Natasha Warner
University of Arizona, Dept. Linguistics

Processes in Connected Speech

Cambridge Handbook of Phonetics

5.0 Abstract

For this chapter, connected speech will be defined as any speech in units larger than single words, including phenomena that happen at word boundaries even in careful speech, as well as phenomena of spontaneous or conversational speech. The former includes abstract phonological processes that are triggered by word boundaries (e.g. insertion of /r/ in some English dialects, as in *Australia[ɹ] is*), but that are accompanied by subphonemic, phonetic effects. Because this only has the opportunity to occur at word boundaries, it is by definition a connected speech phenomenon, even though it is often studied in very limited, careful speech materials.

The latter topic covers acoustic characteristics and perception of connected speech, regardless of word boundaries. For example, the vowel space appears to shrink in more connected and/or spontaneous speech, phonemically voiced stop consonants are often reduced to approximants, and segmental deletions and reduction in the number of syllables are common. The chapter offers examples of such connected speech reductions from the author's and collaborators' recordings. It is often difficult to believe that speakers really could produce phrases with the degree of reduction that one finds in casual, spontaneous speech. Even when listening to the recordings, one frequently fails to notice the reductions until one zooms in and examines individual syllables. Providing an array of examples (audio available online) may help to demonstrate the pervasiveness of reduction in connected speech.

5.1 Introduction

How is phonetics different in connected speech than in isolated words? One could begin by comparing two extreme cases: a careful speech production of nonsense syllables /ba, da, ga/ as separate utterances vs the speech from a casual conversation in Figure 5.1. Many things differ: the latter has various expected segments deleted and merged into each other, and it can be difficult to locate syllable or even word boundaries in it, or even to determine which words are deleted.

INSERT FIGURE 5.1 APPROXIMATELY HERE

First, it will be helpful to clarify some terminology, since many terms regarding speech style are used inconsistently in the literature. I take 'connected speech' to refer to speech that is produced in at least fluent sentence or utterance-sized units, not pronounced as isolated words. This is a generous definition, allowing even frame sentences like *Now I say X again* to count, marginally, as connected speech. This definition also includes speakers reading stories or scripts, as well as Map Task speech, interview speech, lectures, and conversation. That is, 'connected speech' covers a large part of the continuum from more careful 'lab' speech to spontaneous casual

conversation; it is not just at the extreme end of the continuum. The most surprising and least understood connected speech processes are common at the far end of the continuum, in spontaneous casual conversation between well-known interlocutors.

In order to discuss phonetic processes of connected speech, we need to know what phonetics as opposed to phonology is. For purposes of this paper, I will assume that phonological processes can reasonably be described by categorical insertion, deletion, or change of a segment to another. For example, r-insertion in some dialects of English (Cruttendon 2014: 317) in phrases like *Australia[ɹ] is* is a phonological process. On the other hand, if a segment is lengthened by some number of milliseconds in some environment, that is a phonetic phenomenon. This definition is not perfect, but it will be useful for this paper.

This chapter has two relatively separate foci: phonetic effects related to phonological patterns that are conditioned by word boundaries (Section 5.4.1), and ways that connected speech, including spontaneous and casual speech, differs overall from carefully pronounced speech (Sections 5.2 and 5.4.2). Section 5.3 furthermore provides an array of examples of reductions (e.g. Figure 5.1) from the author's and collaborators' recordings of spontaneous casual conversations between friends. In Figure 5.1, *can get* is pronounced as [hn̪ŋ] (to the author's best ability to transcribe it). Such reductions may be fleeting, unlikely to be produced with the same deletions and substitutions twice. Or they may be common: the author's research group has informally concluded from examining our recordings of spontaneous American English conversations that almost any sibilant noise that isn't obviously part of some other word is sufficient to be parsed as the word *just*. These massive reductions occur primarily in connected speech, but some degree of reduction is possible even in words read in isolation: the author's group has a recording of *treaties* read in isolation that is homophonous with *trees*. Within connected speech, reduction is not limited to casual speech either: the author recently heard a newscast in which the newscaster produced several tokens of *senator* with little trace of the medial /t/ (expected [ɾ]) other than perhaps some lengthening of vowels. It is often difficult to believe that so much reduction happens in normal speech, and that what we are hearing really has so many deleted and merged segments and syllables, so that the surface string does not remotely resemble the canonical form of a word (where 'canonical' means the pronunciation found in a pronouncing dictionary, with any regular phonological processes such as flapping in American English applied as appropriate). It is difficult to believe partly because we are so good at interpreting reduced speech strings as listeners, and this follows from the fact that reduced spontaneous speech is normal, and is the main means of human communication.

5.2 Historical Overview

Connected speech has, for the most part, only become a topic of phonetic investigation within the past 20 years. Before that, almost but not quite all phonetic research focused on carefully articulated target words, often in focus position of frame sentences. This focus partly reflects technological limitations, such as the difficulty until the 1990's of making large numbers of spectrograms quickly zoomed in and out to an array of time windows. However, some researchers did make early attempts to investigate speech in a more connected, even conversational, context.

Crystal and House, in a landmark series of papers, examine many aspects of segmental durations in connected speech scripts read by several speakers of American English (1982, 1988a, 1988b, and several related articles by the same authors). These papers provide a staggering array of reference data on durations of segments in English connected speech. This work is highly noteworthy for initiating large-scale study of any type of connected speech, while using detailed, laborious measurements. These studies do not reveal the extreme degrees of speech reduction found in some of the studies of spontaneous casual conversation below. However, they do point out that some things researchers believed based on isolated words of careful speech are not true, for example by finding that many stops do not have bursts as stops would be expected to.

Barry and Andreeva (2001) offer an unusually broad cross-linguistic qualitative survey of reduction in six languages (Russian, Bulgarian, Italian, Greek, Polish, and Czech), using quasi-spontaneous dialogues such as a map task. They examine the speech for instances of several types of consonant reduction, such as consonant cluster simplification, realization of stops as approximants, deletion of non-cluster consonants, etc. They find tokens of similar phenomena in all six languages. Arai (1999) provides similar descriptive data showing a variety of types of reduction (many of them the same ones) in Japanese.

An early and thorough study on vowels and speech style is Koopmans-van Beinum (1980). She examines a total of eight levels of carefulness, varying in speech style and stress (vowels produced as isolated segments, stressed vowels in words read in isolation, stressed vs unstressed vowels in a story-reading task, stressed vs unstressed vowels in the non-read speech of a story-retelling task, and stressed vs unstressed vowels in conversation). Koopmans-van Beinum shows that overall, the vowel space shrinks and all vowels centralize as the speech styles progress through the eight levels listed above. This work also includes a perception study examining how poorly native listeners recognize vowels excised from spontaneous speech. This represents very early work on perception of reduced speech.

Greenberg (1999) points out the massive reduction (see below for this term) that is common in casual spontaneous speech. He uses the Switchboard corpus (informal phone conversations) and compares transcription of what was actually produced to a dictionary listing of the canonical pronunciation. The speech in the corpus contains at least 80 pronunciations of the word *and* (all transcribed with distinct strings of phones). The four most common pronunciations found for *and* are [æn, ɛn, ɨn, ʌn] in that order, followed by [n̩] and [n]. Uncommon realizations include [ɨ, ɚ] (a single reduced quality vowel). This work also offers a listing of the most common pronunciation variant for each of the 100 most frequent words in the corpus. For example, the most common pronunciation for *that* is [ðæ] (without final consonant) and the most common for *them* is [ʌm] (without initial consonant). Broad-based listings such as this are eye-opening. Greenberg also demonstrates a much greater preponderance of 1- and 2- syllable words in the Switchboard corpus relative to the lexicon. Furthermore, many words which have consonant clusters in their lexicon listing are produced with simpler CVC or CV syllables in the Switchboard recordings, bringing English closer to maximal CVC syllable structure in its phonetic realizations.

