in press JEP:LMC

How phonological reductions sometimes help the listener

Holger Mitterer
Max Planck Institute for Psycholinguistics

Kevin Russell
University of Manitoba

In speech-production, high-frequency words are more likely to be phonologically reduced than low-frequency words. We tested in an eye-tracking experiment whether listeners can make use of this correlation between lexical frequency and phonological realization of words. Participants heard prefixed verbs in which the prefix was either fully produced or reduced. Simultaneously, they saw a high- and low-frequency verb with this prefix—plus two distractors—on a computer screen. When hearing a reduced prefix, participants were more likely to look at the high-frequency verb than when they heard a fully produced prefix. Listeners hence exploit the correlation of lexical frequency and phonological reduction and assume that a reduced prefix is more likely to belong to a high-frequency word. This shows that reductions do not necessarily burden the listener but may in fact have a communicative function, in line with functional theories of phonology.

In order for speech to convey meaning successfully, the listener must be able to decode which words the speaker intended to utter, but speakers do not make the task easy for the listener. Spontaneous speech is peppered with phonological reductions and deletions (e.g., the reduction of German ‘we have’ /wɪr habən/ to [wɪr ham]).

A long tradition in phonetics and phonology (e.g., Grammont, 1933; Lindblom, 1990) elaborated by current theories of functional or evolutionary phonology (Blevins, 2004; Boersma, 1998), sees the actual pronunciation of an utterance as resulting from a balancing act between the opposing interests of the listener, who wants clear and easy to perceive differences between words, and the speaker, who wants to minimize articulatory effort, resulting in words that are harder for the listener to distinguish.

At first glance, the ubiquity of reductions seems problematic for this picture: the speaker seems to be rather selfish, giving no weight to the needs of the listener in the supposed balancing act. There is ample evidence that listeners find it more difficult to recognize reduced words than non-reduced words (Ernestus, Baayen, & Schreuder, 2002; Sumner & Samuel, 2005; Tucker & Warner, 2007). Some have argued, however, that even when producing reductions, speakers are taking the needs and abilities of their listeners into account, for example, by reducing especially those segments that would have been difficult for listeners to perceive correctly anyway (Kohler, 1990) or reducing words only if the reduced version is perceptually very similar to the non-reduced version (Steriade, 2001).

For example, the alveolar stops /t/ and /d/ are especially subject to reduction, since the perceptual effects of changing the place of articulation are smaller for stops than for fricatives (Hura, Lindblom, & Diehl, 1992) and smaller for alveolars than for other places of articulation (Cho & McQueen, 2008). Mitterer and colleagues (Mitterer, Csépe, & Blomert, 2006; Mitterer, Yoneyama, & Ernestus, 2008)

Holger Mitterer, Max Planck Institute for Psycholinguistics, PO Box 310, 6500AH Nijmegen, The Netherlands. Kevin Russell, Linguistics Department, University of Manitoba, Winnipeg, MB R3T 2N2, Canada.

This research was supported by a grant of the Deutsche Forschungsgemeinschaft, DFG, to the first author as part of priority programme SPP 1234.

Correspondence concerning this article should be sent to: Holger Mitterer, Max Planck Institute for Psycholinguistics P.O. Box 310, 6500 AH Nijmegen, The Netherlands; or via E-mail: holger.mitterer@mpi.nl.

showed that speakers reduce segments especially in those contexts in which the listeners have a hard time noticing the change.

A common thread in these approaches is that listeners may not be burdened by reductions because they fail to note them. This cannot be the whole story, however. While listeners may often be unaware of reductions, there is no way that the auditory system can miss the difference between, for example, the full four-syllable [vərykələk] and the reduced two-syllable [vryklək] versions of the Dutch word *verrukkelijk* ‘delicious’. Additionally, the acoustic effects of the reduced stimuli used in studies like Tucker and Warner (2007) were obviously noticeable enough to impair the participant’s word recognition.

