submission for *Speech Communication*, third revision August 6, 2025

Phonetic Reduction is Associated with Positive Assessment and other Pragmatic Functions

Nigel G. Ward, Raul O. Gomez, Carlos A. Ortega, Georgina Bugarini[1]

Department of Computer Science, University of Texas at El Paso

Abstract

A fundamental goal of speech science is to inventory the meaning-conveying elements of human speech. This article provides evidence for including phonetic reduction in this inventory. Based on analysis of dialog data, we find that phonetic reduction is common with several important pragmatic functions, including the expression of positive assessment, in both American English and Mexican Spanish. For American English, we confirm, in a controlled experiment, that people that speaking in a positive tone generally do indeed use more reduced forms.

Keywords: reduced articulatory precision, hypoarticulation, prosody, pragmatic functions, English, Spanish, corpus study, annotation, perception

Highlights

- We undertook the first perception-based study of the pragmatic functions of reduction
- We used dialog data in American English and Mexican Spanish
- Phonetic reduction is a common correlate of positive assessments in both languages
- Phonetic reduction is also associated with other pragmatic functions
- Phonetic reduction seems to function as a prosodic feature

1 Introduction

A fundamental goal of speech science is to inventory the meaning-conveying features of human speech. Despite great progress — the inventory of possible phonemes has long been essentially complete, and various prosodic-feature inventories are good enough to be often useful — one important feature has been lacking. Phonetic reduction, also known as hypoarticulation or reduced articulatory precision, needs, we claim, to be added.

Such a claim may be unfashionable. With the rise of machine learning and increasing amounts of data available, it has became clear that more features bring more power, even if they lack perceptual or linguistic motivation. This can be seen in the influential OpenSmile set of 6552 features, including, for example, the kurtosis of the second derivative of the shimmer, and the

[1]nigelward@acm.org, romargomez04@gmail.com, carlos.ortega2001@hotmail.com, georgina.bugarini20@yahoo.com

third quartile of the zero-crossing rate [Eyben et al., 2010]. Currently, the mainstream approach is to ignore considerations of feature meaning, and instead use features learned by pretraining, such as the 1024 of the last HuBert layer [Hsu et al., 2021, Mohamed et al., 2022], which work very well for many tasks, including some previously thought to require handcrafted features [Yang et al., 2021, Lin et al., 2022].

Nevertheless, there is still value in features which are directly hearable, as they can support explainable models, whose behavior we can fully understand. Further, such features can serve as understandable control parameters for speech synthesis, can enable comprehensible descriptions of language patterns to support learners, and can provide a leg-up for machine learning to achieve good performance from less data. We present evidence that phonetic reduction is such a feature: not only hearable but also meaningful and useful.

The rest of this paper is structured as follows: Section 2 surveys related research, Section 3 describes the data used, Section 4 describes our initial explorations, Section 5 describes a systematic corpus study that identified many functions associated with reduction, Section 6 describes a production study confirming one connection, and Section 7 concludes.

2 Background and Motivation

While the existence of phonetic reduction is well known, speech technology researchers have not generally seen it as useful. This section overviews related research, sketching out ways in which reduction is important in human language, and why it nevertheless has had no applications impact. We leave a thorough treatment to more comprehensive surveys [Ernestus and Warner, 2011, Jaeger and Buz, 2017, Cangemi and Niebuhr, 2018, Zellers et al., 2018].

It is common experience that speakers sometimes produce words that are difficult to recognize, mumbled, sloppy, or just less well articulated. These phenomena have been studied under various technical terms, most commonly the term we will use here: phonetic reduction or, for short, just "reduction." We can group related work in three categories.

First, there are many studies of reduction that build on the perspective that speakers tend to minimize effort, and in particular, often provide only just enough phonetic accuracy to successfully convey the intended content to the hearer [Lindblom, 1990, Levshina and Lorenz, 2022], a perspective that reflects a general tension in language between economy and redundancy [Horn, 1993]. This is seen in the well-established correlation between durational reduction and lexical predictability [Jurafsky et al., 1998, Bell et al., 2003, 2009]. Dialog systems developers may thus think of reduction, not as a capability worth emulating, but as a reflection of the foibles of human speakers, better to avoid. Certainly it is also true that pervasive reduction and poor intelligibility can be due to a lack of language proficiency, to depression, or to neural and physical impairments [Helfer et al., 2013, Jiao et al., 2016, Tu et al., 2018, Wynn et al., 2022]. Reduction can also be a indicator of disengagement and can cue perceptions of negative personality attributions such as laziness. Reduction can also result from cognitive factors, as when specific words are reduced due to priming effects [Gahl et al., 2012, Kahn and Arnold, 2015] or due to properties of the mental lexicon [Johnson, 2004, Munson and Solomon, 2004, Schubotz et al., 2015, Warner, 2019, Mulder et al., 2022].

Second, many studies have elucidated the structural factors involved in reduction, including

the ways that it is conditioned on the prosodic properties of specific words, on syntactic structure, on grammatical roles (for example, more reduction for function words), and on prosodic structure [Zellers et al., 2018, Machac and Fried, 2025, Bodur et al., 2023]. While these factors are important for natural-sounding speech, in speech synthesis such reductions have generally not been seen as worth modeling explicitly, but rather left to be learned implicitly from data.

Third, a few studies have elucidated the connections between reduction and specific pragmatic functions. Beyond marking things as not new, these include turn hold in English, German and possibly Swedish [Local and Walker, 2012, Niebuhr et al., 2013, Zellers, 2017], repair and agreement in Dutch [Plug, 2005, 2011], and sarcasm in German [Niebuhr, 2014]. Further, in English, reduction may be common in self-deprecating asides [Gustafson et al., 2023b], cues for backchannel feedback, expressions of a self-sufficient stance, and topic closings [Ward, 2019]. Reduction is also known to be involved in certain functions in specific phrases — for example English *gonna* (c.f. *going to*) in statements of intent, *why dontcha* (c.f. *why don't you*) for suggestions, *dunno* (c.f. *I don't know*) to convey indifference, and so on [Hawkins, 2003, Schubotz et al., 2015]. None of these pragmatic functions, however, are top-of-mind for speech applications builders.

Indeed, for the classically-framed problems of text-to-speech and speech-to-text, reduction is just a nuisance: a source of noise in the mappings between speech sounds and word sequences. However, there is an increasing realization that, for spoken dialog systems to be useful in newer scenarios, such as those involving interpersonal sensitivity or situated action, they will need to exploit more of the pragmatic information that spoken language can convey [Marge et al., 2022]. This has led to increased interest in elucidating the extra information that is present in speech. This challenge has been taken up by the speech synthesis community, where, with the problems of intelligibility and naturalness largely solved, there is much interest in finding ways to improve the communicative range of synthesizers and to make them controllable, as surveyed in [Wagner et al., 2019, Mohan et al., 2021, Lameris et al., 2023]. In particular, there is interest in modeling prosody as a way to improve synthesis quality, either for human-controllable parameters, or to build better loss functions for training [Huang et al., 2023], but this interest has not yet spilled over to the topic of reduction. For example, in the 90-some papers on speech synthesis at Interspeech 2023, over half included discussion of prosody, but the features considered were mostly pitch and duration, with 10% mentioning intensity and 1% voicing properties, and none mentioning phonetic reduction.