Arai (1999) provides an early investigation of listeners' perception of highly reduced speech. Arai excises a string of highly reduced speech (the underlined portion) from a larger context

(/hayaku ieba hoka ni/ 'in short, other' and presents it to listeners either alone or with various amounts of the surrounding context included. Listeners transcribe what they hear in Japanese orthography, which facilitates counting the number of moras perceived. The underlined portion (the target string) consists of five moras in its canonical lexical form. When listeners hear only the target string, they perceive approximately two moras (typically /kebo/). When they hear the target with any surrounding context, even just /aku ieba hoka/, they perceive the target as containing an average of almost five moras. This small amount of context does not make the surrounding words clear, and yet it was enough to greatly improve perception of the reduced target portion. This is a one-item experiment, but it demonstrates how much even a small amount of context can help listeners with recovering information from reduced speech. Pollack and Pickett (1964), in a very early study of perception of fast, slow, conversational, and read speech, find that perception of a single word out of context is much weaker than in context, especially for conversational speech.

Mehta and Cutler (1988) use phoneme monitoring in spontaneous vs read speech and find that the coherent intonation of planned speech facilitates processing, relative to the pauses and repairs of spontaneous speech. The focus of this work is on listeners' processing while hearing speech, rather than on what string they ultimately perceive.

5.3 Critical Issues

This section will give examples of connected speech from the author's recordings. This serves as a way to introduce the most critical issue of connected speech: the surprising degree of reduction and variability often found in it. These are the largely unpredictable reductions that happen in connected speech, especially in casual, spontaneous, conversational speech. Audio is available at nwarner.faculty.arizona.edu as well as in the resources associated with this publication.

(1) Examples of spontaneous speech reductions. Parentheses around a segment indicate that it is not clear whether the segment was realized or not, but one cannot rule out that there may be slight acoustic evidence for it. The longer context of the string is given below each example. Spaces are used for word boundaries where they are identifiable, solely for ease of reading. Spaces do not indicate pause. Portions with especially extreme reduction are bolded.

	Realized as	Expected speech string	Phrase
a.	[hn̥ŋ]	[k^{h}ãn gɪt]	**'can get'**

'I think I can get down to a 445 this summer' (swimming) (see Figure 5.1)

b. [əʒ lʌ̟i̟k] [aɪ wʌz (dʒʌst) laɪk] **'I was (just)** like'
'I was (just) like, yeah, right' (continuation of example (a), some listeners hear 'just')

c. [bi̟ ʌ ʒz lʌ̟i̟k] [bʌɾ aɪ wʌz (dʒʌst) laɪk] **'but I was (just)** like'
'but I was (just) like, why wouldn't you just go home?' (Some listeners hear 'just')

d. [p^{h}ʌ i̟ z̥ lʌɪk] [bʌɾ aɪ wʌz laɪk] **'but I was** like'
'...but I was like, oh, no, so he has to cut someone, 'cuz he already told Steve he was in the wedding'.

e. [(l)ɨʊɨ kʰpʰyɾɚ] [ə lɪɾl̩ kʰə̃mpʰjuɾɚ] **'a little computer'**
'converted into, I guess like a little computer room'

f. [ʃɨ z ɪ vl̩lis ɔvɨsɚ] [ʃi wɑnts tu bi ə pəˈlis ˈɔfəsɚ] 'she **wants to be a** police officer'
'She wants to be a police officer, I think.' (Figure 5.2)

g. [ʌ̃ɛ̃ːɚ ɹl̩ĩ noʊ ðæːt̃][aɪ ˈdɪdənt ˈrɪli noʊ ðæt] **'and I didn't** really know that' (Figure 5.3)
'And like, one of my classes is, like, a writing-intensive class, and I didn't really know that, but I need to take it to graduate.'

h. [mʌ fɹɛ̃ː ɯ̟ʊs sʊz sːtjʊpg̠d] [maɪ ˈfraɪdeɪ naɪt wʌz soʊ stupɨd]
'My **Friday night** was so stupid'

i. Dutch:
[m̩sxɨ̃n(d)ɪk(ə)m] [mɪsxĩn dɑt ɪk ə̃m] 'misschien **dat ik** hem' 'maybe that I it'
'dacht ik van misschien dat ik hem wel daar gewoon kan kopen'
'I thought that maybe I can just buy it there after all.'

j. Japanese:
[soiçəʃjono] [so: ju: ç̟i̥totatʃi jo: no] 'soo **iu hitotachi** yoo-no' (Figure 5.4)
'a, soo iu hitotachi yoo-no supeesu-ga aru-no-ka?'
あ、そういう人たち用のスペースがあるのか
'Oh, is there space for people like that?'

INSERT FIGURES 5.2-5.4 APPROXIMATELY HERE

These examples bring to the fore many interesting points about connected speech. If one were to ask a native English speaker or a phonetician what the possible realizations of *can get* are, it is unlikely that anyone would think of [hn̩ŋ̩] (Figure 5.1). In fact, the author often finds that readers do not believe reductions like this are plausible until they hear the sound file. One might have to record an extremely large amount of speech before encountering a similar realization again.

On the other hand, examples (b)-(d) show a surprising degree of similarity across three reduced tokens of *I was like* or *I was just like* used as a quotative. Some of these tokens are ambiguous as to whether there is a *just* present or not. The author has sometimes not heard *just* herself, but then on playing the example to a class of young listeners of this dialect, the majority of the class is certain there is a *just*. All three tokens shown here have in common that *I was (just)* is realized as a central vowel followed by some type of voiced sibilant, and *like* is realized with a perceptible [l], a centralized version of the /aɪ/ diphthong that may have creaky voice, and a [k]. This is a surprising degree of consistency despite extreme reduction. Furthermore, both (c) and (d) have been reduced from four or possibly five syllables to two by reducing *but* to just an onset consonant, merging its vowel with the following *I was*, and deleting the final /t/ instead of producing a flap. The various phrases *(But) I was (just) like* are very high frequency for speakers of this dialect. One can imagine that a high frequency sequence with a relatively consistent

extremely reduced production might eventually become lexicalized as the reduced form, as *prolly* seems now to have been lexicalized from *probably*.

Example (e) contains extreme reduction of *a little*. The reduction of number of syllables is not unusual: the recording sounds like it might be two syllables, a reduction of just one. This speech string is unusually difficult to transcribe phonetically and unusually difficult to understand out of context. It gives a percept of lateralization, possibly throughout the string, and there are non-front vowels. Beyond that, it is very difficult to find identifiable segments or features. The transcription given above may not fully reflect how reduced this recording is.

Reduced speech tokens frequently contain segment strings that are phonotactically illegal in the careful speech of the language. Here *computer* is reduced to the point of having [k^hp^h] as an onset cluster. Example (a) above also contains a highly illegal phonotactic string, where it is unclear how many nasals in a row are even present. It really should not be surprising that phonetic spontaneous speech reduction does not respect phonotactic rules that apply to the lexicon and careful speech. What is probably more important is perceptual cues to a given feature: often one perceives substantial nasalization throughout a string, but cannot say how many nasal consonants (if any) are present. The *a little* case in (e) has a percept of something lateral without a clear [l] consonant present. Niebuhr and Kohler (2011, below) point out that a long nasalized palatalized stretch is sufficient for listeners to perceive *eigentlich* 'actually' in German, and although there are few recognizable segments and no consonants, the nasalized palatalized stretch conveys substantial information. This use of a pervasive feature such as nasalization, palatalization, laterality, or rhoticity without clear segmental units makes it clear that phonotactic legality of segment strings is really not the question in this sort of speech.

