Cynthia G. Clopper and Rory Turnbull

2 Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors

Abstract: Substantial empirical research has revealed that temporal and spectral phonetic vowel reduction occurs in "easy" processing contexts relative to "hard" processing contexts, including effects of lexical frequency, lexical neighborhood density, semantic predictability, discourse mention, and speaking style. Theoretical accounts of this phonetic reduction process include listener-oriented approaches, in which the reduction reflects the talker's balancing the comprehension needs of the listener with production effort constraints, talker-oriented approaches, in which reduction is argued to result entirely from constraints on speech production processes, and evolutionary approaches, in which reduction results directly from long-term interactive communication within a community. Recent research in our laboratory has revealed complex interactions among the linguistic, social, and cognitive factors involved in phonetic vowel reduction processes. These interactions reveal variation in the robustness of phonetic reduction effects across linguistic factors, as well as different patterns of interactions among linguistic, social, and cognitive factors across acoustic domains. These interactions challenge aspects of each of the three existing models of phonetic reduction. We therefore propose that a more complex view of the relationship between processing demands and phonetic vowel reduction processes is necessary to account for these observed patterns of variation.

Keywords: lexical frequency, neighborhood density, contextual predictability, speaking style, regional dialect

2.1 Introduction

Phonetic reduction is one of many processes contributing to variation in the acoustic-phonetic realization of speech. We define phonetic reduction as the phenomenon in which linguistic units (e.g., segments, syllables, or words) are realized with relatively less acoustic-phonetic substance (e.g., shorter duration and/or less extreme articulation) in a given context relative to other contexts. We assume that phonetic reduction involves acoustic-phonetic variation in realized

Cynthia G. Clopper, Ohio State University
Rory Turnbull, University of Hawai'i at Mānoa

https://doi.org/10.1515/9783110524178-002

segments along a continuum from hypoarticulated or reduced to hyperarticulated or enhanced. We therefore consider phonetic reduction to reflect a reduced degree of acoustic-phonetic substance in comparison to more hyperarticulated or enhanced forms (Johnson, Flemming, and Wright 1993). This variation in the degree of acoustic-phonetic substance is assessed using measures of segment and word duration, vowel space expansion, and f0, among others.

We limit our discussion in this chapter primarily to phonetic variation along measurable acoustic dimensions and therefore do not include categorical reduction processes, such as segmental alternations (e.g., full vowels alternating with schwa or stop consonants alternating with flap) or the deletion of segments, syllables, or words (cf. Ernestus 2014; Johnson 2004; Schuppler et al. 2011). We similarly focus on lexical and contextual factors that have been described in the literature as contributing to phonetic reduction as we have defined it here and therefore do not include phonetic reflexes of phonological properties such as segmental context (cf. Klatt 1976; Luce and Charles-Luce 1985; Peterson and Lehiste 1960), lexical stress (cf. de Jong 1995, 2004; Fourakis 1991; van Bergem 1993), or prosodic structure (cf. Lehiste 1971; Wightman et al. 1992). This division between continuous, phonetic reduction and categorical, phonological processes provides us with a more clearly circumscribed focus of discussion in this chapter, but it most likely does not reflect a true, natural division in language processing.[1] We therefore expect our conclusions to extend to categorical reduction processes (see also Cohen Priva 2015) and encourage more work that examines the intersection of prosodic structure and the lexical and contextual factors we discuss in this chapter (see, e.g., Baker and Bradlow 2009; Burdin and Clopper 2015; Turnbull et al. 2015; Watson, Arnold, and Tanenhaus 2008).

We further focus in this chapter primarily on phonetic vowel reduction, which involves both temporal (i.e., duration) and spectral (i.e., vowel space peripherality) dimensions, although we also discuss some preliminary findings in the domain of prosodic (i.e., f0 and timing) reduction. The phenomena we discuss are not unique to vowels, however, and we expect our conclusions to be applicable to phonetic reduction in other domains, including consonantal phenomena (see, e.g., Baese-Berk and Goldrick 2009; Bouavichith and Davidson 2013; Goldrick, Vaughn, and Murphy 2013; Warner and Tucker 2011), coarticulatory phenomena (see, e.g., Lin, Beddor, and Coetzee 2014; Scarborough 2013), and other dimensions of prosodic structure in which duration, vowel quality, and f0 play a critical role (see, e.g., Arnold, Kahn, and Pancani 2012; Calhoun 2010a, 2010b; Watson, Arnold, and Tanenhaus 2008).

1 For discussion of the essentially arbitrary division between categorical and continuous aspects of phonetic and phonological structure, see Ladd (2011) and Munson et al. (2010).

2.2 Phonetic reduction in "easy" contexts

The unifying observation in previous work on phonetic reduction processes is that linguistic forms are reduced in "easy" contexts relative to "hard" contexts, where easy and hard are defined with respect to the assumed processing demands imposed by the context on the talker and/or the listener. The linguistic factors that have been shown to contribute to phonetic reduction include lexical properties (e.g., lexical frequency and neighborhood density), contextual properties (e.g., semantic predictability and discourse mention), and speaking style. The definitions of easy and hard contexts for each of these factors are summarized in Table 2.1.

For each of these factors, phonetic reduction is observed in the "easy" contexts relative to the "hard" contexts, although the identification of easy versus hard contexts differs across factors. For the lexical factors, easy and hard contexts are typically defined with respect to the processing demands of the listener and, although the lexical factors are themselves continuous, "easy" and "hard" contexts are typically treated categorically (e.g., Luce and Pisoni 1998; Munson and Solomon 2004; Wright 2004; cf. Baese-Berk and Goldrick 2009; Gahl, Yao, and Johnson 2012). Speaking style is an explicitly listener-oriented manipulation and is also typically defined categorically (Picheny, Durlach, and Braida 1985, 1986). In contrast, for the contextual factors, easy and hard contexts are more often defined with respect to the processing demands of the talker, which are often treated continuously as a reflection of continuous measures of predictability or accessibility (e.g., Bard et al. 2000; Bell et al. 2009; Kahn and Arnold 2012, 2015; cf. Aylett and Turk 2004). Thus, the observed relationship between processing ease and phonetic reduction has been defined in different ways and has been argued to result from a number of different processing mechanisms.

Lexical frequency is typically defined as the number of occurrences of a target word per million words in a corpus of written or spoken language. Early research on speech intelligibility revealed that high-frequency words are easier for listeners to identify than low-frequency words (Broadbent 1967; Howes 1957).

Table 2.1: Linguistic factors contributing to phonetic reduction.

Factor	"Easy"/Reduced	"Hard"/Unreduced
Lexical frequency	High frequency	Low frequency
Neighborhood density	Low density	High density
Semantic predictability	More predictable	Less predictable
Discourse mention	Second mention/given	First mention/new
Speaking style	Plain	Clear

High-frequency words are also produced more quickly than low-frequency words, suggesting an effect of lexical frequency on lexical access and/or motor planning in production (Balota and Chumbley 1985). Thus, high-frequency words exhibit fewer processing demands than low-frequency words for both talkers and listeners. Phonetic reduction is also observed for high-frequency words relative to low-frequency words. This effect of lexical frequency on phonetic reduction has been observed in both the temporal domain for words and vowels (Arnon and Cohen Priva 2013; Aylett and Turk 2004; Bell et al. 2009; Gahl, Yao, and Johnson 2012; Munson and Solomon 2004; Myers and Li 2009; Pate and Goldwater 2011; Pluymaekers, Ernestus, and Baayen 2005b) and the spectral domain for vowels (Munson 2007; Munson and Solomon 2004), and in both isolated word production (Munson 2007; Munson and Solomon 2004; Myers and Li 2009) and continuous speech production (Arnon and Cohen Priva 2013; Aylett and Turk 2004; Bell et al. 2009; Gahl, Yao, and Johnson 2012; Pate and Goldwater 2011; Pluymaekers, Ernestus, and Baayen 2005b).

Lexical neighborhood density is a measure of phonological similarity across words in the lexicon and is typically defined as the number of words that differ from a target word by one phoneme insertion, deletion, or substitution (Luce and Pisoni 1998). Competition during lexical access among phonologically similar words leads to more difficult perceptual identification of words with many phonological neighbors (i.e., high neighborhood density) than for words with few phonological neighbors (i.e., low neighborhood density; Luce and Pisoni 1998; Vitevitch and Luce 1998, 1999). However, the activation of multiple similar word forms leads to faster and less error-prone production for high-density words than low-density words (Vitevitch 2002).[2] Thus, high-density words are difficult to perceive and easy to produce, whereas low-density words are easy to perceive and hard to produce. Consistent with the processing demands exhibited for neighborhood density in perception, phonetic vowel reduction is typically observed in low-density words relative to high-density words in read speech. This effect of neighborhood density on phonetic reduction has been observed for read speech in both the temporal domain for stop consonants (Fox, Reilly, and Blumstein 2015; Peramunage et al. 2011) and the spectral domain for vowels (Clopper and Tamati 2014; Munson 2013; Munson and Solomon 2004), but not in the temporal domain for vowels (Munson and Solomon 2004). Further, in conversational speech, the effect of neighborhood density may be more consistent with the processing demands exhibited for neighborhood density in production: temporal

2 High neighborhood density also facilitates perceptual processing in tasks involving nonwords (Vitevitch and Luce 1998, 1999).

and spectral vowel reduction is observed for high-density words relative to low-density words (Gahl, Yao, and Johnson 2012).

Several composite measures of neighborhood density and lexical frequency have also been developed to provide a single metric to account for the combined effects of these two lexical factors on phonetic reduction. Similar to the results with the simple measures, these composite measures reveal phonetic vowel reduction in the temporal and spectral domains for high-frequency words with few, low-frequency neighbors (i.e., "easy words") relative to low-frequency words with many, high-frequency neighbors (i.e., "hard words"; Munson and Solomon 2004; Scarborough 2010, 2013; Wright 2004). Thus, both individually and in combination, the two lexical factors consistently predict greater phonetic reduction for easy words relative to hard words.

Turning to the contextual factors, semantic predictability captures a range of phenomena related to the syntactic, semantic, and nonlinguistic context that a target word is produced in. Words that are predictable given the preceding sentence context are more intelligible when presented in context than less predictable words (Kalikow, Stevens, and Elliott 1977; Miller and Isard 1963). Similarly, in production, predictable words are less likely to be preceded by a hesitation indicating disfluency than less predictable words (Beattie and Butterworth 1979). Thus, predictable words exhibit fewer processing demands than less predictable words for both talkers and listeners. Words that are predictable in their context also exhibit phonetic reduction relative to words that are less predictable in their context. This effect of semantic predictability on phonetic reduction has been observed in the temporal domain for words and vowels (Aylett and Turk 2006; Bell et al. 2009; Clopper and Pierrehumbert 2008; Engelhardt and Ferreira 2014; Gahl and Garnsey 2004; Hunnicutt 1987; Jurafsky et al. 2001; Lieberman 1963; Moore-Cantwell 2013; Pate and Goldwater 2011; Pluymaekers, Ernestus, and Baayen 2005a; Tily and Kuperman 2012), the spectral domain for vowels (Aylett and Turk 2006; Clopper and Pierrehumbert 2008; Jurafsky et al. 2001), and the prosodic domain for words (Kaland, Swerts, and Krahmer 2013; Wagner and Klassen 2015; Watson, Arnold, and Tanenhaus 2008). These effects of semantic predictability are consistent across a range of measures of predictability, including syllable *n*-gram conditional probabilities (Aylett and Turk 2006), lexical *n*-gram conditional probabilities (Bell et al. 2009; Jurafsky et al. 2001; Pate and Goldwater 2011; Pluymaekers, Ernestus, and Baayen 2005a; Tily and Kuperman 2012), syntactic structure probabilities (Gahl and Garnsey 2004; Moore-Cantwell 2013), cloze probabilities (Clopper and Pierrehumbert 2008; Hunnicutt 1987; Lieberman 1963), information structure (Wagner and Klassen 2015), and nonlinguistic contextual information (Engelhardt and Ferreira 2014; Kaland, Swerts, and Krahmer 2013; Watson, Arnold, and Tanenhaus 2008).