Thus, phonetic reduction is not currently a mainstream topic in speech technology. At the same time, a few individuals and groups have, over the years, called for its inclusion into the prosodic features inventory [Beller et al., 2008], stressed that speech can be noticeably reduced and still adequately intelligible [Niebuhr, 2016], noted its importance in conversational-style speech [Adigwe et al., 2024], found it to be relevant in judgments of pragmatic similarity [Ward et al., 2025a]. Further, three groups have developed experimental synthesizers with the ability to control the level of reduction [Picart et al., 2014, Birkholz et al., 2016, Gustafson et al., 2023a]. However, none of these systems are ready for general use, no commercial synthesizers support such control, and, in general, reduction remains a niche topic.

Overall, while there is some scattered knowledge of the pragmatic functions that reduction can serve, none of these have been of compelling importance for applied speech researchers. There is something of a chicken-and-egg problem here: detailed studies of reduction functions are difficult without tools, but tool builders have no motivation to add reduction modeling capabilities without

evidence of potential utility. Here we aim to jump-start a virtuous cycle, of greater appreciation of the importance of reduction leading to better tools, in turn leading to greater knowledge and to practical applications.

Thus we set out to obtain a more complete view of the functions that reduction serves.

3 Data and its Annotation for Reduction

As our aim was discovery, we started with a corpus-based approach, with the first step being annotation for perceived reduction.

3.1 Data

The first critical choice was of what data to use. Our most fundamental decision was the choice to use dialog data, not read speech, because the most important potential applications relate to dialog. We chose to examine two languages: English, as a well-studied language where we still hoped to find something new, and Spanish, as a first look at a non-Germanic language.

Our specific choice of data was based on five desiderata: 1) exhibiting diverse pragmatic functions, 2) having diverse speakers, 3) having good audio quality, 4) being among friends or peers (as reduction is more common in informal conversation), 5) having mostly dialects, topics, and dialog styles with which we are familiar, to enable confident annotation. To meet these needs we selected a few conversations from our own data, originally collected for another purpose (namely, subsequent partial re-enactments in another language) [Ward et al., 2023]. This was collected locally in El Paso, Texas. The dialects were General American English and Mexican-American Border Spanish. The selection was made on the basis of interestingness and diversity, and all of the conversations selected happened to be ones where the participants were given neither prompts nor scenarios. Rather, the topics were freely chosen, and included, for example, talk about upcoming internships and classes, the aftermath of a spat, talk about participating in sports, stories of working in a jail, and discussion of anime. As will be seen below, this data supported all phases of the work, in different subsets and via different processes.

3.2 Reduction Annotation Procedure

The second critical choice was how to annotate the data for reduction. We did this by hand, because, at the time, there were no good alternatives: The then-available automatic reduction-detection methods being untested for temporally fine-grained discriminations [Tu et al., 2018, Lubold et al., 2019, Lee et al., 2006, Jiao et al., 2016, Chen et al., 2022, Ward and Ortega, 2024]. While word duration is easy to measure, and shortening is common correlate of reduction [Jurafsky et al., 1998, Kahn and Arnold, 2015], this is not a reliable proxy for all aspects of reduction nor for perceived reduction.

Our specific annotation target was perceived reduction. This is an unusual choice, as most previous work has annotated for correlates or aspects of reduction, such as shorter phonemes, more centralized vowels, more co-articulation, and the effects of various language-specific phonological rules, such as coda-consonant deletion in English [Jurafsky et al., 1998, Turnbull, 2015, Koreman, 2006, Aguilar et al., 1993, Picart et al., 2014, Machač and Fried, 2023, Niebuhr and Kohler, 2011,

Schettino and Cutugno, 2025]. Other work has exploited one of those rare corpora which have both full segmental labels and canonical transcriptions, from which one can infer where certain types of reduction occur [Jurafsky et al., 1998, Niebuhr et al., 2013]. The problems with these methods are that hand labeling does not scale, and that no single feature correlates adequately with all reduction phenomena [Burdin and Clopper, 2015].

Since no previous work seems to have annotated a corpus for perceived reduction, we had to design our own annotation procedure. We started with an initial rough examination of some data and our perceptions, listening together to a few minutes of each language, noting all issues and choices. We then developed a lightweight set of guidelines, and gave them to the annotators in the form seen in Figure 1. The Granularity clause reflects our working assumption that reduction in dialog is mostly suprasegmental. (Of course, reduction occurs at various times scales. Most previous work has focused on "short-term" occurrences, lasting for a phoneme, a syllable, or a word, while other work has examined reduction over the "long-term," across a conversation or an extended speech sample. With our interest in pragmatic functions, we are interested in "mid-term" reduction: reduction at the level of a few words or phrases. In practice, most of the labeling was done at the scale of regions that covered one or a few words, depending on the granularity of the perceived variations in articulation level. (Most words in this data occurred in larger utterances, of various lengths.) The Codes clause reflects our observation that, while reduction is a gradient phenomenon, we could at most reliably annotate for four steps of the continuum: highly reduced, reduced, normal, and enunciated (clearly articulated). The Criteria clause acknowledges the essential subjectivity of these judgments. The Scope clause reflects the fact that non-lexical items [Ward, 2006] — such as *umm*, *wow*, and *uh-huh* — lack reference pronunciations in the same sense as words, so it would make no sense to try to annotate them for reduction. The Confounds clause was intended to prevent annotation by proxy percepts, with the specific cautions based on the literature on the prosodic correlates of reduction in read speech; however, as it turned out, none of these actually correlated with reduction in this data [Ward and Ortega, 2024]. We would recommend that future studies omit such a clause.

Granularity: We are labeling at the level of utterances, phrases, or words. We do not label individual syllables or phonemes.

Codes: 3 = highly reduced, 2 = reduced, 1 = normal, 0 = enunciated

Criteria: Subjective judgments of being poorly articulated or possibly hard to understand without context.

Scope: We only label words. We label all words. That is, we label all word-containing parts of utterances. Every one of these should have one of the four labels above. We do not label laughter, word fragments, tongue clicks, or nonlexical utterances (*uh-huh, umm*, etc.).

Confounds: We know that some things correlate with reduced speech (fast rate, low intensity, low pitch, narrow pitch range), but we try not to let these things affect our judgments.

Figure 1: Reduction Annotation Guidelines

3.3 Yield and Quality Assessment

The third author, a Spanish-English bilingual, annotated a sampling of English and Spanish conversations, totaling 31 and 25 minutes, respectively, for every word in both tracks, left and right, using Elan as the annotation tool. This yielded a total of 3051 annotated regions for English and 1614 for Spanish. For both languages, some degree of reduction was very common, present in 26% and 31% of the regions in English and Spanish, respectively. The annotations are available for download at http://www.cs.utep.edu/nigel/reduction and the audio for the dialogs via http://www.cs.utep.edu/nigel/dral; we note that this is the first publicly released collection of data annotated for perceived reduction.

To roughly gauge the subjectivity of these annotations, we had a second annotator label 2 minutes each of the English and Spanish. Cohen's Kappa between the two annotators, for both English and Spanish, was 0.25 for the four categories, and 0.35 for the two-category discrimination of 0 or 1 versus 2 or 3. Noting that the second annotator tended to perceive reduction more often, we also computed the correlation: it was 0.57, giving a somewhat brighter picture of the agreement.