In example (f) (Figure 5.2), the entire string *wants to be a* seems to be reduced to [zɪ], and yet this utterance is not difficult to understand. Clearly, there is no way that phonological rules can generate reductions like these unless one assumes optional phonological rules that are allowed to delete any segment. Sibilant fricatives do seem to be retained often even in the midst of extreme reductions like this or the *was (just) like* examples above. It is possible that sibilants are especially useful for conveying information, either for lexical reasons or their acoustic distinctiveness from the rest of the mostly sonorant speech. In (g) (Figure 5.3) by contrast, the sequence *I didn't* is reduced to an entirely vocalic string, for which the author's best attempt at transcription is [ɐ̃ɛ̃:ɚ̃]. In this type of reduction, giving a phonetic transcription forces one to choose categorical symbols for the continuous speech signal, even though the production may really not be recognizable as containing segments of the phonological inventory of the language.

Example (h) shows that extreme reduction is not limited to function words or very high frequency content words. Here the entire word *night* seems to be deleted acoustically except for some lengthening of the previous material, and *Friday* seems to have lost a syllable as well. The single long vocalic stretch of *Friday (night)* is quite difficult to transcribe. It has a small amplitude dip in the middle, suggesting a reduced consonant, but it is not recognizable as any consonant symbol of the IPA. The entire vocalic portion, when listened to in isolation, can be perceived as a long [ɛ:].

Examples (i) and (j), both from recordings made by the author's research team and checked by native speakers, confirm that such reduction in spontaneous speech occurs in other languages (cf Barry & Andreeva 2001, above). As far as the author knows, every language that someone has looked for these reductions in contains them. In example (i) in Dutch, the segments transcribed in parentheses have minimal evidence for them. The sequence *dat ik hem* has been reduced by one syllable and several segments. In the more extreme Japanese example (j) (Figure 5.4), vowel qualities have been deleted or merged with neighbouring vowels, multiple consonants and syllables have been deleted, and it is rather difficult to determine what sounds are present and how they map onto the words at all. The only realization of the plural suffix [-tatʃi] seems to be sibilant noise [ʃ] and a [j] that also forms part of the next word. This example shows that even in Japanese, with its almost completely CV syllable structure, reduction does not obey phonotactic rules: [ʃj] is not a possible syllable onset in Japanese, and [ʃ] is not a possible coda unless it is part of a geminate. Ernestus and Warner (2011) provide examples of extreme reductions in several more languages.

What do such examples show about the phonetics of connected speech? First, extreme reductions happen both within a word and across word boundaries. Reduction across word boundary can obscure word boundaries as in (f, g) or delete complete words. Sometimes one can clearly understand the speech but cannot even identify which words have been deleted (f): in the string *wants to be a*, which word is represented by [z] and which by [ɪ]? What these examples show us is that the normal way for humans to communicate information is through reduced speech, with all its unpredictabilities and messiness. This leads directly to wondering how listeners perceive sounds and words and extract meaning from such highly reduced speech.

5.4 Recent Research

Connected speech can be divided into processes that happen specifically at word boundaries, which therefore cannot happen in non-connected speech by definition, and processes that happen more often in connected, spontaneous, or casual speech. The latter are not caused specifically by word boundaries.

5.4.1 Phonetics of Word-Boundary-Triggered Phonological Processes

This section examines several phonological processes that are triggered by word boundaries (thus occurring in connected speech), but that have phonetic, subphonemic aspects to them. Many of the studies discussed here use stimuli consisting of just two words or a short sentence produced very carefully. Thus, this type of speech falls just barely into the range of 'connected speech'.

5.4.1.1 /r/-Insertion or /r/-Linking

One such phenomenon is /r/-insertion or linking in some dialects of English in phrases like *Australia/r/ is*. (For convenience I will write it henceforth as /r/ rather than [ɹ], but this is not meant as a theoretical claim. I will refer to this as insertion, but other analyses are possible, and nothing in this discussion depends on the phonological analysis.) Cruttendon (2014:317) describes the phenomenon of adding an /r/ that is not historically present at word boundary

between two vowels. He gives examples, and states that the inserted or linking /r/ is shorter than word-onset /r/, so that the /r/ in *more ice* is shorter than the one in *more rice*. (Crucially, the final /r/ of *more* in *more rice* is deleted because of being syllable-final and pre-consonantal in the dialects with this pattern, so neither phrase is expected to contain a sequence of two /r/ sounds. The /r/ in *saw ice* would also be predicted to be shorter than the one in *saw rice*.) Tuinman et al. (2011) verify with acoustic measurements that the historically inserted /r/ (as in *Emma/r/ ejected*) is shorter and has less amplitude dip than the historically present /r/ (*Emma rejected*) in the speaker who produced their stimuli. Thus, the inserted /r/ is phonetically weaker, and closer to being vocalic. Tuinman et al. (2011) then use perception experiments to show that at least native listeners of an /r/-insertion dialect of English are sensitive to this and can use the duration to help determine whether they are hearing an inserted vs underlying /r/. They show that this holds true both for cases where the /r/ is inserted (*saw ice* vs *saw rice*) and cases where it is historically present but only retained when pre-vocalic (*more ice* vs *more rice*). They also find that native speakers of a different dialect of English (American English non-/r/-inserting dialects) and even non-native speakers can use duration as a cue to this difference, but to a lesser extent. To summarize, /r/-insertion is a phonological phenomenon that applies at word boundaries (and thus in connected speech). This set of research has shown that there are phonetic effects on the production of this phonological connected speech phenomenon, and that listeners are sensitive to these phonetic effects in perception.

5.4.1.2 French Liaison

One type of French liaison offers a very similar case to English r-insertion, but with a consonant that is clearly underlyingly present and sometimes deleted, rather than one that is historically absent but inserted. Spinelli et al. (2003) discuss phrases like *dernier oignon* 'last onion' vs *dernier rognon* 'last kidney'. Both are pronounced with an [ʁ] at the word boundary, but in the former phrase, it is underlyingly the end of the first word and is resyllabified to the onset of the next syllable, whereas in the latter phrase the [ʁ] is underlyingly the onset of the next syllable. Parallel to the r-insertion English case of *more rice*, if the [ʁ] remains in coda position without being resyllabified it is deleted, so *dernier rognon* does not contain a double [ʁ]. Spinelli et al. show that there are small differences in duration between the two types of [ʁ], with the underlyingly word-final one in *dernier oignon* being shorter by 7-12 ms. Furthermore, Spinelli et al. show that listeners are able to use this difference in duration to help them disambiguate words like *oignon* vs. *rognon*. Spinelli et al. (2002) show this with different consonants, and give further evidence based on splicing that this durational difference is the perceptual cue.

5.4.1.3 Word-Final Devoicing

Incomplete neutralization of final devoicing provides another case. Famously, words such as Dutch *meet* 'measures (1st sg. present)' and *meed* 'avoided' (1st sg. imperfective) sound nearly identical and [t]-like, although they have underlying forms that end in /t/ vs /d/. If the stop is intervocalic, the distinction becomes clear, as in *meten* 'to measure (inf.)' and *meden* 'avoided' (1st pl. imperfective), where the underlying /t/ and /d/ are pronounced as [t] and [d] respectively. What makes this incomplete neutralization is the small difference between forms like *meet* and *meed,* in the direction that would be expected for [t] vs. [d], but on a much smaller order. For Dutch, Warner et al. (2004) find that the vowel preceding the final consonant is 3.5 ms longer

before underlying /d/ (*meed*) than before underlying /t/ (*meet*). Incomplete neutralization is triggered by the phonological environment and appears categorical until one takes a large number of highly controlled measurements, but those measurements reveal a phonetic effect. This pattern is in some sense a connected speech phenomenon because it happens largely at word boundaries. However, incomplete neutralization is usually studied in careful lab speech, even in words produced in isolation, because the effects are so small that adding the variability of any less controlled speech may overwhelm the small effect size. This phenomenon of incomplete neutralization at word boundaries has been found for Dutch (Warner et al. 2004), Russian (Kharlamov 2014, 2015), German, Catalan, and Polish (see Warner et al. 2004 for literature review). Kharlamov offers an exceptionally thorough investigation of the effects of minimal pair status and reading task. Kharlamov (2014) finds that speakers are more likely to produce the incomplete neutralization effect in words that form minimal pairs than in words that do not, and more likely to when reading words in a list than when producing speech in some other way. Kharlamov (2015) finds related patterns in perception. Warner et al. (2006) argue that the effect in Dutch may be largely an effect of orthography rather than underlying form. This suggests that the effect, tiny even in the strongest cases, is probably even smaller or absent in spontaneous speech, further out on the continuum of connected speech.

