Rich Reduction

Sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale

Niebuhr, Oliver

Published in:
Proceedings of the 7th Tutorial & Research Workshop on Experimental Linguistics

Publication date:
2016

Document license
CC BY-ND

Citation for published version (APA):

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to pureresupport@bib.sdu.dk

Download date: 11. Dec. 2019
Rich Reduction: Sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale

Oliver Niebuhr
Institute of Design & Communication, IRCA, University of Southern Denmark, Denmark

Introduction

Managing and, ideally, explaining phonetic variation has ever since been a key issue in the speech sciences. But, it became even more obvious with the beginning of the "acoustic age" after World War II, when the US military declassified the invention of the sound spectrograph. It made speech a precisely analyzable research object. The rapidly developing computer technology made this research object accessible to a growing community of phoneticians (Mattingly 1999), which, in turn, multiplied the number of questions on phonetic variation and their levels of detail and complexity. Phonetic variation supported the development of phonetics and phonology as two different disciplines and later expedited the "divorce" of those disciplines, with phonology taking care of the well-formed structures of clearly defined sound (or intonation) categories and their rule-based changes, and with phonetics measuring the messy, highly variable articulatory and acoustic signals and trying to project them across speakers, genders, speaking styles, and communicative situations onto the "ideal-world" categories of phonology. In this context, it was not surprising that phonology soared to dominate phonetics for decades in the 20th century, and that the joint efforts of the two disciplines primarily aimed at marginalizing or abstracting away from phonetic variation by searching for the invariant characteristics of sound (or intonation) categories.

The motor theory of speech perception, which attracted a lot of attention during that time, is a role model of these efforts (which is not to say that the basic idea of "covered mimicry" has no empirical foundation, see, e.g., Watkins et al. 2003). This theory saw phonetic variation as a troublemaker. Thus, the aim of both listeners and researchers had to be to free the sound segments from their variable acoustic ingredients and the resulting "encumbering auditory baggage that would make them all but useless for their proper role as vehicles of language" (Liberman 1982:148). The research paradigm of categorical speech perception nicely reflects this approach to phonetic variation (see Holt & Lotto 2010 for a summary).

The later-emerging articulatory phonology (Browman & Goldstein 1992) is similar to the motor theory in that it is also rooted in the articulatory
domain, assuming that consistent patterns are to be found only in articulation, whereas the acoustic patterns of speech are intrinsically variable. However, unlike the motor theory, articulatory phonology was explicitly designed to explain phonetic variation. That is, allophonic variation of speech sounds, for example in terms of voicing/VOT, were made conceptually understandable by means of (changes in) the temporal coordination of glottal and supraglottal gestures, conflicting articulatory commands to the same articulator were used to explain strong coarticulation and the blending of sounds, and the disappearance of sounds at the acoustic level were attributed to an extreme overlap of supraglottal gestures, hence postulating that they are only "masked" but nonetheless consistently there at the level of articulation.

Despite its deserved success, articulatory phonology was also criticized, among others by Kohler (1992). One of his major points was that the rules and restrictions according to which the gestural score is organized are probably not able to explain the full range of variations, especially those that relate to common strong speech reduction patterns in spontaneous speech. Moreover, Kohler points out that the rules and restrictions on which articulatory phonology is built themselves need to be externally motivated and supported by independent empirical evidence. Kohler's suggestion in this context is to go beyond the speaker and explain phonetic variation, i.e. its sources as well as its implications for communication, by means of a theoretical framework that also takes into account the listener and his/her cognitive abilities and processes.