Discourse mention is a contextual factor that captures whether the target word is new or old in the context. Repeated words have real-world referents that are already in the common ground of the conversation and are therefore expected to be easier to access for the talker and the listener than new words that introduce new real-world referents (Chafe 1974; Fowler and Housum 1987). Consistent with this hypothesis, phonetic reduction is observed for easier, second mentions of target words than for harder, first mentions of the same word within a given discourse context.[3] This second mention reduction has been observed primarily in the temporal domain for words and vowels in both read speech (Baker and Bradlow 2009; Fowler 1988) and spontaneous speech (Bard et al. 2000; Fowler and Housum 1987; Galati and Brennan 2010; Kahn and Arnold 2012, 2015; Kaiser, Li, and Holsinger 2011; Lam and Watson 2010, 2014; Pate and Goldwater 2011; Sasisekaran and Munson 2012; Shields and Balota 1991). Thus, for both contextual factors, phonetic reduction is observed for easy (i.e., predictable or given) words relative to hard (i.e., less predictable or new) words.

The final linguistic factor, speaking style, refers to the adoption of a particular mode of speaking that is appropriate for the discourse context and the interlocutors. Style can be explicitly manipulated by the talker and is therefore a potentially different type of linguistic factor contributing to phonetic reduction than the lexical and contextual factors discussed above, which are assumed to be largely implicit. In the context of phonetic reduction research, the primary speaking styles that have been investigated are plain lab speech, which is directed toward an imagined friend, and clear lab speech, which is directed toward an imagined hearing-impaired or nonnative listener.[4] Clear lab speech is more intelligible than plain lab speech (Picheny, Durlach, and Braida 1985), reflecting the talker's explicit adoption of a style that is appropriate for a listener who is assumed to exhibit speech processing difficulties. That is, although clear lab speech is easier to perceive than plain lab speech, it is produced in a context in

3 Second mention reduction may also be linked to other aspects of the communicative domain: Hoetjes et al. (2015) observed that co-speech gesturing that accompanies second mentions tends to be reduced in magnitude relative to gesturing which accompanies first mentions. Similarly, Hoetjes et al. (2012) documented second mention reduction effects in Dutch Sign Language.

4 Speaking style is also a focus of a substantial body of work in variationist sociolinguistics (e.g., Eckert and Rickford 2001) and is therefore related to our discussion below of the effect of social factors, including dialect variation, on phonetic reduction. However, we limit our discussion here to clear and plain lab speech styles because this stylistic variation involves a similar continuum of reduced and enhanced speech as the other linguistic factors described in this section.

which perceptual processing is assumed to be difficult, given the characteristics of the listener. Thus, plain lab speech exhibits phonetic reduction relative to clear lab speech, consistent with the unifying claim across domains that phonetic reduction is observed in easy contexts relative to hard contexts. This speaking style effect on phonetic reduction has been observed in read speech in both the temporal domain for words and vowels (Ferguson and Kewley-Port 2007; Picheny, Durlach, and Braida 1986; Scarborough and Zellou 2013; Smiljanic and Bradlow 2005) and the spectral domain for vowels (Ferguson and Kewley-Port 2007; Moon and Lindblom 1994).

In addition to their individual effects on phonetic reduction, the linguistic factors listed in Table 2.1 have also been shown to have independent effects on phonetic reduction when presented in combination. For example, lexical frequency and neighborhood density exhibit independent effects on spectral vowel reduction (Munson and Solomon 2004), neighborhood density and semantic predictability exhibit independent effects on both temporal and spectral vowel reduction (Scarborough 2010), and neighborhood density and speaking style exhibit independent effects on both temporal and spectral vowel reduction (Scarborough and Zellou 2013). The observed independent phonetic reduction effects across linguistic factors suggest a simple additive system related to processing demands. As processing demands are decreased, phonetic reduction is increased, and vice versa. Thus, high-frequency words that are highly predictable are very easy to process and are therefore more reduced than high-frequency words that are less predictable, which in turn are easier (and more reduced) than low-frequency words that are less predictable.

However, interactions between the various linguistic factors have also been observed, suggesting that phonetic reduction may not simply reflect an additive function of the processing demands imposed by the linguistic context. In particular, Baker and Bradlow (2009) observed a three-way interaction among lexical frequency, discourse mention, and speaking style on temporal reduction, in which high-frequency words exhibited more second mention reduction than low-frequency words in plain lab speech, but not in clear lab speech. Baker and Bradlow (2009) attributed this interaction to a maximal reduction in the easiest context (high-frequency, second mention, plain speech), but a lower bound on the permissible degree of reduction in clear speech that reduces the effects of lexical frequency and discourse mention in that style relative to the effects observed in plain speech. Bell et al. (2009) observed a similar interaction between lexical frequency and semantic predictability, in which high-frequency words exhibited a greater effect of semantic predictability on temporal reduction than low-frequency words. As in the Baker and Bradlow (2009) data, this interaction suggests maximal reduction in the easiest context (high-frequency, high-predictability),

but a lower bound on the permissible degree of reduction for low-frequency and/or low-predictability targets.[5]

Taken together, these interactions suggest that a more complex analysis of the phonetic reduction process may be warranted to account for the potential limits on phonetic reduction in various contexts. The observed interactions suggest lower bounds on reduction in some hard contexts, but lower bounds on reduction in extremely easy contexts are also expected. For example, in the temporal domain, the absolute lower bound on phonetic reduction is deletion. That is, the duration of a linguistic unit (i.e., segment, syllable, or word) cannot be reduced to a value less than 0 ms, which may mean that the combined effects of the various linguistic factors contributing to phonetic reduction cannot be additive because the minimum allowable duration is 0 ms (i.e., deletion). Similarly, in the spectral domain, the lower bound on phonetic vowel reduction is potentially a categorical change to schwa. As noted in the Introduction, we consider segmental alternations and deletions to be categorical phenomena that potentially differ from the continuous, phonetic reduction processes we are focused on in this chapter. However, the possibility of segmental alternations and deletions, as well as their effects on how the various linguistic factors in Table 2.1 must interact in promoting phonetic reduction, must be borne in mind as we consider theoretical models of and further empirical evidence for phonetic reduction processes.

2.3 Theoretical approaches to phonetic reduction

A number of theories have been proposed to capture the insight that phonetic reduction emerges in contexts with limited processing demands. As previewed in the previous section, one of the primary dimensions that differentiates these various theories is whether it is the processing demands for the listener (listener-oriented) or the processing demands for the talker (talker-oriented) that are driving the phonetic reduction process.

5 These findings are only partially consistent with Wright's (2004) predictions about the potential interactions among these factors. In particular, although Wright (2004) predicted maximal reduction of "easy" words (i.e., high-frequency words with few neighbors) in easy contexts, as observed by Baker and Bradlow (2009) and Bell et al. (2009), Wright (2004) also predicted maximal enhancement of "hard" words (i.e., low-frequency words with many neighbors) in hard contexts, which was not observed by either Baker and Bradlow (2009) or Bell et al. (2009).

2.3.1 Listener-oriented approaches

From the listener-oriented perspective, phonetic reduction serves the functional purpose of enhancing communicative success while minimizing talker effort. The underlying assumption of this approach is that some segments or words are more likely to be misperceived by the listener than other segments or words, due to acoustic-perceptual factors (such as masking of acoustic cues in certain phonological contexts) and/or linguistic predictability factors (such as the likelihood of an adjective following a noun). According to the listener-oriented perspective, talkers have a tacit awareness of these potential comprehension difficulties and attempt to enhance the acoustic-phonetic prominence of words that are likely to be difficult for the listener to process. Conversely, the talker is free to phonetically reduce easy words, which are likely to be perceived correctly by the listener. These models assume that it is easier for the talker to produce reduced variants than enhanced variants, leading to reduced variants when the listener's successful perception is likely. The listener-oriented perspective, then, claims that hyperarticulation exists to facilitate successful perception by the listener, and reduction exists to ease the articulatory burden on the talker (see also Brouwer, Mitterer, and Huettig 2013; Mitterer and Russell 2013, for evidence that reduction in appropriate contexts can facilitate perception). Successful communication is therefore central to the listener-oriented account (Jaeger 2013; Ramscar and Baayen 2013).

One of the earliest listener-oriented models was Lindblom's (1990) Hyper- & Hypospeech (H&H) theory, which he argued could account for the observation that segmental realization is affected by a range of contextual factors, including those discussed in the previous section. According to H&H theory, speech is produced along a continuum from hyper- to hypoarticulated as a function of the competing goals of the talker to conserve energy (hypoarticulate) and to be understood (hyperarticulate). Contexts in which lexical access is expected to be easier for the listener lead to phonetic reduction relative to contexts in which lexical access is expected to be more difficult.

Aylett and Turk's (2004; see also Aylett 2000; Aylett and Turk 2006; Turk 2010) smooth signal redundancy hypothesis is very similar in spirit to Lindblom's (1990) H&H theory, in that two competing constraints are argued to be at play: reliable communication and conservation of effort. For communication to be reliable, the signal needs to be clear enough for the message to be transmitted successfully. On the other hand, conservation of effort demands that the talker exert minimal effort to convey the message. Too much focus on reliability and the talker's speech provides unnecessarily redundant information; too much focus on brevity and the talker is not understood. According to Aylett and Turk (2004), redundancy in speech communication is of two kinds. One kind is language

redundancy, which is broadly equivalent to the concept of semantic predictability discussed above. More predictable parts of a message are more redundant than less predictable parts. The other kind of redundancy is acoustic redundancy, which is conceptualized as the likelihood that the signal will be perceived correctly based on the acoustic properties alone. The sum of these two redundancies is the total signal redundancy. Aylett and Turk (2004) proposed that language users strive to ensure that the total signal redundancy is smooth (i.e., constant) throughout an utterance. Thus, the balance between language and signal redundancies accounts for the observed relationships between linguistic factors, such as lexical frequency and semantic predictability, and phonetic reduction. When language redundancy is high, because the target word is highly predictable or frequent, signal redundancy can be low, leading to phonetic reduction.

A number of similar listener-oriented accounts of phonetic reduction have been proposed, which invoke concepts qualitatively similar to smooth signal redundancy, including uniform information density (Jaeger 2010; Levy and Jaeger 2007; Qian and Jaeger 2012), communicative efficiency (van Son and Pols 2003; van Son and van Santen 2005), and Bell's (1984) audience design (Galati and Brennan 2010; Schober 1993). Consistent with the functional underpinnings of the listener-oriented approach, these theories are typically linked to broader claims about the critical role of communication in the functional structure of language (e.g., Hawkins 2014).

2.3.2 Talker-oriented approaches

From the talker-oriented perspective, phonetic reduction arises from interactions in the cognitive architecture of the speech production system. The precise formulation and reasoning behind the process is generally theory specific, but the shared theme of these models is that easy words are accessed or processed more quickly and more easily than hard words, which leads to a faster and less precise (i.e., reduced) production for easy words relative to hard words. By contrast, hard words are accessed or processed less quickly and less easily, resulting in a more effortful and precise (i.e., unreduced) production. From this perspective, the ability of the listener to understand the message plays no direct role in shaping the phonetic realization of the speech signal, and successful communication between interlocutors is therefore not central to the process.

One line of evidence for the talker-oriented approach is research demonstrating that talkers do not always take into account the perspective of their interlocutor, and instead appear to rely on their own perspective, in the implementation of phonetic reduction processes. For example, Bard et al. (2000) conducted an

investigation of second mention reduction in the HCRC map task corpus (Anderson et al. 1991). In the materials of interest in Bard et al.'s (2000) study, after the instruction-giver had finished guiding their partner through a map, their partner changed and they had to guide the new partner through the same map. All of the mentions of the landmarks were, in this context, discourse-given from the perspective of the instruction-giver, but discourse-new from the perspective of the partner being led. Bard et al. (2000) found that the productions in this second trial with the new partner were both shorter in duration and less intelligible in isolation than the productions from the first trial, suggesting that second mention reduction had taken place. That is, despite the instruction-giver's awareness that their interlocutor had changed and was therefore not familiar with the discourse context, the instruction-giver still reduced tokens that were discourse-given from the instruction-giver's perspective. Bard et al. (2000) interpreted this finding as evidence of an egocentric pattern of phonetic reduction, in which the talker's situational knowledge assumes primacy over their modeling of the listener's knowledge (see also Keysar 2008; Keysar and Barr 2005; Keysar et al. 2000).