5.4.1.4 More General Acoustic Cues to Word Boundaries

Mattys and Melhorn (2007) investigate how listeners weight cues to word boundaries. They examine cues in sentential meaning, the lexicon, and acoustics. For example, *The baker looked at the drawing of a plum pie/plump eye* is strongly biased toward the *plum pie* interpretation, while *The surgeon looked at the drawing of the plum pie/plump eye* is biased the opposite direction. Both can be pronounced with the word-final /p/ giving acoustic cues that suggest a word-final or a word-onset. This is a more general case of the phenomena described above, since it does not depend on a specific phonological pattern. Mattys and Melhorn created stimuli with acoustic cues that strongly or mildly favoured each parse, and paired these productions with various preceding sentences and words/non-words. Listeners responded with whether they heard *pie* or *eye*. Mattys and Melhorn find that listeners rely more heavily on sentential meaning and the lexicon if acoustic cues to presence or absence of a word boundary are weak.

5.4.1.5 Across-Word-Boundary Regressive Assimilation

It is well known that consonants frequently assimilate to the place of a following consonant. Since many consonant clusters within a word have already assimilated as a historical process (e.g. /m/ in *impossible*) or are homorganic with the following consonant anyway (e.g. /n/ in *can't*), regressive place assimilation is most noteworthy when it occurs across word boundaries, as in *late calls* pronounced with [k] but no [t] or *ten boys* pronounced with [m] rather than [n] (Browman & Goldstein 1992). Browman and Goldstein (1992) and related work show that this process is gradient and phonetic rather than categorical and phonological, since articulatory traces of the underlying consonant remain even if they cannot be heard. Thus this assimilation is itself a phonetic connected speech process.

Dilley and Pitt (2007) and Zimmerer et al. (2009) both investigate how often such across-word-boundary place assimilation happens in connected speech. The former study uses a semi-

spontaneous appointment scheduling task, similar to a map task in carefulness of the speech. The latter uses the spontaneous conversations of the Buckeye corpus, giving examples such as *gree[m] boat* for *green boat*. Zimmerer et al. (2009) confirm that alveolar consonants are more likely to undergo assimilation than others, and find that final consonants of function words more often undergo assimilation than those of content words. Dilley and Pitt (2007) examine the corpus for how often such across-word-boundary place assimilation occurs, and find that it is a relatively infrequent phenomenon even in appropriate environments.

A long series of psycholinguistic studies have investigated how listeners recognize words that have undergone such assimilation. Mitterer and Blomert (2003) provides one example, investigating Dutch listeners' processing of words such as *tuin* 'garden' pronounced with final [n] vs [m] in a phrase before *bank* 'bench' (which can trigger assimilation) vs *stoel* 'chair' (which cannot). Overall, this stream of research finds that listeners are able to process this type of variation successfully as long as it is realized in a viable environment.

5.4.2 Phonetic Phenomena of Connected Speech not Specifically at Word Boundaries

We turn now to aspects of connected speech that are not specific to word boundaries. If a phonetic effect occurs in spontaneous conversational speech and does not usually occur in words pronounced in isolation, we may consider it a connected speech phenomenon even though it is triggered by speech style rather than by words being concatenated. The literature is too rich to cite every paper studying or using connected speech here, so this section will give a sampling of works on several topics.

5.4.2.1 Realizations of Consonants in Connected Speech

Warner and Tucker (2011) provide a large study of intervocalic consonants before unstressed vowels in American English, comparing words read in isolation, story-reading, and spontaneous casual conversation between a speaker and a well-known interlocutor. They show that when /b, g/ or flap is expected in a word, a large proportion of tokens actually have an approximant, a consonant with no burst, or even deletion of the consonant. Even /p, k/ show fully voiced, approximant-like, or deleted tokens far more often than one might expect for phonemic voiceless stops. Flap shows the most reduction (most approximant-like or deleted tokens), followed by /b, g/, with /p, k/ remaining the least reduced. Warner and Tucker find that as expected, these intervocalic consonants are more reduced (more approximant-like, more often voiced, more often burstless, or more often deleted) in spontaneous speech than in isolated word productions. They find that story-reading is intermediate to the other two speech styles on most measures in most conditions, but when it groups with one of the other styles, it usually groups with conversation rather than with isolated word reading, suggesting that the connectedness of the speech is more important for degree of reduction than the spontaneous nature of the speech planning. Bouavichith and Davidson (2013) show related results on intervocalic phonemically voiced stops, comparing pre-stress vs pre-unstressed environments. (Thus, syllabic affiliation of the consonant may differ depending on one's analysis of potentially ambisyllabic consonants, as in 'judo' vs 'adept'.) They use speech from a story-reading task. They find far more reduction in pre-unstressed stops than in pre-stress ones, and confirm the previous finding that before an

unstressed syllable, reduction of phonemically voiced stops to something resembling an approximant is not unusual.

Davidson (2011) offers a detailed acoustic examination of the releases (or lack thereof) of stops in clusters. She examines the stop in the first position of word-medial clusters (e.g. *oatmeal, napkin*), in word-final phrase-medial clusters (e.g. *up North*), and in phrase-final position (e.g. *I saw her jump rope.*) The stops were followed by either another stop or a nasal. Davidson uses monologue and interview data from radio broadcasts (NPR's Story Corps recordings), as well as a more structured picture description task that allowed a more balanced collection of stops in various conditions. Overall, this study finds that pre-consonantal stops are released relatively often, even though stops in this environment might be expected not to have a burst. By far, the most common realizations in most environments are stop with burst or stop without burst (unreleased). Only a relatively small proportion of the stops are deleted, realized as glottalization, or lenited to an approximant-like form.

Davidson (2016) investigates possible realizations of voicing during stop and fricative constrictions, using American English speech from a story-reading task. She examines only the segments that are phonemically voiced (/b, d, g/ and voiced fricatives). She classifies each token as fully voiceless (meaning less than 10% of the duration is voiced, to prevent classifying a single pulse of voicing as a partially voiced closure), fully voiced (more than 90% of duration voiced), or partially voiced. At least half of the phonemically voiced stop and fricative realizations are only partially voiced, averaged over environments. This work goes into more detail by examining the shape of voicing over time in the partially voiced tokens, classifying them by whether voicing bleeds into the consonant from the preceding sound and then stops (the majority of stops), begins during the consonant (prevoicing), begins and ends during the consonantal constriction (a rare realization), etc. Davidson (2018) gives the comparable results for phonemically voiceless stops and fricatives, showing that the most common outcome for both is partial voicing with the voicing bleeding into the constriction from the preceding segment and then stopping. Barry and Andreeva (2001) and Arai (1999, production portion), discussed above, are more broadly related to this topic.

5.4.2.2 Realization of Vowels in Connected Speech

Several studies have examined vowel space for particular languages in careful vs connected or spontaneous speech, such as Koopmans van Beinum (1980, discussed above). Kuo and Weismer (2016) forms a recent comparison study on American English, showing reduction of the vowel space in conversational speech (conversations with the experimenter) relative to more careful speech, and especially relative to intentionally 'clear speech' (see Section 5.4.2.5 below).

Gahl et al. (2012) examine specifically words with dense lexical neighbourhoods (e.g. *cat*, which has many neighbours such as *mat, bat, sat, cap, catch, coot, coat*). Gahl and collaborators find that in the spontaneous speech of the Buckeye corpus, these words are produced with reduction of the vowel space as well as reduced duration. Kiefte and Nearey (2017) analyse one dialect of Canadian English for carefully read vowels in a controlled *hVd* environment vs vowels in spontaneous monologue speech. They fit a model of change over the course of time to each vowel, and compare the trajectories over time of spontaneous vs careful read vowels.