The latter is exactly what was done by Lindblom (1990) in his very influential H&H theory. "Explaining phonetic variation" (p.403) is the explicit aim of Lindblom's theory. It compares speech communication to a tug-of-war, with speaker and listener pulling the rope that represents phonetic variation in opposite directions, see Figure 1. The speaker follows a basic ethological principle of all mammals, i.e. striving for economy. Accordingly, the speaker's aim is to minimize the articulatory effort invested in speech production and hence reduce the speech signal as much as possible. The extent to which this is possible is defined by the listener at the other end of the rope: The speech signal has to contain at least enough phonetic information to allow the listener understand the message conveyed by the speaker. In other words, speakers want to produce "hypospeech", and listeners want to hear "hyperspeech".
On this basis, the key concept of the H&H theory is that, at each point of the conversation, the level of speech reduction is an implicitly negotiated compromise along the hypo-hyper scale between speaker desires and listener demands. A further key concept is that this dynamic, adaptive compromise takes into account not only basic factors like speaker physiology (e.g., gender, emotions, pathologies) and the environmental acoustics of the communication situation. The compromise is also made with respect to the listener's metalinguistic top-down knowledge and context-driven expectation about which units, functions, and meanings will be contained in the upcoming speech signal. This allows the speaker to be less clear in or even completely omit those acoustic cues which s/he knows that the listener can add in the process of speech perception. This idea was probably the H&H theory’s most important contribution to speech sciences. It replaces invariance by sufficient contrast and hence goes beyond the common picture of speech as a machine-like self-contained code that is encoded on the side of the speaker and transmitted through the air with all elements that the listener requires to decode it. In contrast, all that speakers need to do in Lindblom’s framework is, broadly speaking, to be sufficiently clear, feed their listeners with a sufficient number of acoustic cues, and then let their top-down processes do the rest, i.e. interpret the signal by matching it against knowledge and expectations, and, if necessary, fill in gaps.

Many studies provide empirical support for the H&H theory. For example, Hunnicutt (1985) concluded from the results of a combined production-perception experiment that speakers hyperarticulate more if words are less predictable in a given semantic (sentence-frame) context. Fowler & Housum (1987) showed by means of radio news broadcasts that repeatedly stated words are more hypoarticulated (i.e. reduced) by speakers.
Similarly, Wright (2003) found "easy" words, i.e. frequent words with relatively few lexical competitors, to be more strongly hypoarticulated than "hard" words. Finally, we know from a number of experiments that speech produced under adverse conditions such as noise or greater spatial distances between the dialogue partners is produced with more effort both articulatorily and phonatorily (Traunmüller & Erickson 2000; Junqua 1996).

Despite this converging evidence in favor of H&H, we should not lose sight of one crucial fact: Lindblom's framework never aimed at explaining phonetic variation in general. Rather, the framework was developed to explain that phonetic variation that is relevant to and emerges in connection with "successful lexical access" (Lindblom 1990:405). However, we know at least since the rise of intonational phonology (Ladd 2008) that speech communication is not only about words. Lindblom himself notes that speech is "produced not only in the laboratory but also in its natural, ecological settings" (p.418), and he stresses in this context that the assumption of only two antagonistic forces that create the one-dimensional reduction continuum from hypo to hyper is a "deliberate simplification that is likely to be revised in the course of future work" (p.419).

In fact, Lindblom's H&H theory was taken up and further elaborated, for example, in terms of the smooth signal redundancy hypothesis of Aylett & Turk (2004). In simple terms, the hypothesis states that the total degree of reduction used by speakers is understandable as the sum of two types of redundancy: language redundancy (e.g., due to syntactic order or grammatical agreement) and signal redundancy (e.g., several acoustic cues on the same phonological distinction). Aylett & Turk assume that speakers strive to keep the total redundancy constant, which means that a lower language redundancy is compensated by a higher signal redundancy (i.e. hyperspeech), whereas a higher language redundancy allows for a lower signal redundancy (i.e. hypospeech). As is obvious from these explanations, Aylett & Turk refined rather than revised Lindblom's H&H framework, keeping intact the central tug-of-war metaphor and its two antagonistic forces, which are called "conservation of effort" and "reliable communication" in Aylett & Turk's terminology. The same applies to many other works and concepts that are inspired by H&H, such as uniform information density, communicative efficiency, and audience design, see Clopper & Turnbull (submitted) for a summary.