So it would seem that some reductions have the potential to impair the listener’s processing. The best that a speaker who chooses to reduce can hope for is to mitigate that damage by cleverly choosing *where* to reduce. However, some theoretical perspectives would suggest that listeners may benefit from reduction. Aylett and Turk (2006) argue that speakers reduce words that are to some extent predictable. If listeners are aware of this correlation, they should be able to use the reductions they perceive in order to make inferences about the frequencies of the intended words. This may allow listeners to resolve temporary ambiguities earlier than they would otherwise have been able to.

The relationship between frequency and reduction has been shown by many studies (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Bybee, 2001; Zipf, 1949). For example, Pluymaekers, Ernestus, and Baayen (2005) found that a wide range of Dutch affixes were more likely to be reduced in high-frequency than in low-frequency words, including the prefixes *ge-* and *ver-* that we focus on in this paper. Language learners are excellent at learning covariate relationships, such as the fact that, in stress-timed languages, shorter syllables are more likely to be part of longer words while long syllables are likely to be monosyllabic words (e.g., cap vs. captain, see Salverda, Dahan, & McQueen, 2003). Similarly, listeners might also be able to learn the correlation between word frequency and the likelihood of phonological reduction and use reduction as a cue for lexical frequency during word recognition, in the same way that they are able to draw inferences about the frequency and predictability of an upcoming word from the presence of an immediately preceding disfluency (Corley, Macgregor, & Donaldson, 2007).

Consider a listener who has heard up to the [xdr] point in the incomplete Dutch sentence: *Afgelopen nacht heeft hij veel* [xdr…] (Engl., ‘Last night he has ... a lot’). The fragment [xdr] contains the prefix *ge-* /xə/ from which the schwa has been deleted. The listener will probably recover easily from this reduction, because /xd/ is an impossible onset cluster in Dutch (see Spinelli & Gros-Balthazard, 2007). At this point partway through the word, the listener has all the information necessary to decide both that the lexical representation of the word begins with /xədr/ and that the speaker has reduced the word.

Given the beginning of the sentence *Afgelopen nacht heeft hij veel ...*, there are then two main candidate words starting with /xədr/: *gedronken* /xədrɔŋkə/ 'drank', which has a frequency count of 1098, and *gedroomd* /xədromt/ 'dreamed', which has a frequency of 492 in the CELEX corpus (Baayen, Piepenbrock, & Gulikers, 1995). According to standard accounts of lexical access (Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986; Norris & McQueen, 2008), the listener will have to wait till the vowel to be able to disambiguate fully the word that the speaker intended, but that does not mean the listener is considering both possibilities as equally likely prior to the word's uniqueness point. If, as we hypothesize, the listener uses the relation between reduction and lexical frequency during word recognition, we might find that hearing reduced /xdr/ biases the listener even more strongly toward higher-frequency *gedronken* than unreduced /xədr/ does.

We will test these predictions using the visual-world paradigm. In this paradigm, listeners hear a sentence while they are looking at a screen containing (typically four) pictures or written words and having the direction of their gaze recorded with an eye-tracker. The participants' task is to click on one of the pictures or words based on the content of the auditory sentence, but their ultimate responses in this task are less interesting than which pictures or words their eye-gaze is directed toward while they are in the middle of processing the sentence. Over a decade of research since Allopenna et al. (1998) has found that the average proportion of time that a group of listeners spend looking at one of the pictures or written words on the screen is closely related to the activation strength of the corresponding candidate word as predicted by the major models of word recognition (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Righi, Blumstein, Mertus, & Worden, 2010).

So eye-tracking seems well suited to test the prediction that reduction will affect the lexical activation of high- and low-frequency words differently. If our hypothesis is correct, hearing a reduced prefix should make listeners look more at the high-frequency words, while hearing canonical forms should make listeners look more at the low-frequency words.

Method

Participants

Forty subjects from of the participant pool of the Max Planck Institute participated in the experiment for pay. All participants were native speakers of Dutch with normal hearing and vision.