Vasilyeva et al. (2016) provide a comparison in an underdocumented language by studying the vowels of Yakut in read vs spontaneous speech. Their focus is on phonemic vowel length rather than speech style, and this work does not include measurements of the vowel space, but it does show that phonemically long vowels are longer than phonemically short vowels in both read and spontaneous speech, and that there are other correlates of phonemic vowel length in addition to duration in both speech styles.

5.4.2.3 Studies of Specific Segmental Reduction Processes

Many studies examine a specific reduction process in a specific language. For example, Bürki et al. (2011) examine schwa deletion in French words such as *fenêtre* 'window', where the first vowel can be deleted. Bürki et al. examine connected speech from radio broadcasts. They find that in this speech style, schwas in the relevant environment are deleted in more than half the tokens. They also find that the factors that predict whether the schwa will be deleted are not the same as the factors that predict the duration of a schwa when it is present, and they conclude from this that the deletion is a categorical phonological process rather than the extreme end of a gradient shortening process. Brand and Ernestus (2017) examine listeners' perception and acquisition of these French forms with the schwa present vs absent. They find that native French listeners' ease of recognition of words with and without schwa correlates with how frequently that variant of the given word is used. Native Dutch speakers who are advanced learners of French show the same pattern, but based on their own estimates of variant frequency rather than native listeners' estimates. This suggests that both native and advanced learner listeners' word recognition is influenced by how often they have heard a given variant form.

Hanique et al. (2013) use similar methods to study reduction and deletion of two segments in Dutch spontaneous speech, schwa and /t/. They study both segments in past participles, such as *gemaakt* /xəmaːkt/ 'made'. Both types of reduction and deletion occur in a wide variety of words, not just in a few function words where the reduction might be lexicalized. They find that the factors that favour shortened duration (their measure of reduction) and those that favour deletion are the same for /t/, while the factors favouring these two outcomes are different for schwa. This shows that reduction in spontaneous speech can be gradient, or can have a categorical and a gradient component.

Many researchers have investigated other specific segmental reduction phenomena in a handful of languages. A large proportion of the research on this topic has been carried out in the Netherlands, leading perhaps to a focus on Western European languages. These studies, taken together, have provided a rich basis from which to understand reduction in connected or spontaneous speech (Mitterer & Ernestus 2006, Torreira et al. 2010, Ernestus et al. 2017, and many other publications by these authors, their collaborators, and others).

5.4.2.4 Syllable Deletions and Other 'Massive' Reductions

Spontaneous speech also contains many cases of reduction spanning a long string of consonants and vowels, rather than affecting a single sound. This may result in a percept of fewer syllables than words would be expected to have. This is sometimes termed 'massive reduction' (Johnson

2004). Greenberg (1999), discussed above, is important for bringing the degree of variability caused by massive reduction to the attention of phoneticians. Johnson gives examples of *particular* realized as [pʰtʰɪkʰɚ] and *apparently* realized as [pʰɛɾĩ]. Johnson (2004) uses phonetic transcriptions of the Buckeye Corpus (spontaneous interview/conversational speech with an interviewer), and compares the phonetic transcription of the actual production to the dictionary listing of the canonical form to determine whether there are deletions. This study finds that 6% of all canonical syllables in content words and 4.5% of those in function words are deleted in this type of spontaneous speech. Syllable deletion is more common in longer words: words with a canonical form with three or more syllables showed a syllable deletion in at least 20% of cases. Furthermore, content words with canonical pronunciations having 4 or more phones showed deletion of 20-25 percent of their phones. Four phones is not an especially long content word, so this is quite a high rate of deletion: almost a quarter of all segments in the dictionary pronunciation for these words are deleted in connected speech. Additional examples from the author's own work appear in Section 5.3 above.

Liu et al. (2016) discuss syllable count reductions in Mandarin Chinese connected speech. They examine production of Mandarin words with a disyllabic canonical form. They classify the productions by whether they maintain both syllables, show various types of intermediate reduction, or collapse the form to a single syllable (which they find to be especially likely for high frequency words).

Several articles provide general introductions to the topic of reduced speech. Ernestus and Warner (2011), Warner (2011), and Warner (2012) offer recent reviews of some representative literature, reduction processes, and methods for studying connected, spontaneous, and reduced speech.

Various researchers have investigated how listeners perceive speech and recognize words with massive reduction. Niebuhr and Kohler (2011) point out that the German phrase *eigentlich eine (noun)* 'actually, a (noun)' can sometimes be reduced so drastically that it is realized as [aĩĩ], homophonous with *eine (noun)* 'a (noun)'. They construct a perception experiment in which the duration of such a vocalic, nasalized, palatal-ending stretch is varied. They show that when the palatal (high front vowel) portion of this phonetic string is lengthened, listeners are more likely to perceive the stimulus as the longer phrase *eigentlich eine* (noun). That is, a long duration of nasalized palatalized vowel-like material is sufficient to serve as a perceptual cue to the entire word *eigentlich* 'actually', which has three syllables in its canonical pronunciation. While this study examines a single item pair, it does point out a striking feature of highly reduced spontaneous speech: one often finds long stretches of the waveform that convey a surprising amount of perceptual information, that seem to consist only of nasalized vocalic material, which one might only be able to identify as for example a non-back non-low vowel quality with some change in F2 (cf. Section 5.3). Listeners are apparently able to extract considerable lexical information from such time-varying vocalic stretches.

Ernestus et al. (2002) present related findings to the early Arai (1999) work discussed above, but use a more generalizable larger set of items and stimuli. They choose tokens of high frequency words and phrases such as *natuurlijk* 'naturally', *mogelijk* 'possible', and *op een gegeven moment* 'at a certain time' from a corpus of spontaneous Dutch speech. They classify each token based on

its transcription into a high-, medium-, or low-reduction group. Examples of high-reduction pronunciations for these three items include [tyk], [mɔk], and [ɔpxef] respectively. Ernestus et al. present these items to listeners for identification with or without context. They find that when presented in isolation, the high-reduction tokens are incorrectly perceived approximately half the time, and even the low- and medium-reduction tokens are misperceived 10-20% of the time. However, with context, error rates drop to below 10% even for high-reduction tokens. Brouwer et al. (2012) extend this topic to examine the timing of listeners' processing of the speaker's intended word vs lexical competitors. Eye-tracking data for listeners hearing reduced vs. unreduced speech forms allows them to examine activation of lexical competitors when listeners hear speech where reductions have created ambiguities (e.g. if a listener hears a token of *computer* pronounced as *puter*, does the listener initially activate *pupil* as a competitor, and at what point is that incorrect competitor ruled out?). The results suggest that listeners are more open to interpreting reduced forms as the longer word they are reduced from (i.e. more likely to look at 'computer' when hearing 'pu-') when the stimuli are in a connected speech context.

5.4.2.5 Clear Speech

At the opposite end of a continuum from reduced speech falls 'clear speech', meaning the extra-clear speech speakers produce when they know their interlocutor has hearing loss or does not speak the language fluently, and they want to make their speech easy to understand. Bradlow and colleagues have contributed particularly strongly to our knowledge of perception of clear speech, especially regarding how clear speech interacts with bilingualism and how it is realized in various languages. For example, Smiljanić and Bradlow (2005) show that both Croatian and English speakers produce some of the same acoustic characteristics in their clear speech (in their respective native languages), including expanding the vowel space similarly, even though the languages differ in how crowded the vowel space is. Furthermore, Croatian and English listeners show a similar benefit to clear speech in perception. Smiljanić and Bradlow (2011) and Bradlow and Bent (2002) examine clear speech with L2 speakers and listeners. In most such studies, speakers are instructed to read the sentences once using a 'conversational' style, and once as if they were talking to someone with hearing loss or a non-native speaker. That is, most studies of clear speech do not use speech actually produced in spontaneous conversation, and instead use exclusively read speech, although the non-clear condition is often referred to as 'conversational speech'. In all of these studies and others on clear speech, the purpose is not to examine what makes connected speech different from non-connected speech. Rather, clear speech and regular speech are both examined in sentence context, meaning the topic is normally only investigated using connected read speech.