In summary, despite Lindblom's own expectation, his deliberate simplification of a one-dimensional hypo-hyper scale has not been significantly addressed. The present paper is intended to pave the way for
initial steps in this direction by pointing the readers to two basic aspects in which H&H, in the author's own humble opinion, overestimates and oversimplifies variation in the degree of reduction.¹

The supposed harmfulness of reduction
A key premise of the tug-of-war metaphor in Lindblom's H&H theory is that reduced articulatory effort on the side of the speaker and the resulting reduction phenomena in the speech signal² put pressure on the listener, for example, in that the listener has to rely more on his/her cognitive top-down processes to compensate for missing acoustic cues associated with reduced sound segments or entire meaningful elements. A growing body of production and perception evidence that has accumulated after the H&H theory was published raises doubts about the general validity of this tug-of-war premise.

Nolan (1992) summarizes EPG data collected by W. Barry and P. Kerswill on alveolar-to-velar place assimilation in British English stop consonant sequences like "road collapsed" and "lead covered" (/dK/). The EPG data show a range of more or less strongly assimilated productions, but also cases in without any trace of a tongue contact at the alveolar ridge. These "zero-alveolar" cases (in Nolan's terminology) are indistinguishable from "non-alveolar" cases representing actual "rogue collapsed" and "leg covered" (/gk/) realizations. Yet, Nolan found in a 2AFC word-identification test (based on newly recorded stimuli) that listeners are to a significant degree able to perceive the assimilated /d/ of which there is no EPG trace left and thus keep zero-alveolar cases of "road collapsed" and "lead covered" separate from non-alveolar cases of "rogue collapsed" and "leg covered". These striking results led Nolan to the conclusion that, even in the absence of an alveolar contact or closure, "the tongue configuration in realizing lexical /dg/ sequences [...] is subtly different from that for /gg/ sequences" (Nolan 1992:272). This subtle difference persists as a difference in vowel quality that functions as an acoustic cue to /d/ even when measurements suggest that this sound segment itself has fully disappeared. Later, an acoustic analysis by Local (2003) provided supporting evidence for Nolan's assumption and his impression that, "auditorily, [...] the vowel allophone

¹ All references involving Niebuhr in the following sections 2 and 3 that are not including in the list of references can be found in Cangemi et al. (submitted).
² Note that there are actually two different types of reduction: (1) the amount of energy invested in articulation and phonation, and (2) deviations from full/ideal citation forms of consonants, vowels, and words. The two types are equated here. The author is aware of the fact that this is probably a simplification (Yi Xu, pers. comm.), but one that does not affect the line of argument presented here.
before the lexical velar is slightly closer than before the lexical alveolar" (Nolan 1992: 272).

A very similar case to that of Nolan (1992) and Local (2003) was found for French by Niebuhr & Meunier (2011). They investigated /s/-to-[ʃ] place assimilation in word sequences like "trousse chargée" (full bag) and "fils charmant" (charming son). Their acoustic spectral center-of-gravity measurements revealed a production continuum from weakly to fully assimilated /s/ sibilants. Moreover, Niebuhr & Meunier found differences in the preceding vowel that were there independently of the degree of /s/-to-[ʃ] assimilation and even remained if the original /sʃ/ sequence was acoustically indistinguishable from an actual /ʃʃ/ sequences. Vowels preceding /s/ were shorter, had a higher acoustic energy, and a less breathy voice than those before /ʃ/. A later pilot perception experiment (Clayards & Niebuhr 2011) based on the identification of pseudo names in forename-surname sequences demonstrated, as in Nolan (1992), that listeners used these vowel cues to identify a /s/ even if the sound segment itself became (according to the measurements taken) indistinguishable from /ʃ/.