More recently, Baese-Berk and Goldrick (2009) carried out a series of experiments in which a participant instructed their partner to click on an item on a computer display. Both the instructor and the partner saw the same display of items. In a condition where two of the displayed items were referents of a voice-onset-time (VOT) minimal pair (e.g., *cod* and *god*), more extreme VOT values were observed on the target word relative to a condition in which the target word did not have a real-word minimal pair competitor (e.g., *cog*, where *gog* is not a real word in English). This phonetic enhancement of the aspiration contrast in a potentially ambiguous context is consistent with a listener-oriented perspective. However, when the same target item *cod* was displayed without its minimal pair competitor, VOT enhancement was still observed, albeit to a smaller degree. Baese-Berk and Goldrick (2009) argued that this VOT enhancement in an unambiguous context cannot be accounted for by a listener-oriented perspective and suggested instead that lexical competition in production drives the enhancement effect. Various replications of Baese-Berk and Goldrick's (2009) findings in situations that do not involve a communicative partner (Bullock-Rest et al. 2013; Fox, Reilly, and Blumstein 2015; Kirov and Wilson 2012; Peramunage et al. 2011) provide further evidence against a purely communicative account of the phenomenon. In particular, in the absence of a live interlocutor, the communicative imperative to speak clearly to distinguish minimal pair targets is arguably absent.

However, this line of argumentation suggests that most of the data presented in the previous section should be taken as evidence for a talker-oriented approach to phonetic reduction. In particular, the effects of lexical frequency, neighborhood density, semantic predictability, and discourse mention described

above are all observed in the absence of a live interlocutor. If real-time communication is required for listener-oriented adjustments, such adjustments should not be observed in laboratory settings without an immediate communicative task. That is, if phonetic reduction reflects an adjustment for the listener, phonetic reduction should not be observed when a listener is not physically present. However, numerous studies have shown that talkers can make explicit speaking style adjustments for imagined interlocutors in this kind of noncommunicative laboratory setting (e.g., Ferguson and Kewley-Port 2007; Picheny, Durlach, and Braida 1986; Smiljanic and Bradlow 2005), suggesting that real-time communication is not necessary for listener-oriented adjustments to take place. Furthermore, adjustments in duration and vowel space size can be comparable to those that are produced when a live interlocutor is present, although other processes such as coarticulation and speaking rate show significant effects of real versus imagined interlocutors (Scarborough et al. 2007; Scarborough and Zellou 2013). Although participants in many laboratory studies are not talking to another person, they are producing speech in a laboratory setting and are aware that their speech is being recorded, implying that someone (e.g., the researcher or participants in a future study) will eventually listen to their speech (see also Wagner, Trouvain, and Zimmerer 2015). Thus, recorded speech in a laboratory is not comparable to true self-directed speech with no communicative intent, and the lack of an explicit communicative context may not be sufficient to negate a listener-oriented interpretation of phonetic reduction processes.

2.3.3 Evolutionary approaches

In addition to the listener-oriented and talker-oriented perspectives, a third explanation for the relationship between processing demands and phonetic reduction has been proposed. This set of theories differs from the previous two perspectives in holding that no active force is responsible for phonetic reduction. Specifically, rather than communicative pressure or cognitive architecture producing these effects, phonetic reduction simply exists as a natural consequence of patterns of language acquisition and change over generations. Segments or words that are easy to perceive are generally perceived correctly, whereas segments or words that are difficult to perceive are only perceived correctly if they are sufficiently acoustically prominent. Over time, segments and words that are perceived correctly (i.e., easy words and acoustically prominent hard words) become the principal component of language; all other modes of production fall into disuse (Garrett and Johnson 2012; Pierrehumbert 2001a, 2002; Silverman 2012). Silverman (2012, p. 147) expressed this position as follows (emphasis in original):

> Successful speech propagates; unsuccessful speech does not. Confusing speech tokens may be misunderstood, and thus not pooled with the exemplars of the intended word, and so the system maintains its state of semantic clarity. Anti-homophony is thus not an *active* pressure for which there is an abundance of overt evidence. Rather, it is a *passive* result of the pressures that inherently act upon the interlocutionary process.

One of the few explicit formulations of this perspective comes from Pierrehumbert's (2001a, 2001b, 2002, 2003a, 2003b) work on exemplar-based phonology. Her description involves an exemplar model (Goldinger 1998; Johnson 1997; Tenpenny, 1995) in which each perceived word token has its own representation in a perceptual cloud (see also Blevins and Wedel 2009; Tupper 2014; Wedel 2006, for refinements and extensions of these mechanisms). In Pierrehumbert's (2002) model, phonetic reduction effects emerge as a simple consequence of the acquisition process. In particular, when a high-frequency word is uttered, the listener can guess the word's identity with relative ease, even if it is not acoustically prominent, due to its high frequency. When the word is identified, the token is added to the listener's exemplar cloud and becomes part of that word's representation. However, when a low-frequency word is uttered, the listener cannot as easily guess the word's identity (due to its low frequency), and the token therefore needs to be more acoustically prominent than the high-frequency word for its identity to be ascertained correctly. When the word is not correctly identified, the token is not added to the listener's exemplar cloud and does not become part of the target word's representation. Thus, the low-frequency word token will only be added to the exemplar space if it is sufficiently acoustically prominent (Tupper 2014). Over time, then, the exemplar space will contain acoustically prominent low-frequency words, and both prominent and nonprominent high-frequency words. In speech production, the talker selects a token at random from the exemplar space of the target word (see Pierrehumbert 2001a, 2002, for mathematical details of the implementation). High-frequency words will tend to be reduced in production relative to low-frequency words because their exemplar clouds contain both reduced and unreduced variants, whereas the exemplar clouds of the low-frequency words contain primarily unreduced variants, leading to unreduced productions of these targets. Within a speech community, this behavior facilitates a positive feedback loop leading to clear productions of low-frequency words and reduced productions of high-frequency words.

Additional support for this evolutionary perspective comes from animal behavior research, suggesting that nonhuman animal communication systems are structured to allow for maximal information transmission with minimal effort (see, e.g., Bezerra et al. 2010; Semple, Hsu, and Agoramoorthy 2010; Semple et al. 2013, on primates and Luo et al. 2013, on bats). Ferrer-i-Cancho et al. (2013) explicitly argued that all communication systems, including human language, are

governed by basic distributional properties that enhance efficiency of coding. For a communication system to persist, successful communication with the lowest possible energy expenditure is necessary (see Ferrer-i-Cancho and Elvevåg 2010; Ferrer-i-Cancho and Moscoso del Prado 2011, for statistical approaches to this reasoning). Under this view, the observed linguistic effects on phonetic reduction are a necessary consequence of natural selection and no appeal to cognitive or psychological mechanisms is needed.

2.4 Complexifying our understanding of phonetic reduction

The listener-oriented, talker-oriented, and evolutionary approaches differ considerably in their assumptions regarding the root cause of phonetic reduction. However, these models share the assumption that one underlying factor (e.g., cognitive processing demands) drives the phonetic reduction effects that are observed across temporal and spectral acoustic domains and across lexical, contextual, and stylistic contexts. However, recent research in our laboratory has revealed variation in phonetic reduction processes, suggesting that a simple relationship between processing demands and phonetic reduction may not be sufficient to account for these various effects on segmental realization. In particular, we have observed complex interactions among linguistic factors, social factors, and cognitive factors in temporal, spectral, and prosodic phonetic reduction processes. These interactions reveal variation in the robustness of the linguistic effects on phonetic reduction, as well as different patterns of interactions among linguistic, social, and cognitive factors in temporal and spectral reduction, suggesting diverse phonetic reduction processes across acoustic domains. These findings challenge the notion of a simple linear mapping between phonetic reduction and processing difficulty.

2.4.1 Interactions among linguistic factors

One component of our recent research on phonetic reduction has explored interactions among linguistic factors on temporal and spectral vowel reduction. This work builds on previous research demonstrating both independent and interactive effects of these factors on temporal and spectral reduction (e.g., Baker and Bradlow 2009; Bell et al. 2009; Munson and Solomon 2004; Scarborough 2010; Scarborough and Zellou 2013), and extends the analysis to consider the relationships among more factors simultaneously.

We conducted a large experiment to explore phonetic reduction in read passages in which we manipulated lexical frequency, lexical neighborhood density, semantic predictability, discourse mention, and speaking style in a fully crossed design (Burdin, Turnbull, and Clopper 2015; Clopper, Turnbull, and Burdin in press). The materials were a set of short stories read by Midwestern undergraduates and containing target words with the stressed vowels /i, ɛ, æ, ɑ, ɔ, u/. The target words varied in lexical frequency (as presented in the Hoosier Mental Lexicon; Nusbaum, Pisoni, and Davis 1984), lexical neighborhood density (as presented in the Hoosier Mental Lexicon; Nusbaum, Pisoni, and Davis 1984), and semantic predictability (as assessed by an independent cloze task with Midwestern undergraduates). Each word was included twice in the same story to elicit discourse mention effects. The complete set of stories was read twice by each talker – first to an imagined friend and then again to an imagined hearing-impaired or nonnative listener – to elicit plain and clear lab speech, respectively. For each target word in each story, vowel duration and vowel dispersion, defined as the Euclidean distance from the center of the F1 × F2 vowel space in Bark, for the primary stressed vowel were obtained.

Mixed-effects regression models predicting vowel dispersion from the five linguistic factors (lexical frequency, lexical neighborhood density, semantic predictability, discourse mention, and speaking style) and their interactions revealed the expected main effects of lexical frequency, discourse mention, and speaking style, as well as a four-way interaction between lexical frequency, lexical neighborhood density, semantic predictability, and discourse mention. None of the other main effects or interactions were significant for the vowel dispersion measure. As shown in the top left panel of Figure 2.1, vowels in high-frequency words exhibited less dispersion in the vowel space than vowels in low-frequency words, and vowels in plain speech exhibited less dispersion in the vowel space than vowels in clear speech. Vowels in second mention words also exhibited less dispersion in the vowel space than vowels in first mention words (2.08 vs. 2.14 Bark, respectively). The four-way interaction further revealed effects of lexical neighborhood density and semantic predictability in the expected directions: vowels in low-density words exhibited less dispersion than vowels in high-density words and vowels in high-predictability words exhibited less dispersion than vowels in low-predictability words. Unlike the main effects of lexical frequency, discourse mention, and speaking style, however, these effects of lexical neighborhood density and semantic predictability were more variable across contexts, and thus, significant main effects did not emerge.