We note that one could certainly obtain higher agreement by using a more structured process, or by leveraging more objective correlates of reduction. However this would run counter to our strategy of using direct perceptions, and would prematurely constrain what we might find [Aroyo and Welty, 2015]. Further, detailed analysis of all differences of opinion between the annotators [Ward and Ortega, 2024] revealed that these usually affected at most one or two words per phrase, and only rarely differed by more than one point. For the analysis below, which relies only on average behavior across multi-word phrases and across many such phrases, such differences are therefore just an acceptably small minor noise source. Further, none of the complicating factors [Ward and Ortega, 2024] seem likely to much relate to any of the pragmatic functions that we ended up investigating.

4 Initial Analysis

With the annotated data as a resource, we began analysis, using qualitative inductive methods to develop hypotheses. For this we used a structured process, which we describe in some detail in the hope that this will help those contemplating similar studies, but without expecting this initial analysis to be replicable in any sense. More on this procedure and our initial observations appears in the companion technical report [Ward and Ortega, 2024] and at the companion website, `https://www.cs.utep.edu/nigel/reduction`.

For English, our first step was to examine all regions in ten minutes of dialog that were labeled as reduced and note down the functions observed at these times. Second we grouped these into several categories. Third we refined these as we went back to re-examine the instances, leading to nine categories: 1) fillers, interjections, and backchannels, 2) prosody carriers like *like* and *you know*, 3) uncertainty markers, 4) recapitulations, 5) predictable words, 6) downplayed phrases, including parentheticals, 7) topic closing moves, 8) turn grabs, and 9) personal feelings, including preferences and desires. These categories are neither exhaustive nor mutually exclusive; for example, there was a case of reduction occurred as part of yielding to resolve a moment of speech inadvertently overlapping that of the other speaker.

Fourth, to winnow out the categories not truly associated with reduction, we wrote short descriptions of the nine categories and used these to systematically label all regions in the remaining 21 minutes of data for these functions, plus two controls: negative assessment and positive assessment. Examining the strength of the associations with reduction, we found, to our surprise, that the positive assessment category (defined below) was strongly associated with reduction. Of the 9 functions, all except prosody carriers and fillers showed up as significantly correlated with reduction. For recapitulations and predictable words, this was unsurprising, based on the literature, but the connections with the others were novel.

The procedure for Spanish was similar. We examined dialogs from the same corpus, but not overlapping the reduction-labeled set, and after some iteration arrived at 7 categories: uncertainty markers, topic closings, turn grabs, personal feelings, downplayed phrases, positive assessment, and agreement, with only the latter something not also seen in English.

4.1 More on Positive Assessment in English

This section briefly explains what we mean by positive assessment, our focus in the rest of this article. We chose to focus on this for three reasons: its connection with reduction was the most surprising, it was the second-best supported in the initial analysis, and it is most obviously of practical importance.

We defined "positive assessment" as "positively assessing or expressing positive feelings about something, some happening, or someone, including the interlocutor or the speaker himself." Thus we intended it to be taken quite broadly, and in fact the annotator did so. Here are four examples; in each, underlining marks the reduced region, * marks the start of the region where the pragmatic function (here positive assessment) is clearest, and / marks speaker change.

- *I want to work with the inmate population / *oh wow, that's interesting* (006@1:53)
- *Yeah, because I got the research position here, and I thought, *it was a good, opportunity, because I want to do my Ph.D.* (006@2:39)
- *I was in taekwondo . . . no, but. *It was, it was pretty cool. I liked it.* (033@0:15)
- *I've seen Bleach, I've seen Akame ga Kill / *That one's so good; I love that one. That one's like, a nice short one ...* (043@2:28)

While there is a fair body of work on the prosodic properties of positive assessment, and positive sentiment more generally, as surveyed elsewhere [Niebuhr, 2010, Ward and Jodoin, 2019], the connection with reduction been not been previously noted. There is an even larger body of work on the properties of positive emotion, affect, and valence. While positive assessment relates at best marginally to these phenomena — since in these conversations the participants were generally in a good mood, and certainly not reacting emotionally to anything — it is interesting that the correlation found here does not align with the finding of the only work connecting reduction to emotion: that there is more reduction in sad speech and depression [Lee et al., 2006, Hall-Lew, 2024]. Interestingly, reduction often affects only some of the words in positive phrases, sometimes excluding the most valenced words, such as *good*, *love*, and *liked* in the examples above.

5 Corpus Study

To determine which of the functions actually have a statistically significant connection with reduction, we did corpus studies, for both English and Spanish.

Our method was to obtain an independent annotation of the data for 8 pragmatic functions, derived from our initial analysis, as shown in Figure 2. Of these, we expected 7 to correlate with reduction in one or both languages, and included negative assessment as a control. Annotation was done by the fourth author, a Spanish-English bilingual, who at the time was naive to the purpose of the annotation. She received about 10 minutes of explanation, with examples, and learned the value of considering all factors: the words spoken, the prosody, and the context. She then annotated the same 21 minutes of English and all 25 minutes of Spanish.

We must digress briefly to note that this work was unavoidably subjective, and especially so due to the brevity of our category descriptions. We can illustrate this with differences in how two of us interpreted two of the categories. For Uncertainty Markers, one of us interpreted the definition to include many occurances of the discourse marker *like* and some false starts and hesitations, which the other did not. For Turn Grabs, one of us interpreted the definition to include many instances of supportive simultaneous talk and cases where the speaker suddenly sped up and produced a new talk-spurt, and the other did not. We would suggest that future annotation efforts might use more explicit definitions, to explictly include or exclude such subcategories. However there were fewer differences for the other categories, and in particular, the definitions of Positive Assessment and Topic Closings did not appear to have as much scope for alternative intepretations.

In any case, after obtaining these functional annotations, we applied a script to compare them with reduction annotations. This computed two types of indication. The first type was the average reduction level for each functional region, and, derived from this, the average across all regions for each function. These are shown in the second column of Tables 1 and 2 below. We also used an unpaired, one-sided t-test to estimate the evidence for the distribution of average reduction levels for a function being significantly above the distribution for all regions. The second indication was, aggregating across all regions with a given functional label, the percent of the time they exhibited each reduction level. As noted above, these ranged from 0 (enunciation) through 1 (normal) and 2 (reduced) to 3 (highly reduced).

5.1 Results and Discussion: English

The results for English are seen in Table 1. The most strongly associated function was Positive Assessment, and the other strongly associated function was Topic Closings, as, for example, in

- *. . . it's like cloud services, it's like AWS / oh, okay / *that kind of stuff yeah yeah, you know, I don't know, this is what happens I guess. / Which summer . . . ?* (013@0:57)

Four other functions had tendencies to connection with reduction, namely Downplayed Phrases, Agreement, Turn Grabs, and Uncertainty Markers. (Examples for these appear in [Ward and Ortega, 2024].) With the overall standard deviation being 0.89 steps, the effect sizes ranged from modest to reasonably large: for example, the effect of being in a positive assessment was 0.28 standard deviations, and of being in a topic closing 0.60.

PO Positively assessing or expressing positive feelings about something, some happening, or someone, including the interlocutor or the speaker himself.