Stimuli

We chose forty pairs of high- and low-frequency Dutch past participles: twenty pairs where both verbs began with the prefix *ge-*, such as *gedronken* 'drunk' and *gedroomd* 'dreamed', and twenty with the prefix *ver-*, such as *verliefd* 'in love' and *verliederlijkt* 'neglected, run down'. Within each pair, the two participles were phonologically identical up to the onset of the syllable following the prefix and continued with a vowel that was either also identical or at least similar (e.g., /o/ and /ɔ/). The low-frequency member of each pair had a count of under 50 (mean: 12.71) in the CELEX lexical database, with the exception of the word *gedroomd* (491), which was paired with the more frequent *gedronken*

(1089). The high-frequency member had a count of at least 300 for the *ver-* participles (mean: 599) and at least 400 for the *ge-* participles (mean: 2930).

For each pair, a sentence frame was constructed such that both participles would be semantically natural completions, for example *Afgelopen nacht heeft hij veel* ___ 'Last night he has ___ a lot' for the pair *gedroomd* 'dreamed'/*gedronken* 'drunk'. The participle was the final word in 38 of the sentence frames and followed only by a particle in the other two.

To estimate predictability, an online experiment with 34 participants measured the cloze probability for the high- and low-frequency word in these sentences. Participants read the sentence frame and typed between two and seven words that could end the sentence. After this open cloze test, they also rated the cloze-probability of the high- and low-frequency word on a 7-point scale. As there was a preference for the high-frequency words in the open cloze test (high-frequency mentioned on 9.4% of the trials versus 4.4% for low frequency word) and in the rating task (high-frequency words: 5.4, low-frequency words: 4.1 on a scale from 1 to 7), we used these measures as covariates in the data analysis.

Two versions of each sentence were recorded by a female native speaker of Dutch, once with the instruction to articulate the prefixes of the participle in their full form ([xə] and [vər]) and once in their reduced form ([x] and [və]). The recorded sentences were cut at a positive zero crossing before the onset of frication of [x] and [v] at the beginning of the prefix. To verify that the amount of reduction was similar for high- and low-frequency words, we performed a two-factor ANOVA on the prefix duration with the factors Reduction and Frequency (see Table 1). The results showed an effect of Reduction ($F(1, 157) = 68.04$, $p < 0.001$) but neither an effect of Frequency nor an interaction (Fs < 1). A similar analysis of stem durations showed that they were independent of prefix reduction (F < 1) but longer for low-frequency words ($F(1, 157) = 13.3$, $p < 0.001$). This reflected the difference in number of segments between low- and high-frequency words. These analyses showed that the rate with which information about the stem became available is approximately similar in the four cells of the design.

	Lexical Frequency	
	Low	High
Number of Segments	8.0	7.3
Prefix Duration in ms		
Reduced Realization	120	120
Unreduced Realization	163	159
Stem Duration inms		
Reduced Realization	669	616
Unreduced Realization	675	602

Table 1. *Properties of the stimuli as used in the four cells of the experiment.*

As the speaker tended to speak the entire sentence with a faster pace when producing the reduced-prefix versions, eight variants were created for each sentence by crossing the following three factors: 1. Was the initial part of the sentence taken from a recording with a reduced or a canonically produced prefix? 2. Was the prefix reduced or canonically produced? 3. Was the target verb high- or low-frequency? This gives rise to 320 sentence stimuli.

Apparatus and Procedure

Participants were seated in a chair in front of a computer screen at 68 cm distance with their chins resting on a tower mount of an EyeLink 1000 eye-tracker from SR Research. The stimulus presentation was controlled by a different computer using ExperimentBuilder software from SR Research. After calibration, instructions presented on the computer screen told participants that they would be hearing sentences and simultaneously seeing four printed words on the computer screen. The instructions stated (correctly) that they would hear one of these four words in the sentence and that their task was to click on that word.