Although previous research has not examined vowel dispersion as a function of discourse mention, significant effects of lexical neighborhood density and semantic predictability on vowel dispersion have been reported in previous work (e.g., Aylett and Turk 2006; Clopper and Pierrehumbert 2008; Jurafsky et al. 2001; Munson and

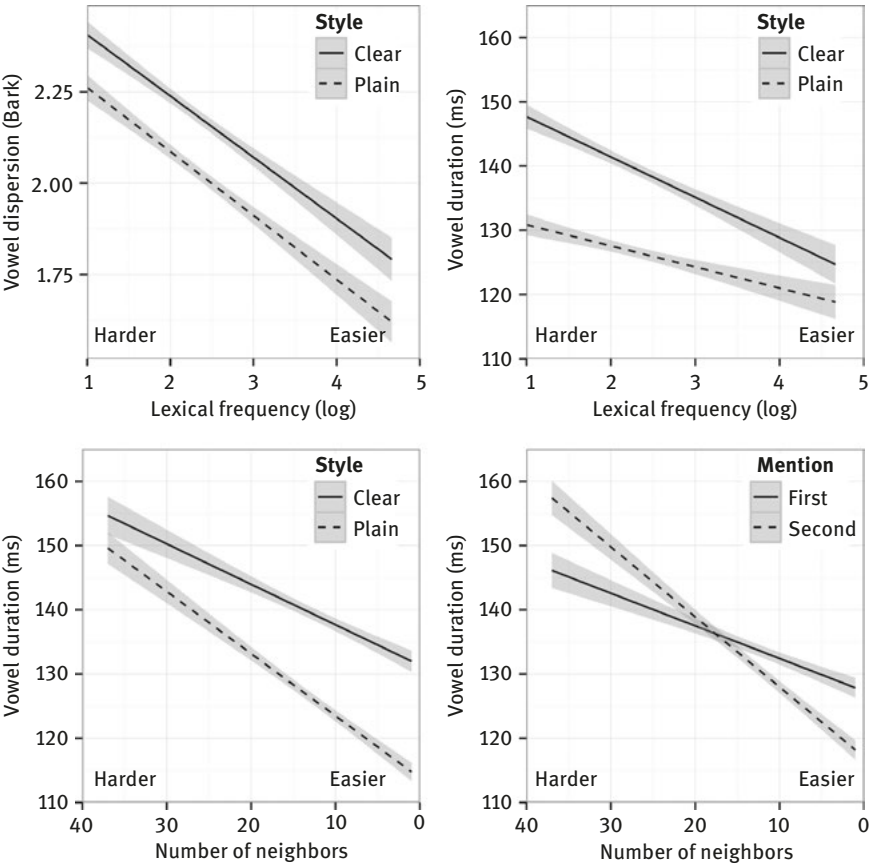

Figure 2.1: Lexical frequency (log occurrences per million words) and speaking style effects on vowel dispersion, defined as the Euclidean distance from the center of the F1 × F2 Bark space (top left), lexical frequency and speaking style effects on vowel duration (top right), lexical neighborhood density and speaking style effects on vowel duration (bottom left), and lexical neighborhood density and discourse mention effects on vowel duration (bottom right) in read short stories. Adapted from Burdin, Turnbull, and Clopper (2015).

Solomon 2004). Thus, the lack of significant main effects of lexical neighborhood density and semantic predictability on vowel dispersion in Burdin, Turnbull, and Clopper's (2015) study is somewhat surprising. This null result may reflect variability in these effects across vowel categories (see, e.g., Clopper and Pierrehumbert 2008; Scarborough 2010; Wright 2004) or the relative sizes of the effects. Munson and Solomon (2004) reported a much larger effect size for lexical frequency than lexical neighborhood density on vowel space dispersion and Clopper et al. (2017) reported more robust effects of speaking style than neighborhood density or discourse mention on vowel space dispersion (see below). Thus, the significant effects

of lexical frequency and speaking style in Burdin, Turnbull, and Clopper's (2015) study may have swamped any smaller effects of the other factors.

Mixed-effects regression models predicting vowel duration from the five linguistic factors and their interactions also revealed the expected main effects of lexical frequency and speaking style. As shown in the top right panel of Figure 2.1, high-frequency words had shorter vowels than low-frequency words and vowels in plain speech were shorter than vowels in clear speech. None of the other main effects were significant, although a number of significant interactions were observed for vowel duration. As shown in the top right panel of Figure 2.1, lexical frequency and speaking style interacted such that the lexical frequency effect was larger in clear speech than in plain speech. This pattern of interaction contrasts with the interaction observed for lexical neighborhood density and speaking style, shown in the bottom left panel of Figure 2.1, in which the lexical neighborhood density effect was larger in plain speech than in clear speech. The interaction between lexical neighborhood density and discourse mention, shown in the bottom right panel of Figure 2.1, parallels the neighborhood density × speaking style interaction and reveals a larger lexical neighborhood density effect for second mentions than for first mentions.[6] Thus, consistent with previous findings in which the effects of lexical frequency and discourse mention were greater in plain speech than in clear speech (Baker and Bradlow 2009), we find evidence for a larger effect of lexical neighborhood density in plain speech than in clear speech and for second mentions than for first mentions. These results are consistent with the maximization of temporal reduction in easier (i.e., low-density, second mention, plain speech) contexts relative to harder (i.e., high-density, first mention, clear speech) contexts. In contrast, the interaction between lexical frequency and speaking style is not consistent with this interpretation and may reflect a lower bound on temporal reduction in easy contexts. That is, high-frequency words in plain speech may not be maximally reduced because further temporal reduction would lead to deletion.

Three findings emerge from these results that suggest that phonetic reduction may reflect a more complex process than simple additive effects of processing difficulty. The first of these findings is that the patterns of phonetic reduction differ

6 This interaction between lexical neighborhood density and discourse mention also exhibits a cross-over effect, suggesting that first mentions were phonetically reduced relative to second mentions for words with many phonological neighbors. This apparent reversal of the discourse mention effect for words with many neighbors may reflect other factors contributing to vowel duration, including prosodic structure (see Burdin and Clopper 2015) or information structure (see, e.g., Wagner and Klassen 2015).

across acoustic domains. Although most previous research has focused on the temporal reduction of words and vowels in easy contexts relative to hard contexts (e.g., Aylett and Turk 2004; Bell et al. 2009; Fowler and Housum 1987; Gahl, Yao, and Johnson 2012), studies that have examined spectral reduction have typically observed spectral reduction in the same easy contexts in which temporal reduction is typically observed (e.g., Aylett and Turk 2006; Clopper and Pierrehumbert 2008; Munson and Solomon 2004; Scarborough 2010, 2013; Wright 2004). However, Burdin, Turnbull, and Clopper's (2015) results reveal comparable main effects of lexical frequency and speaking style on temporal and spectral vowel reduction, but different patterns of interactions among these and other linguistic factors in the two acoustic domains, suggesting that temporal and spectral reduction exhibit different linguistic constraints and may arise from different processes associated with processing difficulty.

Further evidence for differences in phonetic reduction across acoustic domains comes from Turnbull's (2017) analysis of data obtained in an experiment conducted by Ito and Speer (2006). This experiment involved a naïve participant instructing a confederate in the decoration of a Christmas tree. The type of ornament to be hung and its location on the tree were presented to the participant on a computer screen, but no explicit instructions were provided about how to phrase the instructions to the confederate. Thus, the speech elicited in this task was truly spontaneous and interactive. Ornaments varied in both color and shape, necessitating their description as adjective-noun phrases, like *blue drum*. The target words were coded for discourse mention as either the first or a subsequent mention in the decoration of the Christmas tree. The effect of discourse mention on vowel duration and peak f0 of the stressed syllables of the target adjectives and nouns were examined, after controlling for variation in phonological pitch accent type. As shown in Figure 2.2, reduction in duration for subsequent mentions was observed relative to first mentions, consistent with prior research (Baker and Bradlow 2009; Fowler and Housum 1987). However, no effect of discourse mention was observed for peak f0, suggesting that second mention reduction may not extend to the domain of intonation.

The second critical finding from Burdin, Turnbull, and Clopper's (2015) results is that the interactions observed for temporal reduction suggest both maximization of reduction in some easy contexts (e.g., low-density, plain speech), as suggested in previous research (Baker and Bradlow 2009; Bell et al. 2009), and a lower bound on reduction in other easy contexts (e.g., high-frequency, plain speech). Additional evidence from our laboratory for a lower bound on reduction comes from a recent analysis of segmental deletion in interview speech (Turnbull 2015a, in press). Previous studies of segmental deletion in interview speech have revealed widespread deletion in words of all sizes (Johnson 2004), as well as more frequent

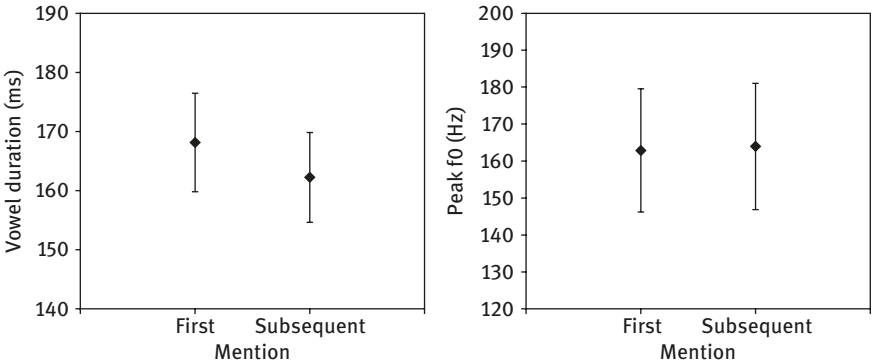

Figure 2.2: Effect of discourse mention on mean vowel duration (left) and peak f0 (right) in Ito and Speer's (2006) Christmas tree decorating task. Error bars are standard error of talker means. Adapted from Turnbull (2017).

/t, d/ deletion in easy words (i.e., high-frequency or high-predictability words) than in hard words (i.e., low-frequency or low-predictability words; Raymond, Dautricourt, and Hume 2006; Jurafsky et al. 2001). Turnbull's (2015a) analysis of the Buckeye Corpus of Conversational Speech (Pitt et al. 2007), which is a phonetically aligned corpus of approximately 40 hours of spontaneous, interview speech, extended this previous work and considered the roles of lexical frequency and lexical neighborhood density on segmental deletion. Each word in the Buckeye Corpus is tagged with both a phonemic (dictionary) transcription and a phonetic (narrow) transcription. By comparing these transcriptions, the number of deleted phones in each of the 282,435 word tokens in the corpus was determined.

A mixed-effects Poisson regression model predicting the number of deleted phones from the target's lexical frequency, lexical neighborhood density, and the number of phonemes in the target's citation form revealed the expected effect of number of phonemes – longer words tended to exhibit more deletions than shorter words, because shorter words can only delete so many phonemes before the word is unintelligible. The expected effects of lexical frequency and lexical neighborhood density were also observed. Harder words in denser neighborhoods tended to have fewer deleted phones than easier words in sparser neighborhoods and harder, less frequent words tended to have fewer deleted phones than easier, more frequent words. However, as shown in Figure 2.3, these two factors interacted such that lexical frequency effects were observed for words in denser neighborhoods (i.e., more than 3 neighbors), but no effect of lexical frequency was observed for words in sparser neighborhoods (i.e., 0–3 neighbors). These low-density words, regardless of lexical frequency, exhibited a high mean phone deletion rate of just over 0.6 phones per word. Thus, the easy, low-density

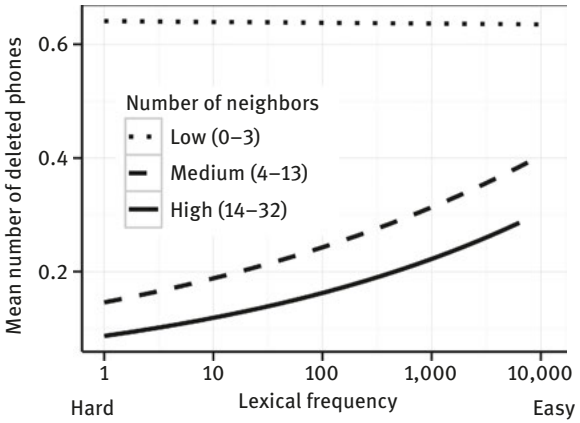

Figure 2.3: Effects of lexical frequency (number of occurrences in the Buckeye Corpus) and lexical neighborhood density on the mean number of deleted phones across words in the Buckeye Corpus. Adapted from Turnbull (2015a).

words exhibit an upper bound on deletion that is comparable to the lower bound on temporal reduction observed for the high-frequency words produced in plain speech in Burdin, Turnbull, and Clopper's (2015) study. That is, low-density words exhibit the maximal number of deleted phonemes and high-frequency words exhibit the minimal vowel duration that the production system allows. This parallel in bounds on reduction suggests a strong connection between the phenomena that we have characterized as categorical versus continuous (see also Cohen Priva 2015), further suggesting that such a distinction is ultimately arbitrary.