NE Negatively assessing or expressing negative feelings, like *super dangerous*, and *I peaked back then*. More often, the negative feeling is not in the words but in the tone.

UC Marking Uncertainty, for example with phrases such as *I don't know, hopefully, I'm not sure yet*, and *I feel like*.

DP Downplaying something, such as in a parenthetical, aside, self talk, or in a side comment that is not intended to be responded to. Repair markers are also in this category, as in *I'm taking Machine Learning, no, sorry, uh, Deep Learning*.

TG Making a Turn Grab, when a speaker takes the turn before knowing quite what he wants to say. Also in this category we include "rush-throughs," where a speaker seems to revoke a turn yield by speaking quickly to forestall the other from taking the turn after all.

TC Making Topic Closing moves, for example with cliched phrases and other ways to show that the speaker has nothing more to say, and no further interest in, a topic.

AG Agreeing with something the other has said, or confirming what they have asked, for example with words like *exactly* or *that's true*.

PF Expressing Personal Feelings, including simple statements of wants and preferences, and also more thoughtful or introspective statements.

Figure 2: Pragmatic Function Annotation Categories

5.2 Results and Discussion: Spanish

The results for Spanish are seen in Table 2. For Spanish also, the connection with Positive Assessments was the strongest. The other functions with strong connections were Turn Grabs and Downplayed Phrases. With an overall standard deviation of 0.83, the effect sizes were modest, for example, 0.19 standard deviations for positive assessment.

To illustrate the three functions with statistically significant connections:

1. Positive Assessment
   - *abrieron una sección nueva de Star Wars, que *está bien chida vato.* (001@1:02)
     *they opened a new Star Wars section, which is so *cool dude.*
   - *. . . *si disfruto ver a bastantes amigos y todas las cosas.* (001@4:15)
     *. . . *yes I enjoy seeing a lot of friends and all that.*
   - *es que *están ricos, ella los hace bien ricos.* (008@0:17)
     *it's just that *they are delicious, she makes them very delicious.*
2. Turn Grabs (many of which here involved actually cutting off the other speaker)

	average	n	p	percent with reduction label: 0	1	2	3
all regions	0.98	3051	–	35%	38%	21%	6%
Positive Assessments	1.23	40	.008	26%	35%	30%	9%
Topic Closings	1.51	8	.015	4%	49%	40%	7%
Downplayed Phrases	1.17	26	.101	30%	30%	33%	7%
Agreement	1.13	32	.122	32%	33%	26%	9%
Turn Grabs	1.08	37	.137	33%	33%	26%	8%
Uncertainty Markers	1.09	41	.156	36%	28%	26%	9%
Negative Assessments	0.96	65	.571	44%	25%	22%	9%
Personal Feelings	0.89	39	.850	43%	29%	22%	5%

Table 1: Reduction Statistics for Various Pragmatic Functions in English. The second column indicates the average reduction levels. The third column shows the number of occurrences: the count of function-labeled regions that overlapped at least one region with a reduction-level label. Functions are ordered by strength of evidence for a relation to reduction, as measured by p values for the t-tests, in the fourth column. The remaining columns show the percentages of 10-millisecond frames at each reduction level. The "all regions" statistics are for all regions with reduction-level labels, not limited to those at times for which a functional label was assigned.

- *. . . esas si estan hechas a mano y luego / *pues es que no no es como que hechas a mano. . .* (008@0:46)

  *. . . it's just that those are made by hand and then well / *It's because no not that like it's handmade. . .*

- *El huevo con tocino no va en sandwich / Pues tienes razon. / Si no / *O que te digo. . .* Pues a ti, ¿Que te gusta el café o el té? (012@1:24)

  – *Bacon and egg don't go together in a sandwich / Well, you're right./ Yes / *Oh what do I tell you. . .* Well, do you like coffee or tea? (012@1:24)

- *que no ibas a aplicar a eso lo de / *pues si pero pues no me, como que no tiene. . .* (012@3:26)

  *weren't you going to apply to that of / *well yes but well it doesn't, like it doesn't have. . .*

3. Downplayed Phrases

- *Le fui a preguntar que si / al profesor/ * el profe, el profe. Bueno le fui a preguntar que si podía buscar en mi mochila, a ver si tenía otro (lapiz) porque él no tenía nada.* (003@0:24)

  [talking about needing a pencil during an exam] *I went to ask him / the teacher / *the teacher, the teacher. Well I went to ask him if I could look in my backpack, to see if I had another one because he didn't have anything.* (003@0:24)

	average	n	p	percent with reduction label: 0	1	2	3
all regions	1.03	1614	–	31%	38%	29%	2%
Positive Assessments	1.19	39	.022	24%	37%	35%	4%
Turn Grabs	1.25	27	.029	19%	43%	32%	6%
Downplayed Phrases	1.36	12	.036	16%	38%	41%	5%
Negative Assessments	1.15	72	.054	26%	36%	36%	3%
Topic Closings	1.27	4	.198	24%	25%	51%	0%
Personal Feelings	1.10	23	.294	32%	29%	36%	3%
Agreement	1.02	22	.522	21%	55%	22%	1%
Uncertainty Markers	0.94	39	.819	35%	38%	26%	2%

Table 2: Reduction Statistics for Spanish.

- *Yo apuesto que me fue medio bien. Hay muchas cosas que pues yo digo que sí me sé, si me supe. *Porque, lo que estudié ayer. Pero no estudie, no estudie, me faltaron estudiar dos capítulos del libro.* (003@1:28)

  *I bet it went pretty well for me. There are many things that, well, I say that I do know, I did know. *Because, of what I studied yesterday. But I didn't study, I didn't study, I missed two chapters of the book.* (003@1:28)

6 Experimental Verification

Based on the corpus study, we hypothesized that English speakers would reduce more in utterances said in a positive tone. As the corpus study involved subjectivity, we designed a controlled experiment. This has two phases: we obtained both positive and neutral versions of the same phrases, and then we obtained judgments of which were more reduced. Both of these procedures were deemed exempt from human-subjects review by our local Institutional Review Board. This section presents the experiment design and the results.

6.1 Phase 1: Elicitation of Positive and Neutral Productions

Motivation: Wanting to control all factors other than the level of positivity, we had participants create two versions of several phrases, one positive and one neutral. Thus we obtained a collection of positive-neutral pairs, each with the same lexical content and produced by the same speaker.

Stimuli: Because reduction in monologues may pattern differently from reduction in dialog [Ward and Ortega, 2024], we chose to collect dialog-style samples. We biased the participants towards this style by 1) having them listen to phrases from actual conversations, with enough context to make them clearly real, and 2) asking them to "say the phrase in the way you might if you were part of this conversation." Thus the participants served as reenactors. Table 3 shows the phrases we chose. These were taken from the same corpus of student conversations [Ward et al., 2023], since we wanted our reenactors to be able to comfortably imagine themselves talking to a peer about these topics. Our primary criterion was to select phrases whose valence could easily

and you want to work with kids? /
I want to work with the inmate population /
oh, wow [that's interesting]

I was born in El Paso but raised in Juarez. I was here literally for a day and then I moved back /
[me too, actually]

I think I will be boasting, but I was on the varsity team /
ah, okay, okay /
I actually went, like, for [two tournaments] to Cathedral

what class are you taking in the summer? /
Applied Agile Software Development. *It's like cloud services, [like AWS], that kind of stuff*

Do you know what classes you're taking next semester? /
yeah I'm taking [computer security]

next semester? /
in the summer, hopefully. /
happens, I guess. *[which Summer is it] gonna be?* the first one or the second one?