Each trial started with a fixation cross presented for 500 ms in the center of the screen. After that, words appeared in the center of each of the four quadrants of the screen, and the mouse cursor was made visible at the center of the screen. After a preview of one second, the auditory sentence began. The trial finished when the participant clicked on one of the words. Every ten trials, the calibration of the eye-tracker was checked.

The four words on the screen were the **target** word used in the auditory sentence (e.g., *gedronken*), its opposite-frequency pair-mate as the main **competitor** (e.g., *gedroomd*), plus two **distractors** from the pool of participles with the other prefix (e.g., *verladen* ‘shipped’ and *verlaten* ‘left’). In this way, each printed word appeared twice during the forty trials, once as a member of a target-competitor pair and once as a distractor.

Each participant heard 40 sentences, ten sentences in each combination of full vs. reduced prefix and high- vs. low-frequency target word. Across all participants, each sentence was presented equally often in each of the four conditions. Randomization was done on an individual basis.

This experiment was combined with another experiment intended to explore the findings of Mitterer and McQueen (2009) that it can be possible under some conditions to repeat items in eye-tracking experiments without altering the results. After the initial block of trials reported on here, the participants took part in additional blocks to test the effects of repeating the same stimuli. As it turned out, repetition did alter participants’ responses in later blocks in ways that will not be discussed here.[1]

Each participant hence heard ten sentences of each combination of target word (high-low frequency) and pronunciation (full vs. reduced). Within each cell, half of the carrier sentences came from an utterance with a full or a reduced production of the target word. As there was no effect of this variable we will not consider it further. After the experiment, participants filled in a questionnaire rating how frequent each of the 40 words was in their (active and passive) language use. These ratings will also be used as covariates in the data analysis.

Design and Analysis

For the purposes of time-locking the eye-gaze data, we took into account that our speaker pronounced the reduced prefixes faster than the the unreduced ones. This makes the uniqueness point for the reduced verbs fall earlier than that for the unreduced verbs. In order to correct for this, we measure all times from the onset of the verb's stem rather than the onset of the entire word (e.g., the beginning of the [d] in [xədromt] and [xdromt]).

The main factors in this experiment are the frequency of the target word, with the levels High vs. Low, and realization of the target word, with the levels Full vs. Reduced. Additional models are considered that also include as co-variates the cloze probability of the target and competitor words (from the on-line experiment) and each participant's subjective frequency ratings of the target and competitor word.

The dependent variable is the amount of time participants spent looking at the target word relative to its competitor during the time window from 200 ms to 600 ms after the onset of the verb stem. If a participant looked at the target for 100 ms on a given trial, this would lead to a fixation proportion of 0.25 (100 ms / 400 ms). The fixation proportions were transformed logistically, and zeros and ones were replaced by 0.005 and 0.995 (cf. Macmillan & Creelman, 1991). The dependent variable is the difference between the transformed target and competitor looks.

Linear mixed effect models were used to determine the relation between the dependent variable and the independent variables (Baayen, Davidson, & Bates, 2008) to allow the use of trial-specific covariates. Reported P-values were based on Monte Carlo Markov Chain simulation (MCMC, cf. Baayen, et al., 2008).

Results

Figure 1 shows how the listeners' eye-gazes changed during the course of listening to the word for all conditions. All panels show the typical three-way split for the time course of target, competitor, and distractor fixations. There was no preference for a word in the pre-target baseline; 200 ms after the onset of the stem, participants looked towards target and competitor to the same degree. By 600 ms after the onset of the stem, there was enough acoustic information for the listener to determine that the stimulus began, for instance, [xdrom...] rather than [xdrɔŋ...], so a preference for the target over the competitor arose.