The third critical finding from the Burdin, Turnbull, and Clopper (2015) study is that cloze predictability was unexpectedly not a significant independent predictor of either temporal or spectral reduction. As noted above, this null result may reflect variability in the magnitude of the effect across vowel categories or a relatively small effect size. However, in a series of recent studies, we have explored alternative dimensions of semantic predictability and their relative contributions to phonetic reduction in the temporal and prosodic domains. In one of these studies, Turnbull (2017) analyzed a set of data from an experimental investigation of focus marking in English to explore potential effects of semantic predictability on the realization of word duration and peak f0. Crucially, as in the analysis of the Christmas tree data described above, this analysis took phonological pitch accenting into account, so the results cannot be reduced to phonological effects of accent choice, but rather must be attributed to adjustments at the phonetic level. The data set was drawn from an experiment conducted by Turnbull et al. (2015) and Burdin, Phillips-Bourass et al. (2015), which featured a naïve participant instructing a confederate in an object-placing task. The task involved placing

tiles depicting colored objects into numbered boxes on a game board. The objects depicted on the tiles were differentiated in both color and shape, and the participants' instructions were of the form "put the ADJECTIVE NOUN in box NUMBER." The order of the tiles to be placed was manipulated to elicit focus on different linguistic expressions across utterances. Following Rooth (1992), we consider focus to be a semantic property denoting a set of alternatives to the asserted content, not a phonological prosodic property of the utterance. Thus, for example, in the sequence *green lion … blue lion*, the adjective *blue* is focused as a contextually relevant alternative to *green*, whereas in *blue train … blue lion*, the noun *lion* is focused as a contextually relevant alternative to *train*. The set of available tiles was finite and visually salient to both interlocutors, which meant that, as more tiles were played, the individual probability of any one tile being played increased.

Turnbull's (2017) analysis of these data revealed an inverse relationship between peak f0 and utterance probability given the number of remaining available tiles, as shown in Figure 2.4. This result extends previously established effects of probability on duration (Aylett and Turk 2004) to the f0 dimension. Turnbull (2017) also observed an effect of utterance probability on word duration, such that more probable items were produced with a shorter duration, as expected, but this effect held only for nonfocused nouns. An effect of utterance probability on word duration was not observed for focused nouns or for any adjectives. This result suggests that semantic predictability and focus can interact, with focus essentially "blocking" temporal effects of predictability. This interaction is similar to the interaction that Baker and Bradlow (2009) observed in which the effects of lexical frequency and discourse mention were reduced (or "blocked") in clear speech relative to plain speech.

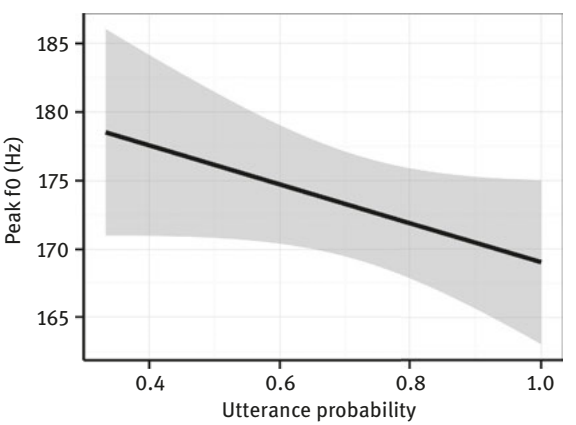

Figure 2.4: Effect of utterance probability on syllable peak f0. Adapted from Turnbull (2017).

A final relevant manipulation in Turnbull et al.'s (2015) study was that the constituent in the instructions that would be focused was either predictable or unpredictable from the global context of each game board. For example, one game board consisted of all red tiles, in which case the noun in each instruction was focused (*red* LION, *red* TRAIN). In other boards, the focused constituent was not predictable from the global context, and which constituent was focused changed from utterance to utterance. The hypothesis under investigation was that phonetic cues to focus, such as word duration and peak f0, would be less prominent in the predictable condition than in the unpredictable condition, due to the contribution of the context to the listener's interpretation of the utterance. The analyses presented by both Turnbull (2017) and Turnbull et al. (2015) found support for this hypothesis. Differences in word duration and peak f0 were larger across focus conditions in the unpredictable condition than in the predictable condition, independent of phonological pitch accent type or phrasing. As shown in the top two panels of Figure 2.5, the effect of context was more robust for peak f0 than for word duration, which was more variable within and across conditions. However, taken together, the results demonstrate that when the context provides information about the relevant semantic contrasts, the talker produces smaller prosodic cues to indicate those semantic contrasts, consistent with Aylett and Turk's (2004, 2006) proposal for a trade-off between language and acoustic redundancies to produce a constant signal redundancy.

The results of Turnbull's (2017) study therefore provide further evidence for variation in phonetic reduction across acoustic domains, as well as evidence for variation in phonetic reduction across different dimensions of semantic predictability. Whereas context condition (predictable vs. unpredictable) exhibited consistent effects across acoustic domains (word duration and peak f0), utterance probability exhibited a robust effect only in the f0 domain. Thus, different dimensions of semantic predictability reveal different phonetic reduction patterns within the same data set. Further, Turnbull's (2017) analysis of utterance probability revealed complex interactions between semantic predictability and other linguistic factors (i.e., focus and word class), which exhibit independent prosodic effects of pitch-accenting and phrasing on phonetic prominence. Given these complex patterns of interaction, the analysis of phonetic reduction must involve careful consideration of all potentially relevant linguistic and contextual factors that contribute to the realization of acoustic-phonetic prominence.

To explore the cross-linguistic generalizability of the American English findings, Turnbull et al. (2015) also examined data from Paraguayan Guaraní in the same tile-placing game that was used with the American English participants. Paraguayan Guaraní has a similar overall prosodic structure to American English, including lexical stress and phrase-level prominences realized through

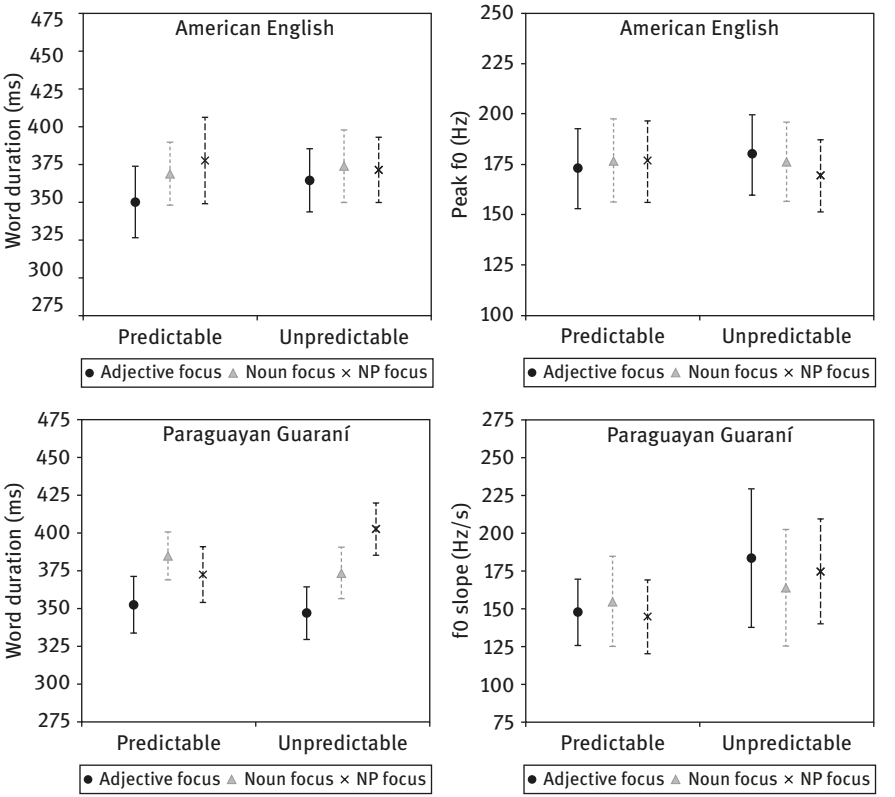

Figure 2.5: Mean word duration (left panels) and f0 prominence (right panels) of adjectives and nouns in American English (top panels) and Paraguayan Guaraní (bottom panels) noun phrases as a function of the focused expression in the noun phrase (adjective, noun, or noun phrase) and experimental context (predictable or unpredictable). Error bars are standard error of talker means. Adapted from Turnbull et al. (2015).

pitch accenting, but differs in the size of its pitch accent inventory (two in Paraguayan Guaraní vs. five in American English) and the number of levels of prosodic phrasing above the word (one in Paraguayan Guaraní vs. two in American English; see also Turnbull et al. 2015; Burdin, Phillips-Bourass et al. 2015). The similarities in the overall prosodic structure allow for a meaningful comparison across languages, whereas the differences allow variation in the realization of contextual effects to emerge. As shown in the bottom two panels of Figure 2.5, word duration in Paraguayan Guaraní was affected by focus condition, with shorter words in adjective focus and longer words in noun and noun phrase focus, independent of pitch accent type and phrasing, but word duration was not significantly affected by context or its interaction with focus condition. However,

context had a significant effect on the f0 slope of the Paraguayan Guaraní pitch accents, independent of phonological pitch accent type. The slope of the pitch accents was steeper when the focused expression was not predictable from the context relative to when the focused expression was predictable from the context. Thus, although both American English and Paraguayan Guaraní exhibit a pattern that can be interpreted as prosodic reduction in an easier (i.e., more predictable) context, the patterns differ considerably across languages. In Guaraní, prosodic prominence was globally reduced through shallower f0 slopes in the easier predictable context relative to the harder unpredictable context and these effects of context did not interact with focus.

Taken together, the recent findings in our laboratory suggest substantial variation in phonetic reduction across acoustic domains and across linguistic factors. Phonetic reduction is most robust in the temporal domain and effects are more variable in the spectral and prosodic domains. These differences across acoustic domains may reflect the relative contributions of these sources of information to phonological contrasts in English. Whereas vowel spectral information and f0 information are critical for distinguishing vowel quality and intonational contrasts, respectively, duration plays a relatively minor role in distinguishing vowel contrasts and may therefore be available for conveying other lexical or contextual information. With respect to linguistic factors, we have observed variation in the strength of phonetic reduction effects across dimensions of semantic predictability, as well as different patterns of interactions among the linguistic factors that contribute to phonetic reduction and between those factors and other factors that contribute to variation in acoustic-phonetic prominence. Segmental duration in particular is shaped by numerous linguistic and contextual factors (Klatt 1976) and these factors must be considered when phonetic reduction is examined. Finally, our results from Paraguayan Guaraní suggest that phonetic reduction processes may also differ in their implementation across languages.

We interpret these results as strong evidence that a simple dichotomy between easy and hard processing contexts, such as that presented in Table 2.1, is insufficient to account for phonetic reduction patterns within or across languages. Minimally, the distinction between easy and hard contexts must be elaborated to account for the observed variability in effect sizes across acoustic domains and linguistic factors. For example, processing demands could be conceptualized as a continuum from easy to hard, with different linguistic factors covering different ranges of the continuum or exhibiting different constraints on their possible realization along the continuum. This idea is consistent with Baker and Bradlow's (2009) proposal for a lower bound on phonetic reduction in clear speech: if clear speech is constrained to a particular range of the "hard" end of the processing demands continuum, the combined effects of other linguistic factors contributing

to phonetic reduction may not lead to as much reduction in clear speech as in plain speech if plain speech has fewer constraints on its possible range. Similarly, the permissible range of variation may vary across acoustic domains, so that larger differences in processing demands are required for phonetic reduction effects to emerge in spectral or prosodic domains than in the temporal domain.