Table 3: The phrases, including the context that the reenactors heard. The re-enacted parts are in italics, and the parts heard by the judges (the "payloads") are in brackets. In the corpus, the first three of these were positive in tone and the last three neutral.

vary, depending on how they were said. We did not consider whether their appearances in the corpus included reduction nor whether they included phonemes that are likely to become reduced. To improve the likelihood that the results would generalize, we aimed to include some diversity, of three types: diversity of topic, of syntactic form, and of level of positivity in the original conversation. Regarding the latter, 3 phrases originally appeared without any positive feeling, and 3 with some degree of positive feeling, from weakly expressed to strongly expressed.

Procedure: The recording protocol was straightforward apart from steps added to deal with a potential confound: speaking rate. Pilot experiments showed a tendency for the positive renditions to be generally faster than the neutral ones. Increased rate is a known cause of reduction, and if not controlled, could explain away any effects. As our aim here is to show a *direct* connection between positivity and reduction, we needed to rule out this confound. In order to get positive clips that were no shorter than the neutral ones, we did two things. First, we ordered the re-enactments so that for each phrase each subject did the positive rendition first, thinking that the second version, as a repetition of something already said, would tend to be faster. This appeared to have only a small effect. Second, we added a step, such that if the neutral payload came out slower than the positive one, the reenactor would redo the former until it was about as fast as, or faster than, the positive rendition. (We also considered obtaining the desired rate relation in the opposite way, by having the reenactors say the positive version slower, but found that these often came out sounding condescending.)

Thus, for each phrase, each reenactor

1. Heard the phrase as used in a real conversation.
2. Re-enacted it with a positive frame of mind.
3. Re-enacted it with a neutral frame of mind, then waited for the experimenter to measure

the durations, and if necessary produced another neutral version, this time faster.

4. Read aloud each of the words in the payload, as if making sample recordings for a dictionary.

Before starting, each participant received an explanation of the procedure and gave consent to release their recordings to the public domain. They then donned a Shure BRH441M head-mounted microphone connected via a VOLT1 USB audio interface to a laptop. Their audio was recorded using Reaper, which enabled realtime monitoring of volume, and the ability to subsequently mark the duration of each payload.

Participants: The reenactors were 6 university students, recruited by word of mouth, mostly from among friends, plus one who had participated in previous studies and had asked to be invited back. There were no exclusion conditions. The reenactors were half male and half female and mostly Hispanic. While most spoke Spanish as a second native language, all were fully proficient in American English. Each was compensated US $20 for their time.

Observations: Overall this took about 20 minutes for most reenactors, but #12 found it difficult to produce neutral versions short enough, so he took more time. The dictionary style recordings were generally clearer, but we noticed that Reenactor #1 produced *interesting* with 3 syllables in dictionary style, but 4 syllables in both her positive- and negative-tone versions.

Regarding the durations, in the end, the average duration of a positive clip was 916 milliseconds, 57 milliseconds longer than the average negative clip, 859 milliseconds. For only 6 of the 36 pairs was the positive version shorter/faster than the negative, and most of these were by less than 10 milliseconds, which is approaching the limits of measurement accuracy. The pair for which the positive was relatively shortest was Reenactor #12's production of *which summer is it*, where the positive and neutral durations were 578 and 600 milliseconds, respectively. Thus, despite repeated attempts, he was only able to get the neutral to within 22 milliseconds (4%) of the duration of the positive.

6.2 Phase 2: Evaluation of Which Versions were More Reduced

Motivation: Given these collected productions, we needed a way to determine which were more reduced. As noted above, no suitable automatic methods were available, so we chose to use perceptions.

Procedure: Since absolute judgments of reduction are difficult, we only asked our participants to judge relative reductions between stimulus pairs. Since judgments of reduction generally reflect both top-down and bottom-up factors, we had participants rate at the phoneme level, to be closer to a pure perception task.

Participants entered their judgments though screens like that in Figure 3. The instructions were "please mark which of the pair is more clearly articulated (which is generally the one more similar to the careful-reading-style reference)." The parenthetical was included due to the issue with one production, mentioned above. We asked them to judge which was "more clearly articulated," to be less confusing than "which is more reduced."

Because reduction is often subtle, we gave subjects the ability to play each of the clips as often as they wanted. Since we wanted to glean every perception, no matter how weak, we gave the judges no option to skip a pair. Rather, the instructions specified "if not sure, please just make your best guess."

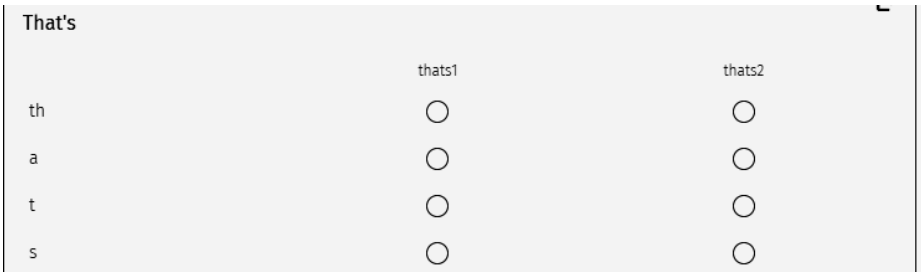

Figure 3: Data entry screen for judgments. For each phoneme, subjects selected whether it was clearer in the first or second stimulus.

To control for possible order bias, we randomized whether the positive or neutral clip was presented on the left or right. To reduce the chance that participants could perceive the positive or negative feeling, which might affect their judgments in some way, we randomized the order of presentation of the words, so they never heard a full phrase. This was not entirely successful, in that, for a couple of word pairs, it seemed obvious, to us, even from the word in isolation, which was more positive (usually by high pitch and faster rate); but of course the subjects weren't looking for this, as we didn't tell them our hypothesis until the debriefing stage.

Judges listened through headphones. They were proctored by the experimenter, sitting by their side and watching, to encourage them to continue to devote the high level of attention needed for these subtle judgments.

Stimuli: Since phonemes in isolation often don't sound like anything, each stimulus was an entire word with enough context to be able to perceive all the phonemes of the word: 20ms both before and after the word, or more, as needed. As an anchor we gave them references, namely the dictionary-style productions. Since our subjects were phonetically-naive, we gave them the list of phonemes in conventional spelling.

Participants: We chose to use the perceptions of untrained subjects, to avoid any possible theory-induced bias. All the judges were again university students, recruited and compensated in the same way as the reenactors. There was no overlap between the two groups. The judges were half male and half female, all Hispanic, and, again, mostly had Spanish as a second native language. There were 6 judges. Each judge made 150 judgments: 75 phoneme pairs each across the productions of 2 reenactors. Thus every phoneme pair production was judged by two judges.

Observations: Overall, including explanations, this process took 15 to 50 minutes, with Judge #22, the one who by far took the most time, very saliently putting in a lot of effort, listening to all the stimulus pairs again and again.