Examples like those above stress the relevance of a concept that was developed by Kohler (1990) and is hence as old as the H&H theory: "Articulatory prosodies". At the heart of the concept lies the statement that reduction processes do not necessarily cause a loss of acoustic cues and in this way undermine the richness of the speech signal. In spontaneous speech, reduction is the rule rather than the exception, and the phenomena subsumed under reduction represent processes by which the packaging scheme of acoustic cues in the form of sequences of linear sound segments is broken up. The affected sound characteristics or acoustic cues are then reshaped as long-term resonances, i.e. articulatory prosodies, that are superimposed on the remaining sound segments. Niebuhr (2008) elaborated this concept by adding the notion of "phonetic essence", see Niebuhr & Kohler (2011) and Kohler & Niebuhr (2011). Phonetic essence is a feature of complex sound sequences like words, and the assumption is that, in speech reduction, those sound characteristics of the sequences are maintained and reshaped as articulatory prosodies that belong to the sequence's phonetic essence.

For instance, the German modal particle "eigentlich" (actually) is characterized by palatality that pervades virtually the entire word: [œiqntlc]. An analysis of the Kiel Corpus of Spontaneous Speech (Peters 2005) showed that "eigentlich" can be severely reduced, with only the initial diphthong and, maybe, the middle nasal being left at the segmental level: [æiʃ(ɲ)]. However, in these cases the palatality of the lost sound segments is maintained by strengthening and lengthening the palatality in the initial diphthong. That is, the closed-vowel element is produced longer and with a higher F2 frequency. A perception experiment conducted by Niebuhr &
Kohler (2011) showed that listeners have no problems interpreting this articulatory prosody of palatality and distinguishing highly reduced "eigentlich" from the segmentally similar unreduced word "ein" (indef. article). Likewise, the study Kohler & Niebuhr (2011) addressed the word "ihnen" - [iːnʲnʲ] (to you) - whose separate segmental representation completely disappeared in the sentence frame "ich kann ihnen das ja mal sagen" (I can mention this to you) produced by speaker TIS in the Kiel Corpus of Spontaneous Speech. Despite the loss of all segments, the phonetic essence of palatality of "ihnen" was kept and superimposed by the speaker on the segments of "kann" and "das" that, as a result, change from [kʰannas] to [k̟ʰɛ̈nʲnʲə̟s]. Evidence from a perception experiment showed that listeners can reliably perceive the entire word "ihnen" on this basis of [k̟ʰɛ̈nʲnʲə̟s] in the sentence frame "Ich ___ ja mal sagen". Moreover, as the sound segments of [k̟ʰɛ̈nʲnʲə̟s] were successively replaced by those of [kʰa̠nna̠s], the perceived wording of the stimulus sentence changed to "ich kann das ja mal sagen" (I can mention this), without "ihnen".

Further phonetic essences that are reshaped as articulatory prosodies and whose perceptual relevance was been experimentally demonstrated are velarization (Niebuhr 2008), glottalization (Kohler 1999), and lip rounding (Niebuhr & John 2014). In all these examples, the articulatory prosodies were reduced representatives of at least entire syllables, and in the case of Niebuhr (2008) the velarization even represented two full words, i.e. "auch noch" (as well).

Articulatory prosodies almost always co-occur with duration cues in the form of a compensatory lengthening of segments in the vicinity of disappeared segments. However, while articulatory prosodies are sufficient to make listeners perceive segmentally disappeared syllables or words, mere segmental lengthening is not sufficient (cf. Niebuhr & Kohler 2011). It must be temporally coordinated with the articulatory prosodies and/or affect those remaining sound segments that reflect the relevant phonetic essence in order to function as a cue to disappeared syllables or words. Therefore, the question of whether duration or segmental lengthening may be considered a separate articulatory prosody is not yet settled. Interesting in this context is the work of Dilley & Pitt (2010). They manipulated the relative duration of vowels like [ɔː] in the middle of phrases like "leisure time", for example, by means of lengthening the vowel. The results of the corresponding perception experiment showed that this change in relative duration makes listeners perceive an additional "or" in the between the two words. That is, "leisure time" became "leisure or time", without adding any further phonetic sound features.