Interestingly, between those two times, after the listener was sure that what the prefix was, but before they were sure what the stem was, the pattern differed between the reduced and unreduced conditions. Panel A shows the data for reduced prefixes, with a preference for high-frequency over low-frequency words (be it target or competitor), and Panel B shows the data for unreduced prefixes with no differences between high-and low frequency words. Panels C and D show the same data, but now allowing for a direct comparison between reduced and unreduced forms in one panel. Panel C shows that, for high-frequency targets, reduction did not lead to a recognition cost, while Panel D

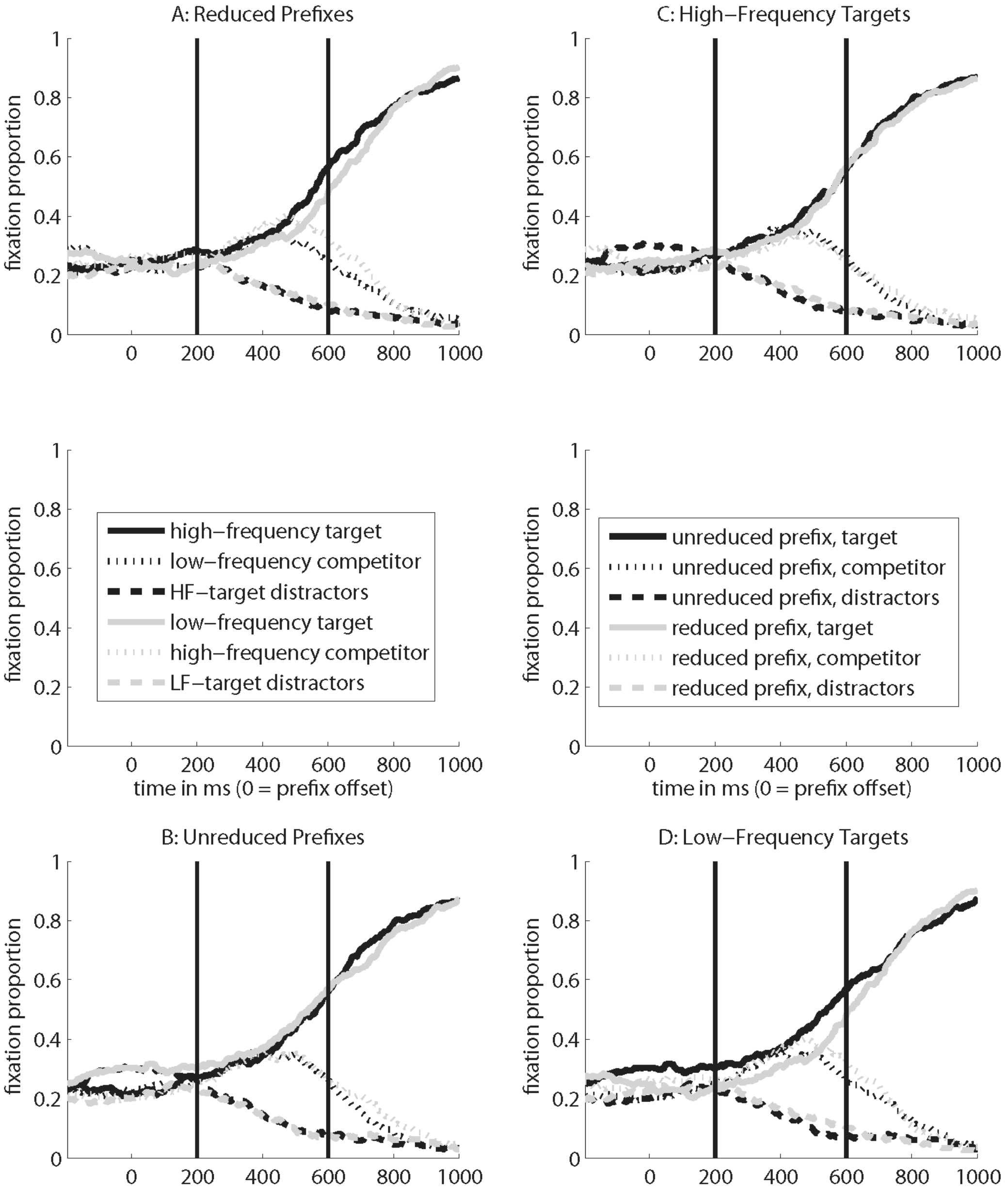

Figure 1. Comparison of fixation proportions to the targets (solid line), phonological competitors (dotted lines), and unrelated distractors.(dashed lines). Panel A compares the fixation proportions for high- and low-frequency words given a reduced prefix in the auditory input. Panel B shows the same comparison given a unreduced prefix in the input. Panels C and D show the same data grouped differently. Panel C compares the fixations to high-frequency targets and low-frequency competitors for a reduced versus unreduced prefix in the input. Panel D shows the analogous comparison for low frequency targets with high-frequency competitors.