5.4.2.6 Topics that can Only be Investigated in Connected Speech

Many works that use connected speech, spontaneous speech, or conversational speech as material are not actually studying connected speech for its own sake, but rather are trying to answer questions that can only be investigated in contexts beyond the word or in natural utterances. For example, one cannot study intonation or prosody without using full utterances. Most studies of intonation use sentence-level careful read speech (e.g. Pierrehumbert & Beckman 1988), while some use a corpus of spontaneous speech (e.g. Maekawa et al. 2002), but practically all use connected speech. Studies of rhythm/timing class similarly require at least

utterance-level speech for the phenomenon to exist (e.g. Nazzi et al. 1998, Warner & Arai 2001, Grabe & Low 2002, White et al. 2012). Most studies of rhythm use careful readings of constructed sentences, whereas Warner and Arai use spontaneous monologue speech. Studies of pausing, laughter in speech, and self-corrections are also limited to connected and generally even spontaneous speech (e.g. Nakatani & Hirschberg 1994, Shriberg 2001, Kouwenhoven et al. 2015). Another topic that is sometimes investigated using spontaneous speech out of necessity is language production by very young children (Lowenstein & Nittrouer 2008, for just one sample), where one method involves analysing whatever speech the child produces while playing with a caregiver. (Young children's speech is not limited to this method, though.)

5.5 Best Practices for Teaching and Learning

The author has found that reduction in spontaneous speech lends itself extraordinarily well to teaching. All the instructor needs to do is to find some recordings in the language of instruction with massive reductions in them, splice out the highly reduced word(s) and save that file, and have the corresponding file with utterance context available. One should not give the files filenames that contain words of the utterance, so that one can display the signals on the screen in the classroom without giving away the answers. (If the language of instruction is English, readers are welcome to use examples from the author's website at nwarner.faculty.arizona.edu, but it is not difficult to find more examples. Any recording of spontaneous conversation is likely to contain them.)

The instructor plays the extracted highly reduced word(s) out of context, and says something like 'This is a phrase of two words, and it's in English. It really is. What do you think it might be?' After letting the students guess, one plays the context file. For most highly reduced examples like those in Section 5.3, the words will be completely unintelligible out of context but completely clear in context, and in fact the reductions are not even noticeable in context. For example, *a little* in (e) above is sometimes guessed out of context as *donut*. Students who are speakers of the same dialect occasionally do identify *but I was (just) like* strings correctly even without context.

Asking the students what they think the human brain is doing to be able to understand the words usually elicits replies of 'context'. For an advanced class, one can then ask the students what kinds of information might be in the context, and one can explore how hard the task was for non-native speaking students vs native speakers. For a phonetics course, one can ask students to transcribe some portion of highly reduced connected speech, and then compare the various transcriptions generated by the class. Students' transcriptions are often strongly influenced by the expected canonical form, missing many deletions, and this provides good material for discussion. The author has used a basic demonstration (without transcription exercise) successfully in classes from first-year General Education level and high school outreach visits to graduate seminar level. Students at all levels seem to find this topic highly engaging.

5.6 Future directions

The literature on connected speech thus far (above) has established that reduction is pervasive in every language studied and is more drastic than phoneticians often expect, that listeners cannot

understand connected speech forms when presented out of context, and that second language learners perceiving L2 connected speech in some ways resemble native speakers and in other ways do not. There are still several topics that we have only begun to explore in connected speech. Is reduction in spontaneous or casual speech the same in all languages, and caused primarily by constraints of the vocal tract, or is it partially language-specific? How are intonational and tonal distinctions modified in connected speech, moving beyond phonological sandhi processes to phonetic reduction of pitch distinctions? How much of connected speech reduction is contributed by speech rate, how much by familiarity with the interlocutor and casualness of the topic, how much by word frequency, etc.? How variable are speakers (within a speaker over time and across speakers) in how much they reduce? How is the reduction of connected speech realized in child speech, child directed speech, L2 speech and speech directed to an L2 listener, etc.? How does spontaneous speech reduction interact with presence of background noise? Returning to the more constrained topic of sub-phonemic effects at word boundaries (Section 5.4.1), the question remains of whether all phonological processes that affect segments at word boundaries have such sub-phonemic variability. Every one of these questions has been addressed in at least a few papers, but there is a lot we do not know about these questions.

Acknowledgements

The author would like to acknowledge Mirjam Ernestus and Benjamin V. Tucker for long discussion of the topics here, Seongjin Park, Mike Hammond, Skye Anderson, and other members of the University of Arizona Department of Linguistics for discussion of this paper, and the anonymous reviewer for helpful feedback. Some of the examples in Section 5.3 were collected as part of a larger project funded by NSF Grant #1022313.

5.7 References

Arai, T. (1999). A case study of spontaneous speech in Japanese. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*. Berkeley, CA: Department of Linguistics, University of California, pp. 615-618.

Barry, W. & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, 31, 51-66.

Bouavichith, D., & Davidson, L. (2013). Acoustic characteristics of intervocalic stop lenition in American English. *The Journal of the Acoustical Society of America*, 133(5), 3565-3565.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, 112(1), 272-284.

Brand, S., & Ernestus, M. (2017). Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: evidence from native listeners and learners of French. *The Quarterly Journal of Experimental Psychology*, 1-24.

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, 27(4), 539-571.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4), 155-180.

Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., & Frauenfelder, U. H. (2011). What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *The Journal of the Acoustical Society of America*, 130(6), 3980-3991.

Cruttenden, A. (2014). *Gimson's pronunciation of English*. London: Routledge.

Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *The Journal of the Acoustical Society of America*, 72(3), 705-716.

Crystal, T. H., & House, A. S. (1988a). Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, 83(4), 1553-1573.

Crystal, T. H., & House, A. S. (1988b). A note on the durations of fricatives in American English. *The Journal of the Acoustical Society of America*, 84(5), 1932-1935.

Davidson, L. (2011). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, 53(8), 1042-1058.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, 54, 35-50.

Davidson, L. (2018). Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association*, 1-26.

Dilley, L. C., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *The Journal of the Acoustical Society of America*, 122(4), 2340-2353.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81(1-3), 162-173.

Ernestus, M., Kouwenhoven, H., & Van Mulken, M. (2017). The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics*, 62, 50-64.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(SI), 253-260.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4), 789-806.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, 7, 515-546.

Greenberg, S. (1999). Speaking in shorthand–A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, 29(2-4), 159-176.

Hanique, I., Ernestus, M., & Schuppler, B. (2013). Informal speech processes can be categorical in nature, even if they affect many different words. *The Journal of the Acoustical Society of America*, 133(3), 1644-1655.

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, pp. 29-54.

Kharlamov, V. (2014). Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences. *Journal of Phonetics*, 43, 47-56.

Kharlamov, V. (2015). Perception of incompletely neutralized voicing cues in word-final obstruents: The role of differences in production context. *Laboratory Phonology*, 6(2), 147-165.

Kiefte, M., & Nearey, T. M. (2017). Modeling consonant-context effects in a large database of spontaneous speech recordings. *The Journal of the Acoustical Society of America*, 142(1), 434-443.

Koopmans-van Beinum, F. J. (1980). *Vowel Contrast Reduction: An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions.* Amsterdam: Academische Pers B.V.