The adoption of a gradient, nonbinary interpretation of processing difficulty is relatively trivial and not at odds with any of the previous work in this area. As noted above, many of the linguistic factors are themselves continuous variables (e.g., lexical frequency, lexical neighborhood density, some measures of semantic predictability) or could straightforwardly be transformed to ordinal (e.g., style) or numerical (e.g., discourse mention) variables. The nature of the nonlinear relationships will be more difficult to determine, but primarily requires substantially more data from production and perception to allow us to characterize not only the nature of the processing difficulties imposed by each of the relevant factors, but also the magnitude of phonetic reduction effects for each of the relevant factors in various combinations across acoustic domains. Thus, the next stage of research in this area will require us to untangle the nonlinear relationships among these numerous continuous variables. Understanding these nonlinear relationships is an essential first step toward determining how much of phonetic reduction can be accounted for by this proposed elaboration of the processing demands explanation.

2.4.2 Interactions between linguistic factors and dialect variation

A second component of our recent research on phonetic reduction has explored the interactions between dialect variation and the linguistic factors contributing to phonetic reduction. A small, but growing, literature suggests that talkers produce more marked social information in easy processing contexts relative to hard processing contexts. For example, Oprah Winfrey, an African-American talk-show host, produces more African-American features for easy, high-frequency words than for harder, low-frequency words (Hay, Jannedy, and Mendoza-Denton 1999). Similarly, gender differences are more pronounced in easy, low-density words than in harder, high-density words (Munson 2007; see also Scarborough 2010, and commentary by Flemming 2010, suggesting more extreme dialect-specific variants are produced in low-density words than high-density words).

In a series of recent studies (Clopper, Mitsch, and Tamati 2017; Clopper and Pierrehumbert 2008; Clopper and Tamati 2014; Turnbull and Clopper 2013),

we have confirmed this general observation that more extreme dialect variants are observed in easier (i.e., low-density, high-predictability, second mention, plain speech) contexts than in harder (i.e., high-density, low-predictability, first mention, clear speech) contexts. However, a closer inspection of the results of these studies reveals variation in the interactions between dialect variation and linguistic factors across vowels and across acoustic domains (see Figures 2.6–2.9).

First, in an investigation of the effect of lexical neighborhood density on vowel reduction and dialect-specific variants in the Midland and Northern dialects of American English (Clopper, Mitsch, and Tamati 2017), more extreme dialect-specific variants were observed consistently for low-density words relative to high-density words in the spectral domain, but dialect differences were enhanced in the temporal domain only for /i/. In particular, although no effect of lexical neighborhood density on vowel duration was observed (see the left panel of Figure 2.6), the Northern vowels were longer than Midland vowels overall and this difference was exaggerated for easy, low-density /i/ words relative to hard, high-density /i/ words. In the spectral domain, we observed more extreme dialect-specific variants, including raising and fronting of /æ/ by the Northern talkers and fronting of /u/ for both talker dialects, in the easy, low-density words than in the hard, high-density words, as shown in the right panel of Figure 2.6. Thus, in the spectral domain, the observation that dialect information is marked more strongly in easy contexts relative to hard contexts was robust across vowel categories, but in the temporal domain, lexical neighborhood density interacted with dialect variation only for one of the four vowels examined.

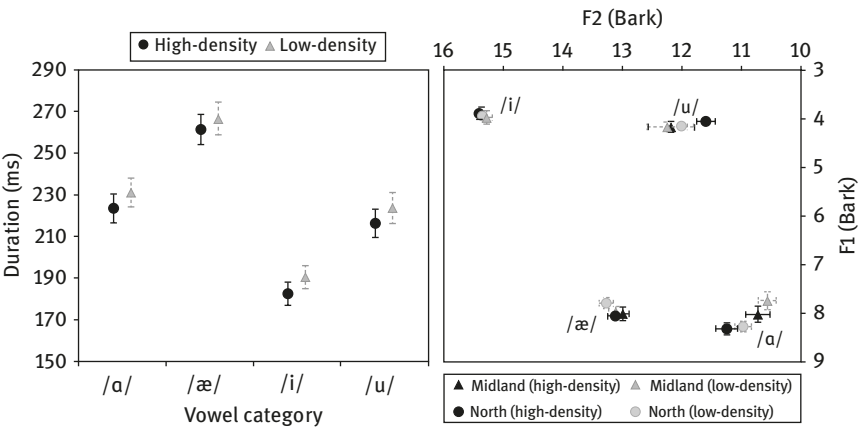

Figure 2.6: Effects of lexical neighborhood density on mean vowel duration (left) and mean vowel formant frequencies (right). Error bars show standard error of talker means. Adapted from Clopper et al. (2017).

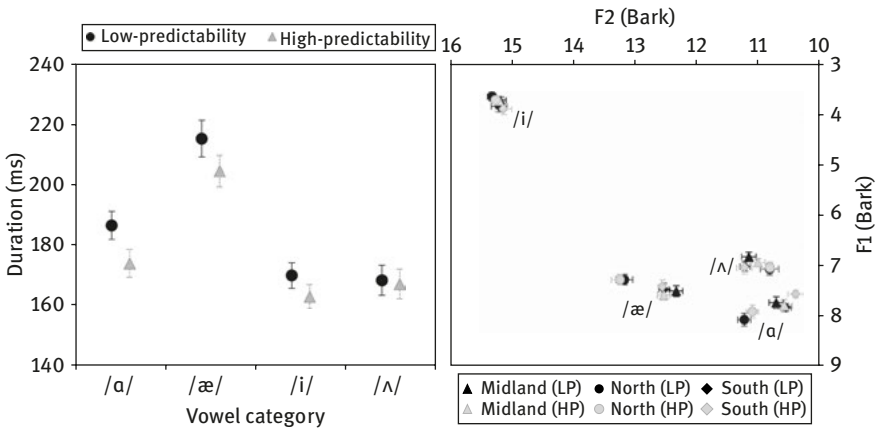

Figure 2.7: Effects of semantic predictability on mean vowel duration (left) and mean vowel formant frequencies (right). Error bars show standard error of talker means. Adapted from Clopper and Pierrehumbert (2008).

Second, in an investigation of the effect of semantic predictability on vowel reduction and dialect-specific variants in the Midland, Northern, and Southern dialects of American English (Clopper and Pierrehumbert 2008), semantic predictability did not interact with dialect variation in the temporal domain and interacted with dialect variation for only one vowel in the spectral domain. In the temporal domain, we observed the expected effect of semantic predictability for three of four vowels (/i, æ, ɑ/, but not /ʌ/), as shown in the left panel of Figure 2.7. Vowels were shorter in easy, high-predictability words than in harder, low-predictability words. In the spectral domain, we observed the expected effect of semantic predictability on vowel dispersion for the Southern talkers for /i, æ, ɑ/ and for the Northern talkers for /i, ɑ/, as shown in the right panel of Figure 2.7. Vowels were less dispersed in the vowel space in easy, high-predictability words than in harder, low-predictability words. In addition, as shown in the right panel of Figure 2.7, we observed greater dialect-specific fronting of /æ/ for the Northern talkers in the easy, high-predictability context relative to the hard, low-predictability context. Thus, for semantic predictability, the interaction between phonetic reduction and dialect variation processes is limited to the spectral domain and to one of the four vowels we examined.

Third, in an investigation of the effect of discourse mention on vowel reduction and dialect-specific variants in the Midland and Northern dialects of American English (Clopper, Mitsch, and Tamati 2017), discourse mention did not interact with dialect variation in the temporal domain and interacted with dialect variation for only one vowel in the spectral domain. In the temporal domain, we observed the expected effect of discourse mention for three of the four vowels (/æ,

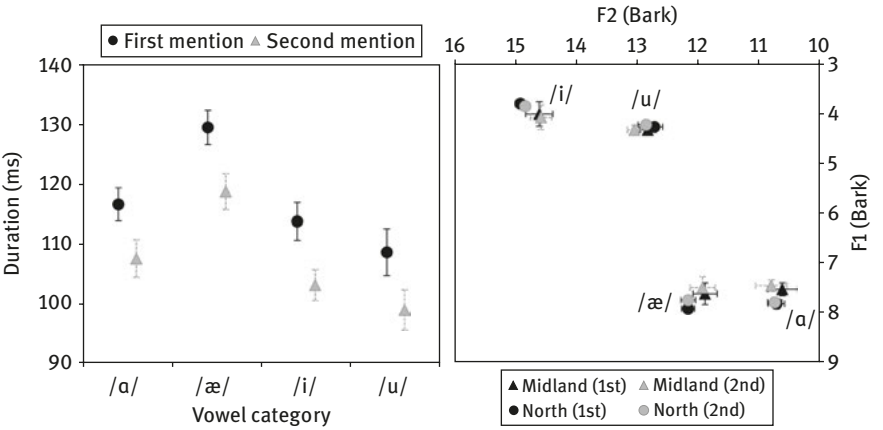

Figure 2.8: Effects of discourse mention on mean vowel duration (left) and mean vowel formant frequencies (right). Error bars show standard error of talker means. Adapted from Clopper et al. (2017).

ɑ, u/, but not /i/), as shown in the left panel of Figure 2.8. Vowels were shorter in easy, second mentions than in harder, first mentions. In the spectral domain, we also observed the expected effect of discourse mention on vowel dispersion for three of the four vowels (/i, æ, u/, but not /ɑ/). Vowels were less dispersed in the vowel space in easy, second mentions relative to harder, first mentions, as shown in the right panel of Figure 2.8. In addition, we observed greater dialect-specific fronting of /u/ for both dialects in second mentions than in first mentions, consistent with the findings for lexical neighborhood density. Unlike the findings for both lexical neighborhood density and semantic predictability, however, no effect of discourse mention was observed for the raising and/or fronting of the Northern /æ/. Thus, for discourse mention, the interaction between phonetic reduction and dialect variation processes is also limited to the spectral domain for a single vowel. This conclusion is qualitatively similar to the conclusions drawn from the analysis of semantic predictability, except that the vowels that exhibit the interaction in the spectral domain differ across linguistic factors (/æ/ for semantic predictability and /u/ for discourse mention).[7]

7 Note, however, that /u/ was not examined in the semantic predictability study, so it is possible that the spectral variation patterns observed for lexical neighborhood density could be replicated with semantic predictability. The studies of lexical neighborhood density and discourse mention examined the same set of vowels, however, so direct comparison between those results is highly interpretable.

Finally, in an investigation of the effect of speaking style on vowel reduction and dialect-specific variants in the Midland and Northern dialects of American English (Clopper, Mitsch, and Tamati 2017), speaking style did not interact with dialect variation in the temporal domain and interacted with dialect variation for two vowels in the spectral domain. In the temporal domain, we observed the expected effect of speaking style for all four vowels, as shown in the left panel of Figure 2.9. Vowels were shorter in plain lab speech than in clear lab speech. In the spectral domain, we observed the expected effect of speaking style on vowel dispersion for three of the four vowels (/i, æ, u/, but not /ɑ/), as shown in the right panel of Figure 2.9. Vowels were less dispersed in the vowel space in plain lab speech than in clear lab speech. In addition, we observed more spectral reduction overall for the Northern talkers than Midland talkers in plain lab speech relative to clear lab speech. As in the previous studies, we also obtained evidence for more fronting of /u/ for both talker dialects and more raising of /æ/ for the Northern talkers in plain lab speech than in clear lab speech. In a separate study investigating the effects of speaking style on /ɑj/ monophthongization in Midland and Southern American English, Turnbull and Clopper (2013) observed the expected effects of talker dialect and speaking style, but the two factors did not interact. Southerners produced more monophthongal /ɑj/ than Midland talkers in both speaking styles and both groups of talkers produced more monophthongal /ɑj/ in plain lab speech than in clear lab speech. Thus, in the spectral domain, the observation that dialect information is marked more strongly in easy contexts relative to hard contexts was robust across vowel

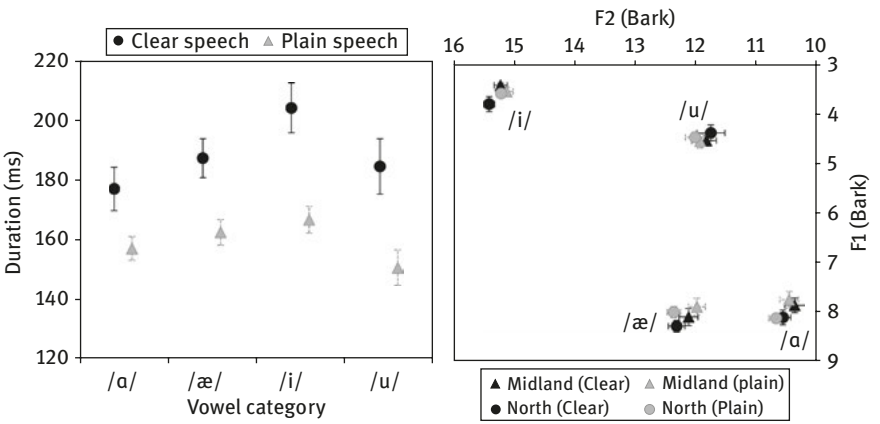

Figure 2.9: Effects of speaking style on mean vowel duration (left) and mean vowel formant frequencies (right). Error bars show standard error of talker means. Adapted from Clopper et al. (2017).

categories, but in the temporal domain, including both vowel duration and vowel trajectory, no interactions between speaking style and talker dialect were observed.