6.3 Results

Across all the judgments, as seen in Table 4, the phonemes from the positive recordings were generally perceived as more reduced (less clear) than the ones from the negative recordings. This was statistically significant according to a generalized linear mixed effects model. Specifically, with random effects for reenactor, judge, and phrase, Matlab's fitglme function gave an estimate of 0.43, with a t statistic of 26 and a p value of 6e-112.

positive	neutral	total
387	**513**	900

Table 4: Number of times that the rendition which was judged clearer came from the positive source utterance, or from the neutral one. Bolding indicates the judgments that aligned with the hypothesis.

6.4 Further Analysis

Although this experiment was not designed for systematic investigation of other factors, we did two post-hoc investigations.

First we investigated individual effects. As seen in Figure 5, results varied by reenactor, from #12, whose productions were strongly aligned with the hypothesis (and who the one who spent the most time on his productions) to #1, whose productions were judged opposite to the predicted direction. It is not surprising that the association between reduction and positive assessment, while prevalent, is not valid for all speakers of English, and it would be interesting to explore this [Ward and Ortega, 2024]. Results also varied by judge, as seen in Figure 6, from #5, whose judgments were near chance, to #22, who overwhelmingly judged the neutral phonemes to be clearer (and who was the one who had devoted the most time to the task).

Reenactor	positive	neutral
1	**82**	68
2	65	**85**
4	60	**90**
12	55	**95**
16	64	**86**
17	61	**89**
Total	387	**513**

Table 5: Overall selection of clearer phonemes per reenactor.

Second, we investigated the per-phoneme tendencies. As seen in Figure 4, while most phonemes were more reduced in the positive case, some had the opposite tendency. Post hoc, we noticed two factors that may explain this. The first factor seems to be various kinds of novelty

Judge	positive	neutral
5	74	**76**
7	60	**90**
8	71	**79**
10	62	**88**
22	49	**101**
1325	71	**79**
Total	387	**513**

Table 6: Overall selection of clearer phonemes per judge.

or stress. This could explain the countervailing tendencies seen: for *AWS* which was entirely new information in the corpus, entirely unpredictable from the context; for the *too* of *me too*, which was, of all the original phrases, the one with the most positive feeling; and for the word *summer*, which was used in an unusual, novel sense, to refer to a branded time period, either the first summer half-semester or the second. The second factor is likely general reducibility, as seen in the unstressed syllables of *interesting*, *actually*, *computer*, and *security*, and in the /m/ of *summer*. Perhaps generally-reduced syllables lack the capacity to be even more reduced in positive-toned productions. Together, these observations suggest the possiblity of a more accurate description, perhaps stating that the increased reduction on positive renditions affects all phonemes except those in positions that are generally highly reducible or those in highly novel or stressed words. Further work will be needed to refine and test this.

To support further analysis, we make all stimuli, judgments, and documentation available at https://www.cs.utep.edu/nigel/reduction/ .

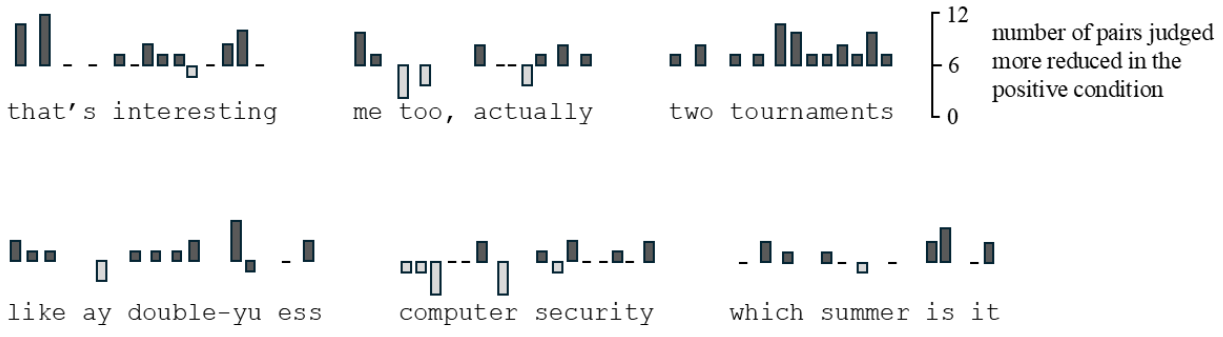

Figure 4: Per-phoneme tendencies to more reduction (dark gray) or less reduction (light gray) for the positive productions versus the neutral ones. As each phoneme pair appeared in 6 stimuli and was judged twice, there were 12 total judgments. For each phoneme, horizontal lines indicate parity, that is, 6 judgments each way, and the bars indicate the number of times, ranging from 3 to 11, that the positive member of the pair was perceived as more reduced.

7 Summary, Implications, and Open Questions

The major contribution of this paper is the finding that reduction is associated with important pragmatic functions, as summarized in Table 7, notably including positive assessment in American English.

Niebuhr (2016) argued that, in contrast to the traditional conceptualization of reduction as merely "the result of the two antagonistic forces 'articulatory economy' and 'sufficient contrast'," a new view is emerging, aware that reduction can serve the expression of various communicative functions, as mentioned in Section 2. The new findings here, of two associated functions in English and three in Spanish, one confirmed with a perception experiment, support this perspective.

This work also may have practical significance, in the form of benefits for future dialog systems. This most clear for the function of positive assessement. In one direction, there are many reasons to want to continuously track the user's state of satisfaction, utterance by utterance, but this

	English	Spanish	German	Dutch
positive assessment	⇑+	⇑		
topic closings	⇑	↑		
downplayed phrases	↑	⇑		
agreement	↑	–		
turn grabs	↑	⇑		
uncertainty markers	↑	↓		
personal feelings	↓	↑		
turn hold	⇑		⇑+	
given/recapitulated/predictable	⇑+			
sarcasm			⇑+	
repair				⇑

Table 7: Summary of Functions Associated with Reduction. ⇑ means good corpus evidence, ↑ means a corpus tendency, ↓ means an inverse tendency, – means no tendency, + means experimental confirmation, and blank means not yet studied. The first seven rows and two columns summarize our findings here; the rest are based on the literature surveyed in Section 2.

has always proven difficult [Schmitt and Ultes, 2015]. Monitoring variations in the user's level of phonetic reduction could enable more accurate tracking and thereby more responsive dialog systems. In the other direction, dialog systems should be able to synthesize utterances in a positive tone, for many reasons. For example, a task-oriented system should be able to clearly mark statements that convey something good, such as completing a successful authentication step, finding a way to satisfy the user's request, or even just having understood what the user said; a tutoring systems should be able to praise the user for a correct answer or a thoughtful guess; and chat system personas should be able to convey when they feel positively about some topic or entity, or in response to the user sharing a happy thought or disclosing something personal. Doing these things better, by appropriate control of the level of reduction, could help dialog systems convey positivity more clearly, more convincingly, and more naturally.