Kouwenhoven, H., Ernestus, M., & Van Mulken, M. (2015). Register variation by Spanish users of English: The Nijmegen Corpus of Spanish English. *Corpus Linguistics and Linguistic Theory*, 14(1), 35-63.

Kuo, C., & Weismer, G. (2016). Vowel reduction across tasks for male speakers of American English. *The Journal of the Acoustical Society of America*, 140(1), 369-383.

Liu, Y. F., Tseng, S. C., & Jang, J. S. R. (2016). Deriving disyllabic word variants from a Chinese conversational speech corpus. *The Journal of the Acoustical Society of America*, 140(1), 308-321.

Lowenstein, J. H., & Nittrouer, S. (2008). Patterns of acquisition of native voice onset time in English-learning children. *The Journal of the Acoustical Society of America*, 124(2), 1180-1191.

Maekawa, K., Kikuchi, H., Igarashi, Y., & Venditti, J. (2002). X-JToBI: an extended J-ToBI for spontaneous speech. In *Seventh International Conference on Spoken Language Processing*.

Mattys, S. L., & Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *The Journal of the Acoustical Society of America*, 122(1), 554-567.

Mehta, G. & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech,* 31, 135-156.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, 65(6), 956-969.

Mitterer, H., & Ernestus, M. (2006). Listeners recover/t/s that speakers reduce: Evidence from/t/-lenition in Dutch. *Journal of Phonetics*, 34(1), 73-103.

Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *The Journal of the Acoustical Society of America*, 95(3), 1603-1616.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756.

Niebuhr, O., & Kohler, K. J. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics*, 39(3), 319-329.

Pierrehumbert, J. & Beckman, M. (1988). *Japanese Tone Structure. (Linguistic Inquiry Monographs*, *15*.) Cambridge, London: MIT Press.

Pollack, I. & Pickett, J. M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior,* 3(1), 79-84.

Shriberg, E. (2001). To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31(1), 153-169.

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, 118(3), 1677-1688.

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America*, 130(6), 4020-4031.

Spinelli, E., Cutler, A., & McQueen, J. M. (2002). Resolution of liaison for lexical access in French. *Revue française de linguistique appliquée*, 7(1), 83-96.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, 48(2), 233-254.

Torreira, F., Adda-Decker, M., & Ernestus, M. (2010). The Nijmegen corpus of casual French. *Speech Communication*, 52(3), 201-212.

Tuinman, A., Mitterer, H., & Cutler, A. (2011). Perception of intrusive/r/in English by native, cross-language and cross-dialect listeners. *The Journal of the Acoustical Society of America*, 130(3), 1643-1652.

Vasilyeva, L., Arnhold, A., & Järvikivi, J. (2016). Phonetic correlates of phonological vowel quantity in Yakut read and spontaneous speech. *The Journal of the Acoustical Society of America*, 139(5), 2541-2550.

Warner, N. (2011). Reduction. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice, eds., *The Blackwell Companion to Phonology.* Malden, MA & Oxford: Wiley-Blackwell.

Warner, N. (2012). Methods for studying spontaneous speech. In A. Cohn, C. Fougeron, & M. Huffman, eds., *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press, pp. 621-633.

Warner, N., & Arai, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *The Journal of the Acoustical Society of America*, 109(3), 1144-1156.

Warner, N., Good, E., Jongman, A., & Sereno, J. (2006). Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics*, 34(2), 285-293.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, 32(2), 251-276.

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, 130(3), 1606-1617.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, 66(4), 665-679.

Zimmerer, F., Reetz, H., & Lahiri, A. (2009). Place assimilation across words in running speech: Corpus analysis and perception. *The Journal of the Acoustical Society of America*, 125(4), 2307-2322.

Figure 5.1. Waveform and spectrogram of spontaneous conversational speech by a female native speaker of American English. The longer context is *...to be on the swim team, so [name]'s like 'I think I can get down to a 445 this summer', I was like, 'yeah, right'.* The portion *I think I can get down*, which has extreme reduction, is shown.

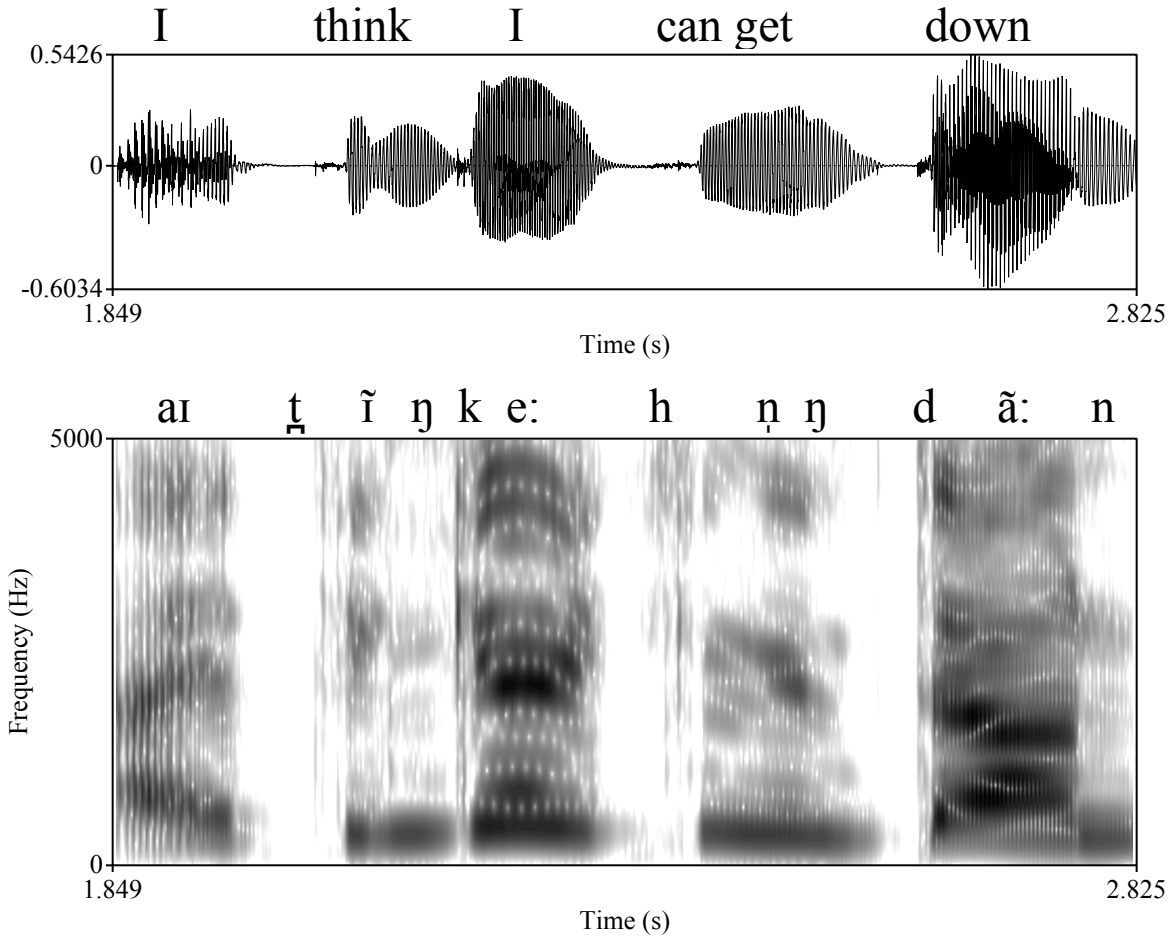


Figure 5.2. Waveform and spectrogram of spontaneous conversational speech by a female native speaker of American English. The longer context is *She wants to be a police officer, I think.* The portion *wants to be a* has extreme reduction.