The question has been raised whether this duration-based appearance of words is a one-step all-or-nothing phenomenon, or whether the number of
additionally perceived words is correlated with the degree of the increase in relative duration. This question was addressed by Dilley together with Evelin Graupe and the author of this paper in a joint study on German (cf. Graupe et al. 2014). The starting point was the fact that, in German, there are several function words each of which can be reduced to a single alveolar nasal [n]. This includes, for example, "in", "ihn", "den", "einen", and "denn". Niebuhr et al. designed stimulus sentences (rhetorical questions) like "Wer braucht Nachrichtensprecher im Radio?" (Who needs newsreaders on the radio) and "Wer findet Nebendarsteller erwähnenswert?" (Who finds supporting actors worth mentioning) and manipulated the relative duration of the initial alveolar nasal of the target nouns, i.e. "Nachrichtensprecher" (newsreader) and "Nebendarsteller" (supporting actor). The semantic contexts of the stimulus sentences basically allowed the relative duration manipulation of /n/ to trigger the appearance of two additional words: "denn" (then, intensifying particle), which corresponds to one syllable, "einen" (indef. article), which corresponds to two syllables, or "denn einen", which corresponds to three syllables. The results clearly show that the stronger the relative duration increase of /n/, the more syllables are perceived by listeners. Weak /n/ lengthening makes the monosyllabic word "denn" appear, strong lengthening the disyllabic word "einen", and very strong lengthening triggers the perception of the whole trisyllable "denn einen".

Results like these emphasize once more that even the most severe segmental reduction need not make the speech signal poorer and ambiguous and speech perception a harder or impossible task for listeners. Moreover, the gradient relationship of relative segment duration and the number of appearing syllables suggests that duration is in fact another independent articulatory prosody rather than just a concomitant feature of palatality, velarization, glottalization, and lip rounding.

**Meaningful variation in reduction**

The tug-of-war metaphor implies that there are only two parties pulling on the rope: speaker or "articulatory economy" and listener or "sufficient signal contrast". Lindblom himself calls this a deliberate simplification; and, in fact, a growing body of evidence supports this assessment. There is at least one more factor that shifts the degree of speech reduction along the hypo-hyper scale: communicative function.

For example, in the domain of prosody, Niebuhr (2008, 2012) showed that the phonetics of voiceless fricative sounds co-varies with intonation such that the spectral-energy distribution and resulting spectral-pitch impression they convey fit in with the level of the adjacent F0. That is, voiceless fricatives sound "brighter" in high-F0 contexts due to more acoustic energy at higher noise frequencies, and "darker" in low-F0 contexts.
due to more acoustic energy at lower noise frequencies. Given the fact that the postalveolar sibilant [ʃ] is creates an intrinsically "darker" sound than the alveolar sibilant [s] (also because [ʃ] is produced with lip rounding in German), Niebuhr et al. (2011) wondered whether the degree of /s/-to-[ʃ] (i.e. "bright"-to-"dark") assimilation in German would be affected by the F0 context. The speech-production study they conducted with native speakers of German confirmed this prediction. The degree of /s/-to-[ʃ] assimilation, determined on the basis of spectral center-of-gravity measurements, was weaker in high F0-peak contexts, in this way giving the entire sibilant sequence a "brighter" sound quality. Assimilation was stronger and made the sibilant sequence sound overall "darker" in low F0-valley contexts. Note that the sibilant sequence's total duration did not differ between significantly between the F0-peak and F0-valley contexts. This fact supports the assimilation interpretation as it means that individual sibilants were not simply produced longer or shorter. Together with the results of perception experiments on the integration of F0-based pitch and fricative-based pitch impressions (Mixdorff & Niebuhr 2013; Welby & Niebuhr 2016), the findings of Niebuhr et al. (2011) show that assimilation and hence speech reduction vary in order to support conveying intonational meanings.