To achieve these benefits, future work should aim to create speech synthesizers that are able to vary the degree of reduction over arbitrary regions. These would, in addition to supporting applications, enable large-scale testing of whether and when reduction actually contributes to perceptions of increased positivity or other functions. Future work should also prioritize the development of tools able to estimate the degree of reduction in speech, phrase-by-phrase. These will necessarily be language-specific [Ward et al., 2025b], and so far there are just two, for English and for Spanish, available open-source at https://github.com/nigelgward/ISG_Reduction_Model. We expect that, just as F_0 estimators are, though not infallible, very useful, such reduction estimators will become widely used. For example, they will make it easy to add a term for faithfulness in reduction-level to the loss function, to enable training speech synthesizers to appropriately control the level of reduction, word-by-word. Such estimators will also greatly accelerate research on reduction, for example, by enabling automatic corpus annotations.

A remaining research question is, what type of feature is reduction? Two aspects of our findings are relevant: reduction is often suprasegmental, as are prosodic features, and reduction seems to interact in complex ways with lexical and phonetic content, as do prosodic features.

Further, reduction often seems to combine with other features to form "prosodic constructions" [Ward, 2019]. This is seen by the fact that, although several functions involve reduction, there is no reason to think they are generally confusable. This is likely because of the contributions of other prosodic features. These may include, for positive assessment, rate and loudness, and for topic closings, slow rate, long pauses, and low volume. Further, uncertainty markers may involve slow speaking rate, turn grabs may involve fast rate and high pitch, and predictable words may involve low pitch and low volume. Thus, we should probably describe reduction not as conveying these meanings by itself, but as being part of various prosodic feature configurations, constructions, which bear meanings. Thus there are good reasons to consider reduction to be a prosodic feature [Beller et al., 2008, Niebuhr, 2016], although further investigation is needed.

Future work aside, our findings already significantly broaden our view of reduction: it not just a matter of careless speech nor a mere side effect of various mental processes, but a "first-class" feature that can help convey meaning in its own right. It should therefore be added to the armamentarium of speech scientists and technology developers.

More generally, our findings indicate there are still interesting things left to discover about the features that are meaningful in speech, and that examination of dialog data, in all its richness, can be a productive strategy for this quest.

Acknowledgments

We thank Divette Marco for the second set of reduction annotations, for help in the initial analysis for Spanish, and for help with the pilot studies. We thank Oliver Niebuhr, Jonathan Avila, Natasha Warner, Visar Berisha, Francesco Cangemi, Rory Turnbull and Harm Lameris for discussion. This work was supported in part by the National Science Foundation through award 2348085, and by the AI Research Institutes program of the NSF and the Institute of Education Sciences, U.S. Department of Education, through Award #2229873 – National AI Institute for Exceptional Education.

References

A. Adigwe, S. Wallbridge, and S. King. What do people hear? Listeners' perception of conversational speech. In *Interspeech*, pages 1210–1214, 2024.

L. Aguilar, B. Blecua, M. Machuca, and R. Mann. Phonetic reduction processes in spontaneous speech. In *Third European conference on speech communication and technology (Eurospeech)*, pages 433–436, 1993.

L. Aroyo and C. Welty. Truth is a lie: Crowd truth and the seven myths of human annotation. *AI Magazine*, 36(1):15–24, 2015.

A. Bell, D. Jurafsky, E. Fosler-Lussier, C. Girand, M. Gregory, and D. Gildea. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113:1001–1024, 2003.

A. Bell, J. M. Brenier, M. Gregory, C. Girand, and D. Jurafsky. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60:92–111, 2009.

G. Beller, N. Obin, and X. Rodet. Articulation degree as a prosodic dimension of expressive speech. In *Speech Prosody*, pages 681–684, 2008.

P. Birkholz, L. Martin, Y. Xu, S. Scherbaum, and C. Neuschaefer-Rube. Manipulation of the prosodic features of vocal tract length, nasality and articulatory precision using articulatory synthesis. *Computer Speech & Language*, 41:116–127, 2016.

K. Bodur, R. Bertrand, J. German, S. Rauzy, C. Fredouille, and C. Meunier. Speech reduction: position within French prosodic structure. In *Interspeech*, pages 117–121, 2023.

R. S. Burdin and C. G. Clopper. Phonetic reduction, vowel duration, and prosodic structure. In *International Congress of the Phonetic Sciences*, 2015.

F. Cangemi and O. Niebuhr. Rethinking reduction and canonical forms. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, and M. Zellers, editors, *Rethinking reduction*, pages 277–302. de Gruyter, 2018.

L. Chen, C. Jiang, Y. Gu, Y. Liu, and J. Yuan. Automatically detecting reduced-formed English pronunciations by using deep learning. In *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*, pages 22–26, 2022.

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

F. Eyben, M. Wöllmer, and B. Schuller. OpenSmile: The Munich versatile and fast open-source audio feature extractor. In *Proceedings, International Conference on Multimedia*, pages 1459–1462, 2010.

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

J. Gustafson, E. Székely, S. Alexandersson, and J. Beskow. Casual chatter or speaking up? Adjusting articulatory effort in generation of speech and animation for conversational characters. In *Workshop on Socially Interactive Human-Like Virtual Agents*, 2023a.

J. Gustafson, É. Székely, and J. Beskow. Generation of speech and facial animation with controllable articulatory effort for amusing conversational characters. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, pages 1–9, 2023b.

L. Hall-Lew. Phonetic cues to depression: A sociolinguistic perspective. *Language and Linguistics Compass*, 18(5):e12529, 2024.

S. Hawkins. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31(3-4):373–405, 2003.

B. S. Helfer, T. F. Quatieri, J. R. Williamson, D. D. Mehta, R. Horwitz, and B. Yu. Classification of depression state based on articulatory precision. In *Interspeech*, pages 2172–2176, 2013.

L. Horn. Economy and redundancy in a dualistic model of natural language. In *SKY: Yearbook*, volume 1993, pages 33–72. Linguistic Association of Finland, 1993.

W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed. HuBert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3451–3460, 2021.

W.-C. Huang, B. Peloquin, J. Kao, C. Wang, H. Gong, E. Salesky, Y. Adi, A. Lee, and P.-J. Chen. A holistic cascade system, benchmark, and human evaluation protocol for expressive speech-to-speech translation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

T. F. Jaeger and E. Buz. Signal reduction and linguistic encoding. *The Handbook of Psycholinguistics*, pages 38–81, 2017.

Y. Jiao, V. Berisha, J. Liss, S.-C. Hsu, E. Levy, and M. McAuliffe. Articulation entropy: An unsupervised measure of articulatory precision. *IEEE Signal Processing Letters*, 24(4):485–489, 2016.

K. Johnson. Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis, Proceedings of the 10th International Symposium*, pages 29–54, 2004.

D. Jurafsky, A. Bell, E. Fosler-Lussier, C. Girand, and W. D. Raymond. Reduction of English function words in Switchboard. In *International Conference on Spoken Language Processing*, 1998.

J. M. Kahn and J. E. Arnold. Articulatory and lexical repetition effects on durational reduction: Speaker experience vs. common ground. *Language, Cognition and Neuroscience*, 30(1-2):103–119, 2015.

J. Koreman. Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *The Journal of the Acoustical Society of America*, 119(1):582–596, 2006.

H. Lameris, J. Gustafsson, and É. Székely. Beyond style: Synthesizing speech with pragmatic functions. In *Interspeech*, pages 3382–3386, 2023.

S. Lee, E. Bresch, J. Adams, A. Kazemzadeh, and S. Narayanan. A study of emotional speech articulation using a fast magnetic resonance imaging technique. In *Ninth International Conference on Spoken Language Processing*, 2006.