Taken together, the results of these studies provide support for the hypothesis that dialect information is marked more strongly in the same contexts that lead to phonetic reduction. In particular, the results of the previous studies (Hay, Jannedy, and Mendoza-Denton 1999; Munson 2007) and the work in our laboratory described above have demonstrated this relationship between sociolinguistic marking and phonetic reduction for lexical frequency, lexical neighborhood density, semantic predictability, discourse mention, and speaking style. Thus, across linguistic factors, when dialect variation interacts with linguistic context in the temporal and/or spectral realization of vowels, more extreme dialect-specific variants are observed in easier processing contexts relative to harder contexts.

However, the recent research in our laboratory has also shown that this interaction between dialect variation and linguistic context does not emerge robustly across vowel categories or acoustic domains. Although the effects are relatively robust in the spectral domain, they are much weaker in the temporal domain. Whereas interactions between linguistic factors and dialect variation have been observed in the spectral domain for at least some vowel categories in all of the relevant studies, the only interactions between linguistic factors and dialect variation that have been observed in the temporal domain are for dialect differences in duration of /i/ in our work and /ɑj/ monophthongization in Oprah Winfrey's speech in Hay et al.'s (1999) study. This difference between the observed effects in the temporal and spectral domains may reflect the relative importance of temporal and spectral information in conveying dialect information in English. However, the pattern presents an interesting contrast to the results discussed in the previous section in which the temporal domain exhibited more robust phonetic reduction effects than the spectral domain.[8]

The analyses of the interactions between dialect variation and linguistic factors in phonetic reduction processes described in this section were necessarily separated by vowel category because different vowels exhibit different patterns of variation across dialects. These by-vowel analyses revealed variation in phonetic reduction processes across vowels in both acoustic domains, as well as

8 Lexical neighborhood density may present an exception to this general observation. Although we observed significant effects of lexical neighborhood density on vowel duration, but not dispersion, in our study (Burdin, Turnbull, and Clopper 2015), some previous studies on lexical neighborhood effects on phonetic reduction have reported the opposite pattern (e.g., Munson and Solomon 2004).

variation across vowel categories in the interaction between dialect variation and linguistic factors. Specifically, temporal reduction due to semantic predictability and discourse mention was variable across vowels, with only three out of four vowels in each study exhibiting a robust effect. Although the sets of vowels differed in the two studies, some direct comparisons are possible. For example, temporal reduction of /i/ was observed for semantic predictability, but not discourse mention. Similarly, spectral reduction due to semantic predictability, discourse mention, and speaking style was variable across vowels, with only three out of four vowels in each study exhibiting a robust effect. Again, although the sets of vowels differed in the three studies, spectral reduction of /ɑ/ was observed for semantic predictability, but not for discourse mention or speaking style. Wright (2004) also observed variation in spectral reduction across vowel categories in his study of lexical neighborhood density effects and concluded that the point vowels are more likely to exhibit spectral reduction than other vowels because they have more space to centralize. However, our results show a mixed pattern of reduction even for the point vowels, suggesting that additional linguistic and/or nonlinguistic constraints beyond those considered here may be at play in phonetic reduction processes (see also Gahl 2015; Holliday and Turnbull 2015).

Further, although more advanced fronting of /u/ was observed in the easy context for both Midland and Northern talkers in all three studies in which we examined /u/ (i.e., lexical neighborhood density, discourse mention, speaking style), the raising and fronting of /æ/ by Northern talkers was more variable across studies. We observed more advanced raising and/or fronting of /æ/ by the Northern talkers in the low-density, high-predictability, and plain speech contexts, but not in the second mention context. Thus, similar to the overall phonetic reduction effects discussed above, the observed interactions between dialect variation and linguistic factors vary across vowel categories and the linguistic factors have different effects on dialect-specific variants within and across dialects. Although dialect variants have different social meanings and may therefore exhibit different patterns of variation across contexts, the variable interactions across linguistic factors suggest that the linguistic factors themselves may reflect different underlying processes that interact differently with dialect variation. As suggested above, the linguistic factors may represent different locations along an easy/hard processing continuum and dialect variation may interact with that continuum in a nonlinear way. Dialect variation is therefore another dimension that must be considered in further explorations of the hypothesis that all sources of phonetic reduction reflect the same underlying processing demands.

Several of our findings also suggest that there is variation in phonetic reduction processes across regional dialects. For example, we observed more temporal reduction of /i/ due to lexical neighborhood density for the Northern talkers than

for the Midland talkers, as well as more spectral reduction due to speaking style for the Northern talkers than for the Midland talkers. Additional evidence for dialect variation in reduction processes, including segmental alternations such as flapping and vowel reduction to schwa, comes from Byrd's (1994) study of the TIMIT corpus and Clopper and Smiljanic's (2015) study of variation in temporal organization in regional dialects of American English. In particular, American English dialects differ in speaking rate and pausing, but Clopper and Smiljanic (2015) observed additional effects of dialect variation on consonant and vowel timing that cannot be attributed to speaking rate variability. Clopper and Smiljanic (2015) hypothesized that this timing variability may be due to variation in reduction phenomena across dialects and provided some preliminary evidence that consonant cluster reduction and coda /t/ deletion and glottalization differ across dialects. We may therefore also expect phonetic vowel reduction to vary across dialects and other social categories, which may lead to further complex interactions among social and linguistic factors in phonetic reduction processes which are independent of the variability we have observed within and across linguistic factors, vowel categories, and acoustic domains.

2.4.3 Interactions between linguistic and cognitive factors

A third component of our recent research on phonetic reduction has explored the interactions between individual cognitive factors and the linguistic factors contributing to phonetic reduction. Within linguistics, the literature on the effects of individual cognitive differences on speech production is largely limited to developmental and clinical studies. However, a small but growing body of work is critically examining the role of individual differences in explaining variation in linguistic behaviors (see also Doherty et al.'s 2013, analysis of the role (or lack thereof) of variation in psychology research).

One recent study was conducted by Yu (2010), who examined individual differences in perceptual accommodation to coarticulation. Previous research demonstrated that listeners adjust their phoneme category boundaries in coarticulatory contexts (Beddor, Harnsberger, and Lindemann 2002). For example, when [s] is adjacent to [u], it has a lower centroid frequency, making it more [ʃ]-like. Listeners are aware of this coarticulatory pattern and are more likely to classify a sound that is ambiguous between [s] and [ʃ] as /s/ when in the context of [u]; that is, they perceptually accommodate the coarticulation (Mitterer 2006). Yu's (2010) study examined the role of autistic traits in neurotypical adults in this kind of perceptual accommodation to coarticulation. Autistic traits were assessed via the Autism-spectrum Quotient (AQ; Baron-Cohen et al. 2001), a short self-report

questionnaire designed to probe the extent to which someone's cognitive style mirrors that of a person with autism. The AQ was specifically designed to assess the dimensions of social skills, attention switching, communication, imagination, and attention to detail, with the notion that people with autism have deficits in the former four dimensions, and a surplus in the latter dimension. In particular, people with autism tend to exhibit strong attention to physical detail, while missing contextual or global cues (Happé and Frith 2006). The literature suggests that people with autism are less able to recognize global properties of speech, such as emotional content (Kleinman, Marciano, and Ault 2001) and regional dialect (Clopper, Rohrbeck, and Wagner 2012) than neurotypical individuals, and that proportionally more of their attention is devoted to acoustic detail over linguistic detail (Järvinen-Pasley, Pasley, and Heaton 2008). With this background in mind, Yu (2010) obtained the result that neurotypical adults with a greater prevalence of autistic traits in their personality (i.e., higher AQ scores) exhibited larger perceptual accommodation effects, while people with very few autistic traits (i.e., lower AQ scores) only accommodated to the coarticulation to a minor degree. This result is somewhat surprising, because rather than ignoring context and focusing on the acoustic signal alone, the participants with higher AQ scores (i.e., greater autistic traits) were instead paying more attention to the context and adjusting their perceptions accordingly. Nevertheless, this result has been replicated by Yu and Lee (2014) and Turnbull (2015a).

Yu (2010) explained these results in terms of an enhanced capacity to "systemize," that is, to create associations between objects and rules, in the participants with higher AQ scores. This capacity allows these individuals to keep track of contextually conditioned phonetic variation, such as coarticulation, which then allows them to perceptually accommodate the variation to a greater degree than other individuals. Yu's (2010) account also posits that these high-AQ individuals expend less cognitive effort on attention to social context and cues, which explains their relative deficits in attention switching and communication skills, and in turn means that these resources are freed up for attending to patterns of phonetic variation. This explanation is theoretically consistent with the mechanisms of perceptual accommodation to coarticulation outlined by Sonderegger and Yu (2010), although the main empirical claims are as yet untested.

Yu's (2010) finding of a relationship between patterns of perceptual accommodation and individual variation in cognitive style, as well as the findings from similar studies by Stewart and Ota (2008), naturally prompt the question of whether other linguistic phenomena are similarly influenced by such individual differences. To the extent that perception is mirrored in production (Beddor, Harnsberger, and Lindemann 2002; Casserly and Pisoni 2010; cf. Pardo 2012), and to the extent that processes of coarticulation are related to processes of

reduction (Deng, Yu, and Acero 2006; Moon and Lindblom 1994; Mooshammer and Geng 2008; cf. Browman and Goldstein 1992; Scarborough 2013), individual variation in cognitive style in general, and autistic traits in particular, may influence phonetic reduction. To explore this hypothesis, interactions between linguistic factors (lexical frequency, lexical neighborhood density, semantic predictability, and discourse mention) and individual AQ scores in phonetic reduction were examined in a series of studies by Turnbull (2015a, 2015b). The results demonstrate that talkers with higher AQ scores tended to have a larger difference between their word productions in semantically predictable versus unpredictable contexts, relative to talkers with lower AQ scores. This effect is depicted in the left panel of Figure 2.10 and is broadly consistent with Yu's (2010) "systemizing" account: the high-AQ talkers are able to determine the subtle systems and patterns within speech, such as noting the statistical trend for phonetic reduction in highly predictable contexts. This pattern is then reflected in their productions. The low-AQ talkers, on the other hand, do not notice the trend or only learn it inconsistently, leading to nonexistent or small reductions in highly predictable contexts. The modeling also revealed no significant interaction between AQ and discourse mention, as shown in the right panel of Figure 2.10: all participants, regardless of AQ, produced shorter words for second mentions than first mentions to the same degree (approximately a 25 ms reduction). This distinction between the effects of semantic predictability and discourse mention highlights their potentially different cognitive sources.