An even better demonstration of the fact that reduction patterns are systematically related to communicative functions are the studies of Local et al. (1986). They examined word-final /ptk/ in British English whose realizations can vary from unreduced post-aspirated stops to highly reduced stretches of glottalization. Local et al. scrutinized the claim in the literature that these differences in the degree of reduction are just random, i.e. free variation (cf. Kreidler 1989). In fact, their analysis of a dialogue corpus of Tyneside English revealed quite the opposite: With only a few counterexamples (1-3%, n=206) unreduced stop variants occurred in turn-final position, whereas all reduced variants, including glottalized ones, were produced turn-internally. Docherty et al. (1997) replicated the findings of Local et al. for a corpus Southern Standard British English. Zellers (in press) analyzed phrase-final consonants in a Swedish dialogue corpus and found that "consonant reduction [...] can further help distinguish between turn change and turn hold contexts in Swedish conversation".

Niebuhr et al. (2013) took up the findings of Local et al. and showed for the Kiel Corpus of Spontaneous Speech that the degree of reduction of the most frequent word ending in German, <-en> /ǝn/, is highly systematically linked with the distinction between phrase and turn boundaries. Among the approximately 5,700 analyzed tokens, the number of schwas was higher in turn-final <-en> realizations, and even if the schwa was absent, the majority of /n/ nasals showed no place assimilation with the preceding consonant. The opposite applies to turn-internal <-en> realizations whose final nasals were
assimilated to either [m] or [ŋ] in about 70% of all cases. Going beyond the studies on English and Swedish cited above, Niebuhr et al. also conducted a perception experiment with degree of <-en> reduction as an independent variable. The experiment was based on a discourse completion task and showed that listeners waited longer with taking the turn and responding to the preceding stimulus if the latter ended in an unreduced <-en> ending. Niebuhr et al. conclude in view of this behavioral evidence that the degree of word-final reduction has a discourse organization - or, more specifically - a turn-taking function. Interestingly, the production data gathered so far point to a language-specific form-function mapping. While in German and probably also in English less or no reduction signals a speaker's turn-yielding intention (in combination with other prosodic cues), Zeller's data suggest the same function could be cued by stronger reduction in Swedish. This is another argument in favor of the non-mechanistic, function-driven nature of phrase-final reduction patterns.

A final example for meaningful variation in reduction is related to speaker attributes or attitudes. Trede (2011) conducted a production study in which she analyzed the phonetic exponents of sarcastic irony in German by means of a comparison of two sets of sarcastic and sincere utterances. Trede replicated previous results (e.g, Bryant 2010) in that she found sarcastic utterances to have lower speaking rates and lower and less variable F0 and intensity contours than sincere utterances. In addition, she counted the number of reductions (assimilations, elisions, lenitions) in each utterance by relating the words' actual realizations to their canonical reference forms. This data showed that sarcastic irony is not only marked by stronger prosodic reductions, but also by stronger segmental reductions. Informal perception tests suggest that these stronger segmental reductions represent a separate cue to sarcasm. This preliminary conclusion fits in well with unpublished perception findings of Niebuhr (in preparation) showing that strong reduction patterns made utterances sound less sincere. These unpublished data also show that the degree of segmental reduction is significantly positively correlated with a speaker's perceived level of education, clumsiness, scattiness, tiredness, and vanity.

"Offshoring" the tug-of-war metaphor

The examples provided in section 2 of this paper aimed to show that reduction, even when it eliminates entire syllables or words in spontaneous speech, does not necessarily pose a challenge for listeners. This troublemaker-view on reduction is driven by the long-established concepts of 'phoneme' and 'canonical form' both of which are currently being controversially discussed (Cangemi et al, submitted). For instance, a phoneme-based, segment-oriented perspective on speech with full canonical
word forms at the starting and end points of the speech chain overlooks that critical features of deleted sound segments can still be present in the form of articulatory prosodies, and that canonical forms may not always be proper references, for example, in the sense of the most frequent realization of a word in spontaneous speech.

