realist theory (DR, Fowler, 1986) employs principles similar to MT in that it considers that the objects of speech perception are gestural. However, DR relies on the notion that this gestural knowledge is garnered from the speech signal directly and not mediated by a separate vocal tract-based module. DR predicts that all listeners attribute all of the coarticulatory effects in speech to a consonantal source, as long as it is present in the signal (Fowler, 2006). This hypothesis accounts for findings that compensation for coarticulation can occur even for speech sounds that are not represented in the grammar of an individual (i.e., compensation for non-native phoneme patterns (Mann, 1986)). Therefore, a DR approach would not predict that compensation is linked to a speaker's idiosyncratic gestural representations.

A third variation of a gestural account for speech perception proposes that language-specific gestural patterns are encoded in the grammatical structure for native speakers (e.g., Beddor & Krakow, 1999; Beddor, Harnsberger, & Lindemann, 2002). Under this perspective, cross-linguistic differences in perceptual compensation indicate that language-specific phonetic patterns shape listeners' expectations for how much acoustic variation is attributed to a source consonant. Beddor and Krakow (1999) investigated how differences in language experience influence patterns of perceptual compensation. Listeners with different native language backgrounds, either native English or native Thai speakers, were presented two types of American English vowels, oral (from CVC words) and heavily nasalized (from NVN words) presented in both C_C and N_N contexts. Contextual nasalization in Thai is far less extensive than in English; And in critical trials, where participants heard heavily nasalized vowels in appropriate N_N contexts, Thai listeners displayed less compensation for coarticulation than English listeners (only in a ratings task, where explicit nasality judgments were elicited, but not in a vowel discrimination task). In other words, Thai speakers displayed more veridical acoustic perception of vowel nasalization in coarticulatory contexts. The authors surmised that Thai listeners' experience with smaller degrees of contextual nasalization in their native language led them to expect, and thus, compensate for, only a small amount of coarticulation. This suggests that listeners have expectations about the extent to which particular segments should overlap based on their language-specific phonetic representations and these expectations shape how their perceptual system attributes acoustic information in the speech signal to underlying gestures.

Hence, an additional possibility that we explore in the current study is that native-language coarticulatory experience guides perceptual compensation, suggesting that language-internal variations might lead to different patterns of compensation in different contexts for listeners from one speech community. For instance, within English there are different coarticulatory patterns depending on whether a vowel is adjacent to one or two nasal consonants: vowels preceded by a nasal coda only (CVN) have less nasalization than vowels flanked by both an onset and a coda nasal consonant (NVN) (Cohn, 1990). Our *context-dependent prediction* holds that if compensation depends on learned community-level phonetic patterns and varies based on the knowledge that all listeners have about context-dependent articulatory structures in their language, we predict more veridical acoustic perception overall in cases where degree of nasalization deviates from what is expected from the consonantal context, e.g., a vowel from NVN in a C_N context.

Such context-dependent variations are indeed part of the perceptual experience an English listener accrues in her native language. Therefore, language-internal context-dependent patterns are a source of potential variation in listener expectations about the talker's intended gestures. An alternative interpretation of the Beddor and Krakow (1999) finding that aligns with our individual differences prediction is that English and Thai listeners differ not only in their linguistic experience, but also, critically, in their own personal representations for coarticulation. Since groups of English speakers and Thai speakers have production targets that specify different amounts of coarticulatory overlap, a production-perception link would predict distinct patterns of perceptual compensation. The current study aims to tease apart the individual differences prediction and the context-dependent prediction.

In contrast to gestural accounts, auditory-theoretic views holds that speech perception proceeds following processes of general auditory perception and spectral contract detection (Diehl, Lotto & Holt, 2004). For example, hearing a nasal