For lexical frequency and lexical neighborhood density, the statistical models revealed a third pattern. For these factors, participants with higher AQ scores were less affected by lexical frequency and lexical neighborhood density than

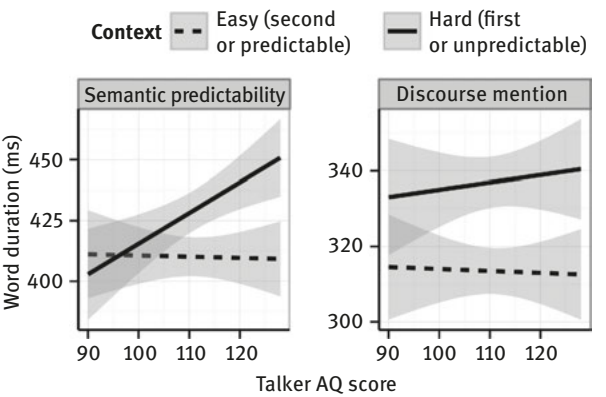

Figure 2.10: Effects of talker AQ score and semantic predictability (left) and discourse mention (right) on word duration. Adapted from Turnbull (2015a).

the lower AQ participants. That is, the acoustic differences – the magnitude of the phonetic reduction – between high- and low-frequency and -density words were smaller for the high-AQ participants than for the low-AQ participants. This result does not immediately appear to be consistent with Yu's (2010) account. However, these results are interpretable in light of the broader research on the autism phenotype. In particular, Stewart and Ota (2008) demonstrated that neurotypical individuals with higher AQ scores exhibit a weaker Ganong effect (Ganong 1980) than individuals with lower AQ scores, suggesting a weaker link between the perceptual system and the lexicon for higher AQ individuals. A weaker link to lexical knowledge could explain the smaller effect sizes for the lexical factors for the higher AQ participants in Turnbull's (2015a) study. Another possible explanation for these results involves an appeal to theory of mind, the ability to impute mental states to others. One of the components of the autism phenotype is proposed to be a weak theory of mind (Baron-Cohen, Leslie, and Frith 1985), and it is therefore possible that higher AQ individuals possess a less well-developed theory of mind than lower AQ individuals. Given a listener-oriented model of phonetic reduction, talkers must have a well-developed theory of mind to model their interlocutor's knowledge, because it is crucial for knowing when to reduce and when to speak clearly. Thus, weaker or more inconsistent phonetic reduction is an expected behavior of individuals with poorer theory of mind and, by extension, a high AQ score. However, this explanation fails to account for the observed interaction with semantic predictability or the lack of an interaction with discourse mention.

Thus, as in our exploration of dialect variation and phonetic reduction in the previous section, we observe considerable variability across linguistic factors in the relationship between cognitive factors and phonetic reduction processes, suggesting that a more nuanced understanding of the relationship between processing demands and phonetic reduction processes is warranted. In particular, the differences we observed across linguistic factors suggest that these factors may reflect different underlying cognitive processes. For example, although the concept of cognitive "accessibility" as a metric of processing difficulty is useful in accounting for both lexical frequency and discourse mention effects, because high-frequency and second mention words are more accessible than low-frequency and first mention words, these phenomena presumably rely on different kinds of accessibility – the former on lexical accessibility and the latter on discourse or referential accessibility. These different types of accessibility may exhibit different effects on processing in different contexts or exhibit different sensitivity to other cognitive or linguistic constraints, which individual differences research could help uncover. Given that the role of individual cognitive differences in speech processing in the neurotypical population is relatively poorly understood, our work in this area represents only a very preliminary step toward

unpacking the potential interactions in this domain, but our initial findings suggest that individual differences may be an important component to understanding phonetic reduction processes.

2.5 Conclusions

We propose that a more complex view of phonetic reduction processes is necessary to account for these observed patterns of variation. As suggested above, this complexification must minimally involve a gradient notion of processing difficulty combined with an allowance for nonlinear relationships between the linguistic factors, the processing difficulty continuum, and phonetic reduction processes. These nonlinear relationships could allow us to capture the apparent limits on phonetic reduction that are observed in some contexts, as well as the variation in the magnitude of phonetic reduction that is observed across acoustic domains and linguistic contexts. This complexification may also involve the differentiation of different kinds of processing demands, including the costs associated with accessing different kinds of linguistic information.

The necessary research to identify the nature of the processing demands that impact phonetic reduction is also likely to help distinguish among the talker-oriented, listener-oriented, and passive evolutionary approaches. Conceptually, all three accounts can be adapted to accommodate the proposed requirements for a gradient notion of processing difficulty that is nonlinearly related to both the linguistic factors and phonetic reduction processes and that differs across acoustic domains. From a listener-oriented perspective, the estimation of potential listener difficulty simply involves more complex computations of processing costs and the appropriate degree of phonetic reduction given the context. From a talker-oriented perspective, processing costs from different levels of representation (e.g., discourse and lexical) must be combined nonlinearly to drive the observed variation in production. From an evolutionary perspective, the exemplar space of potential production targets must be defined based on a large set of weighted contributing factors so that the selected production target reflects the nonlinear combination of the contextual effects that are experienced over time.

The general pattern of interactions between dialect variation and phonetic reduction can also be accommodated in any of the three approaches under the assumption that the dialect-specific variants are the truly native variants for the talker and are therefore easier for the talker to produce. In a listener-oriented account, the talker provides more dialect information by producing the easy, dialect-specific variants when the listener is likely to understand the message. That is, under easy processing conditions, talkers can afford to provide additional

information indexing social information about themselves. However, under more difficult processing conditions, talkers produce more effortful, standard variants in an attempt to make processing easier for the listener.[9] In a talker-oriented account, when processing is relatively easy, dialect-specific variants are activated most quickly because they are the native variants, but when processing is more difficult, more time is available to allow standard variants to be accessed. Similarly, in an evolutionary account, words that are easy to process can be produced and perceived with greater dialect variation and are therefore represented with more variable distributions than words that are harder to process. Thus, for example, high-frequency words will be represented not only by distributions containing more reduced forms but also by distributions containing more dialect-specific forms, leading to the selection of more extreme dialect-specific variants for high-frequency words than for low-frequency words in production.

Nevertheless, all three approaches also face challenges from some of the findings reported in the literature. The listener-oriented account is challenged by findings such as those obtained by Bard et al. (2000), which show that talkers do not always take the needs of their listeners into account. One proposed solution to this apparent problem for the listener-oriented account is to assume a simpler computation of listener need (e.g., Galati and Brennan's 2010, "one-bit" model of audience design), but this kind of simplification is clearly at odds with the evidence we have presented, suggesting the need for a more complex relationship between processing difficulty and phonetic reduction processes. In contrast, the talker-oriented account cannot easily accommodate the speaking style data, which reveal similar phonetic effects arising from explicit instructions about listener needs. That is, the nature of the speaking style manipulation is difficult to reconcile with the talker-oriented account. One obvious solution to this problem would be to treat speaking style as a distinct phenomenon that is separate from phonetic reduction processes, but the acoustic-phonetic realizations of the two phenomena are so similar that this solution seems to violate the goal

9 This account critically relies on the assumption that standard variants are more intelligible than nonstandard variants. Although standard varieties are more intelligible than nonstandard varieties, regardless of the listener's native dialect (e.g., Clopper and Bradlow 2008; Floccia et al. 2006; Sumner and Samuel 2009), nonstandard varieties are also highly intelligible to native speakers of those varieties (e.g., Floccia et al. 2006; Mason 1946; Sumner and Samuel 2009). Thus, the listener-oriented account may lead to different predictions depending on whether the talker and the listener share a dialect. In particular, when a nonstandard dialect is shared by the talker and the listener, dialect-specific information may be enhanced in difficult processing contexts to maximize intelligibility, contrary to the patterns observed in our data that were collected under conditions in which the dialect of the imagined interlocutor was unspecified.

of parsimony in theoretical accounts of speech production. Similarly, the evolutionary account was developed with a focus on lexical frequency effects. The extension of the model to other linguistic factors contributing to phonetic reduction therefore presents the most significant challenge to this approach. Whereas lexical frequency is straightforwardly represented in an exemplar model by the number of experienced tokens, the implementation of a model that can account for other lexical, discourse, and stylistic factors is less straightforward. Finally, Turnbull's (2015a) individual differences data present a challenge to all three approaches because they reveal different patterns of interaction between the cognitive AQ measure and phonetic reduction across linguistic factors, suggesting that different underlying cognitive processes are at play.

In the same way that different phenomena present challenges to the different approaches, some phenomena may be best accounted for by one of the three approaches. For example, the talker-oriented mechanism provides a strong account for discourse mention as in Bard et al.'s (2000) study, whereas evolutionary mechanisms provide a compelling account of lexical frequency as in Pierrehumbert's (2002) model, and a listener-oriented approach is the most obvious account of speaking style as an explicit adjustment in response to task instructions. These intuitions that different approaches provide compelling accounts of different results, together with the evidence for mixed results across linguistic factors and acoustic domains, have led some researchers to abandon a single account of phonetic reduction in favor of a hybrid approach. For example, Watson (2010) proposed a hybrid account in which temporal reduction reflects talker-oriented processing costs, but reduction in f0 reflects listener-oriented processing costs. Similarly, Turnbull (2015a) argued for a hybrid account of his individual cognitive differences data in which lexical effects on phonetic reduction reflect an exemplar lexicon as in the evolutionary perspective, but contextual effects on phonetic reduction reflect a talker-oriented model of the common ground. Although this kind of hybrid approach is less parsimonious than a single account of phonetic reduction, the complexity of the interactions among linguistic, social, and cognitive factors in the realization of phonetic reduction may ultimately require a model of multiple different processes across linguistic factors and/or acoustic domains.

The extent to which phonetic reduction processes are under conscious control is another area of investigation which may help distinguish among these approaches. For example, it is intuitively clear that some speaking style effects are controlled directly by the talker, whereas lexical frequency effects appear to be largely unconsciously controlled. However, care must be taken in the design and interpretation of such investigations, as research in social cognition suggests that a volitional action is not necessarily a consciously controlled action, and vice

versa (see, e.g., Dijksterhuis and Aarts 2010; Moors and De Houwer 2006). A more explicit understanding of the processing demands associated with the relevant linguistic contexts, potentially through careful individual differences research, may provide insight into the locus or loci of the phonetic reduction phenomenon.

Phonetic reduction must also be examined more carefully in interaction with other domains. Our research has revealed interactions with other linguistic factors (see also Gahl 2015, on segmental effects and lexical neighborhood density), with dialect variation (see also Hay, Jannedy, and Mendoza-Denton 1999; Munson 2007), and with individual cognitive factors. These factors all contribute to the phonetic realization of linguistic units and therefore cannot be completely controlled in any analysis of phonetic reduction. Segmental and prosodic structure have a substantial impact on word and vowel duration (de Jong 2004; Klatt 1976), as well as spectral vowel information (de Jong 1995; Fourakis 1991), adding considerable variability to comparisons across words (as in the lexical frequency and lexical neighborhood density analyses) or comparisons of the same words in different contexts (as in analyses involving spontaneous speech). Dialect variation has a substantial impact on spectral vowel information (Labov, Ash, and Boberg 2006), as well as prosody and timing (Clopper and Smiljanic 2011, 2015), adding variability to comparisons across talkers. Our individual differences research (Turnbull 2015a, 2015b) shows that the implementation and magnitude of phonetic reduction also vary across talkers within social groups, adding further variability to our data. Recent advances in automatic phonetic alignment and acoustic analysis, as well as more powerful statistical modeling tools, give us the opportunity to embrace these complex interactions in the search for a more complete understanding of phonetic reduction and its relationship to other speech processing phenomena.

Acknowledgments: This work was supported by the National Science Foundation (BCS-1056409) and a Presidential Fellowship from the Ohio State University Graduate School. We are grateful to Francesco Cangemi, Benjamin Munson, and two anonymous reviewers for comments on a previous draft.

References

Anderson, A. H., M. Bader, E. G. Bard, E. Boyle, G. Doherty, S. Garrod, et al. 1991. The HCRC map task corpus. *Language and Speech* 34. 351–366.

Arnold, J. E., J. M. Kahn & G. C. Pancani 2012. Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin and Review* 19. 505–512.