Furthermore, a second set of examples in section 3 this paper illustrated that the degree of reduction cannot consistently be conceptualized as the result of the two antagonistic forces 'articulatory economy' and 'sufficient contrast', dynamically negotiated on the basis of environmental, social, psychological, and maybe pathological factors. Rather, variation in the degree of reduction can also be meaningful. That is, communicative functions at the levels of intonation, discourse organization, and speaker attitudes/attributes are associated with systematic reduction differences; and, for an increasing number of these production studies, perception experiments show that listeners process and use these differences like any other segmental or prosodic cues in order to identify the corresponding communicative function.

The major contribution of Lindblom's H&H theory was to replace the futile search for invariance by an explainable variance based on the tug-of-war metaphor. The notion of articulatory prosodies and the functional role of reduction both suggest the next steps along the line of argument opened up by Lindblom. Specifically, we need to supplement Lindblom's explanatory framework and revise the speaker-listener conflict that lies at the heart of the tug-of-war metaphor.

The author's suggestion would be to "offshore" the tug-of-war metaphor and replace it by the ocean metaphor of Bolinger (1964), with the ups and downs at the surface of the ocean representing the speaker's variation along the hypo-hyper scale and wavelength corresponding to the time domain of the reduction variation. As is illustrated in Figure 2, the ups and downs are the combined result of tides, waves, and ripples. Tides are long-term settings in the degree of reduction determined by, for example, the communication channel, the situation, the physiological and pathological properties of speaker and listener and the (acoustic) environment in which their communication takes place. Waves and ripples represent additional meaningful or otherwise systematic (e.g., tailored to integrate the listener's top-down processes) short-term variations along the hypo-hyper scale, associated with phrases, words, or single sounds and syllables. This metaphor is compatible with later refinements of Lindblom's H&H theory, such as the smooth signal redundancy hypothesis.
Figure 2: Reframing the tug-of-war metaphor in the form of the ocean metaphor of Bolinger (1964).

Given the fairly incomplete empirical picture outlined sections 2 and 3, it would be premature to try to associate waves and ripples with different sources of systematic variation in speech reduction. However, as a point of departure, it seems that meaningful reduction variation due to conveying speaker attitudes/attributes as well as reduction variation reflecting the speaker's anticipation and integration of listener knowledge are both more likely to manifest themselves as waves, i.e. at the level of phrases or words, whereas the segmental reduction differences realized in connection with different intonation contours show up as ripples. Reduction variation related to discourse functions, as in the example of turn-final and turn-internal syllables in English and German, could sometimes surface as waves and sometimes as ripples, and may represent a third type of wavelength. The ocean metaphor leaves room for distinguishing additional "wavelengths", for example, longer deep-sea waves and shorter coastal waves; and finding out whether or not such additional distinctions are necessary will be one of the interesting tasks of follow-up studies on speech reduction.

In fact, the ocean perspective on reduction opens up a completely new field of questions concerning, for example, the temporal interplay (superposition, coordination, alignment) of reduction phenomena with similar/different wavelengths, the limits of wave amplitudes, correlations between types of waves and wave amplitudes as well as between wave amplitudes and the overall (sound) energy level that is fed into the wave system, and, finally, geographical and coastal (i.e. in the case of speech cultural and phonological) differences. These and many other questions have the potential to stimulate, reconsider, and inspire research in speech reduction for many more years.
Acknowledgments

First of all, I would like to thank Meghan Clayards and Meg Zellers for their useful and insightful comments on earlier drafts of this paper. Moreover, I am greatly indebted to Meg Zellers for taking the time to proof-read the paper. Finally, special thanks are due to Yi Xue, Antonis Botonis and many other participants of ExLing as well as all authors and co-editors of the "Rethinking Reduction" volume for inspiring discussions and contributions on the issue(s) of speech reduction.

References