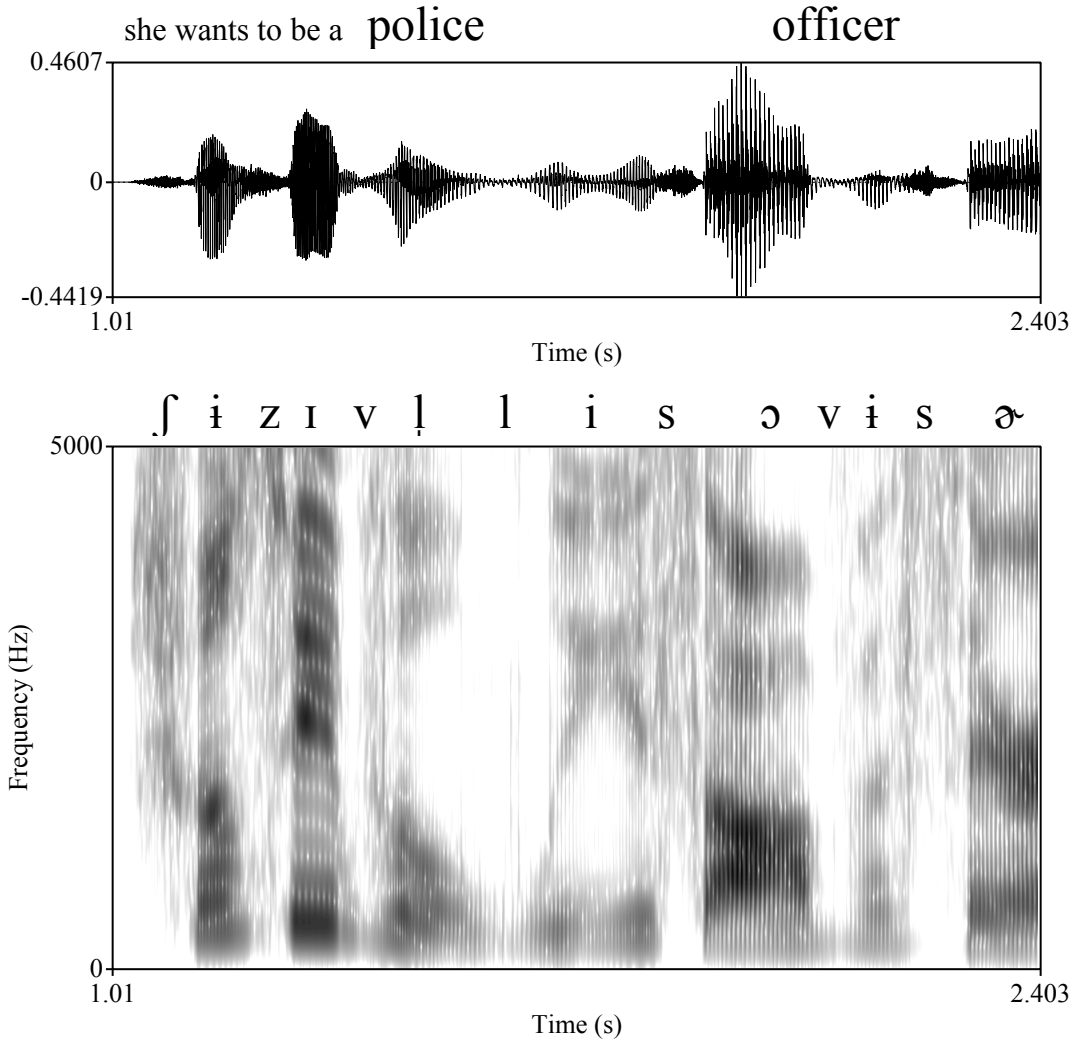


Figure 5.3. Waveform and spectrogram of spontaneous conversational speech by a female native speaker of American English. The longer context is *And like, one of my classes is, like, a writing-intensive class, and I didn't really know that, but I need to take it to graduate*. The portion *and I didn't* has extreme reduction.

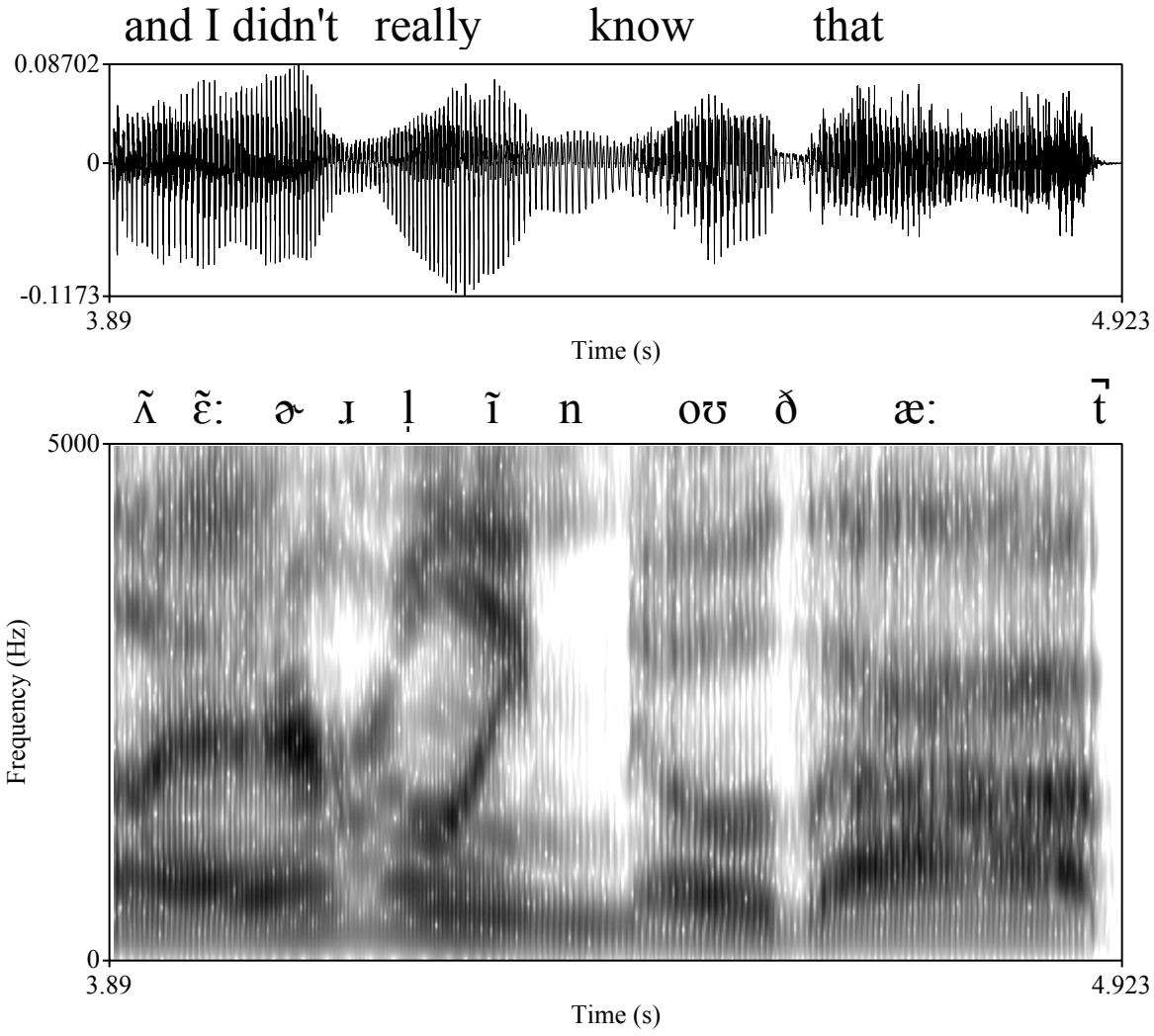


Figure 5.4. Waveform and spectrogram of spontaneous conversational speech by a male native speaker of Japanese. The longer context is *a, soo iu hitotachi yoo-no supeesu-ga aru-no-ka?* あ、そういう人たち用のスペースがあるのか？ 'Oh, is there space for people like that?' The portion *iu hitotachi* has extreme reduction.