N. Levshina and D. Lorenz. Communicative efficiency and the principle of no synonymy: Predictability effects and the variation of want to and wanna. *Language and Cognition*, 14:249–274, 2022.

G.-T. Lin, C.-L. Feng, W.-P. Huang, Y. Tseng, T.-H. Lin, C.-A. Li, H. Lee, and N. G. Ward. On the utility of self-supervised models for prosody-related tasks. In *IEEE Workshop on Spoken Language Technology (SLT)*, pages 1104–1111, 2022.

B. Lindblom. Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle and A. Marchal, editors, *Speech production and speech modelling*, pages 403–439. Kluwer, 1990.

J. Local and G. Walker. How phonetic features project more talk. *Journal of the International Phonetic Association*, 42:255–280, 2012.

N. Lubold, S. A. Borrie, T. S. Barrett, M. M. Willi, and V. Berisha. Do conversational partners entrain on articulatory precision? In *Interspeech*, pages 1931–1935, 2019.

P. Machač and M. Fried. Utterance comprehension in spontaneous speech: phonetic reductions and syntactic context. manuscript, Charles University, Prague, 2023.

P. Machac and M. Fried. Utterance comprehension in spontaneous speech: Phonetic reductions and lexico-grammatical context. In K. Nikiforidou and M. Fried, editors, *Multimodal Communication from a Construction Grammar Perspective*, pages 69–115. John Benjamins, 2025.

M. Marge, C. Espy-Wilson, N. G. Ward, et al. Spoken language interaction with robots: Research issues and recommendations. *Computer Speech and Language*, 71, 2022. paper ID = 101225.

A. Mohamed, H.-y. Lee, L. Borgholt, J. D. Havtorn, J. Edin, C. Igel, K. Kirchhoff, S.-W. Li, K. Livescu, L. Maaløe, et al. Self-supervised speech representation learning: A review. *IEEE Journal of Selected Topics in Signal Processing*, 16(6):1179–1210, 2022.

D. S. R. Mohan, V. Hu, T. H. Teh, A. Torresquintero, C. G. Wallis, M. Staib, L. Foglianti, J. Gao, and S. King. Ctrl-p: Temporal control of prosodic variation for speech synthesis. In *Interspeech*, pages 3875–3879, 2021.

K. Mulder, L. Wloch, L. Boves, L. ten Bosch, and M. Ernestus. Cognate status modulates the comprehension of isolated reduced forms. *Language, Cognition and Neuroscience*, 37(5):576–614, 2022.

B. Munson and N. P. Solomon. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, 47:1048–1058, 2004.

O. Niebuhr. On the phonetics of intensifying emphasis in German. *Phonetica*, 67:170–198, 2010.

O. Niebuhr. 'A little more ironic': Voice quality and segmental reduction differences between sarcastic and neutral utterances. In *7th Speech Prosody Conference*, pages 608–612, 2014.

O. Niebuhr. Rich reduction: Sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale. In *7th Tutorial & Research Workshop on Experimental Linguistics, ExLing*, pages 11–24. International Speech Communication Association (ISCA), 2016.

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

O. Niebuhr, K. Gors, and E. Graupe. Speech reduction, intensity, and F0 shape are cues to turn-taking. In *SigDial*, pages 261–269, 2013.

B. Picart, T. Drugman, and T. Dutoit. Analysis and HMM-based synthesis of hypo and hyperarticulated speech. *Computer Speech & Language*, 28(2):687–707, 2014.

L. Plug. From words to actions: the phonetics of eigenlijk in two communicative contexts. *Phonetica*, 62 (2-4):131–145, 2005.

L. Plug. Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. *Journal of Phonetics*, 39(3):289–297, 2011.

L. Schettino and F. Cutugno. Quantifying and characterizing phonetic reduction in italian natural speech. *Languages*, 10(1):14, 2025.

A. Schmitt and S. Ultes. Interaction quality: Assessing the quality of ongoing spoken dialog interaction by experts—and how it relates to user satisfaction. *Speech Communication*, 74:12–36, 2015.

L. M. Schubotz, N. H. Oostdijk, and M. T. Ernestus. Y'know vs. you know: What phonetic reduction can tell us about pragmatic function. In S. Lestrade, P. de Swart, and L. Hogeweg, editors, *Addenda. Artikelen voor Ad Foolen*, pages 361 – 380. Nijmegen: Radboud Universiteit Nijmegen, 2015.

M. Tu, A. Grabek, J. Liss, and V. Berisha. Investigating the role of L1 in automatic pronunciation evaluation of L2 speech. In *Interspeech*, pages 1636–1640, 2018.

R. Turnbull. Patterns of individual differences in reduction: Implications for listener-oriented theories. In *International Congress of the Phonetic Sciences*, 2015.

P. Wagner, J. Beskow, S. Betz, J. Edlund, J. Gustafson, G. Eje Henter, S. Le Maguer, Z. Malisz, É. Székely, C. Tånnander, et al. Speech synthesis evaluation: State-of-the-art assessment and suggestion for a novel research program. In *Proceedings of the 10th Speech Synthesis Workshop*, 2019.

N. G. Ward. Non-lexical conversational sounds in American English. *Pragmatics and Cognition*, 14: 113–184, 2006.

N. G. Ward. *Prosodic Patterns in English Conversation*. Cambridge University Press, 2019.

N. G. Ward and J. A. Jodoin. A prosodic configuration that conveys positive assessment in American English. In *International Congress of the Phonetic Sciences*, pages 3368 – 3372, 2019.

N. G. Ward and C. A. Ortega. Preliminaries to a study of the pragmatic functions of reduced articulatory precision in dialog. Technical Report UTEP-CS-24-23, University of Texas at El Paso, 2024.

N. G. Ward, J. E. Avila, E. Rivas, and D. Marco. Dialogs re-enacted across languages, version 2. Technical Report UTEP-CS-23-27, University of Texas at El Paso, Department of Computer Science, 2023.

N. G. Ward, D. Marco, and O. Fuentes. Which prosodic features matter most for pragmatics? *IEEE ICASSP*, 2025a.

N. G. Ward, J. Vazquez, and E. Boushka. A tool for estimating the perceived phonetic reduction level. manuscript, 2025b.

N. Warner. Reduced speech: All is variability. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(4): e1496, 2019.

C. J. Wynn, E. R. Josephson, and S. A. Borrie. An examination of articulatory precision in autistic children and adults. *Journal of Speech, Language, and Hearing Research*, 65(4):1416–1425, 2022.

S.-w. Yang, P.-H. Chi, Y.-S. Chuang, C.-I. J. Lai, K. Lakhotia, Y. Y. Lin, A. T. Liu, J. Shi, X. Chang, G.-T. Lin, et al. Superb: Speech processing universal performance benchmark. In *Interspeech*, pages 1194 – 1198, 2021.

M. Zellers. Prosodic variation and segmental reduction and their roles in cuing turn transition in Swedish. *Language and Speech*, pages 454–478, 2017.

M. Zellers, B. Schuppler, and M. Clayards. Introduction, or: Why rethink reduction? In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, and M. Zellers, editors, *Rethinking reduction*, pages 1–24. de Gruyter, 2018.