shows that reduction lead to recognition costs for low-frequency targets. This hints at the hypothesized interaction between frequency and reduction, although not a cross-over interaction. To confirm this statistically, we tested the experimental effects with a number of mixed effects models. The simplest model had only Target Word Frequency and Target Word Reduction as contrast-coded fixed factors, with items and subjects as random factors. In this model, neither the effect of Frequency ($b = 0.3155$, $p_{\mathrm{MCMC}} = 0.38$), nor the effect of Reduction was significant ($b = 0.4238$, $p_{\mathrm{MCMC}} = 0.16$). Crucially, the interaction of Frequency and Reduction was significant ($b = -1.39$, $p_{\mathrm{MCMC}} < 0.05$).

We next investigated the role of three possible covariates: subjective frequency, cloze test rating, and cloze probability. The initial model contained the three covariates and their two- and three-way interactions with the experimental factors Target Word Frequency and Target Word Reduction. Insignificant effects were removed from the model starting with the highest-order interactions. The final model retained a significant main effect of cloze probability (b = -2.18, $p_{\mathrm{MCMC}} < 0.01$) and, as in the simple model, a significant interaction of Target Word Frequency and Target Word Reduction (b = -1.41, $p_{\mathrm{MCMC}} < 0.05$). (The effects of Target Word Frequency and Reduction were again not significant, $p_{\mathrm{MCMC}} > 0.1$, but remained in the model because of the significant interaction of these two factors.)

Given the effect of cloze probability on fixation proportions, we also tested a model in which there was an interaction of Target Word Reduction by cloze probability rather than an interaction of Target Word Reduction by lexical frequency. Given that the high-frequency words were also more predictable ($t(158) = 2.63$, $p < 0.01$, d = 0.4), the model did not include a term for lexical frequency. Even in this case, there was no significant interaction of cloze probability by Target Word Reduction ($p_{\mathrm{MCMC}} > 0.2$).

Discussion

The experiment showed that listeners can take advantage of the correlation between word frequency and phonological reduction. When hearing a reduced prefix, listeners looked more at the high-frequency words than at the low-frequency words. The results, however, did not show a cross-over interaction; that is, with unreduced prefixes there was no difference between high- and low-frequency words. One possible explanation is that the three possible advantages a word can have for word recognition—high frequency, a clear pronunciation, and (as we argue) the expected level of reduction for its frequency—are of roughly comparable strength, at least under the conditions of the present experiment. In three of our four conditions, target words had two of these three advantages, so their activation levels were approximately equal. Unreduced high-frequency words, for instance, have the advantage of their high-frequency and unreduced form, but do not have the expected level of reduction given their frequency. Only in the reduced low-frequency condition did targets lack all three advantages.

The current results from phonological reductions complement those of Salverda et al. (2003) for clear speech, who found listeners used the duration of the syllable [hæm] in clear speech to decide whether it was the entire word *ham* or part of the word *hamster*. Salverda et al. showed that listeners make lexical inferences from duration differences even in the absence of phonological reduction. In their experiment, the duration differences were driven by the realization of an optional prosodic boundary. Our results show that listeners make lexical inferences from phonological reduction even in the absence of prosodically-driven duration differences as they arise in clear speech.

One caveat of the current study is that the reduction affected the prefix of a morphologically complex word. While this allowed us to use very similar forms of reduction across the stimulus set, it raises the question whether the results would generalize to mono-morphemic words. The current results provide little basis to answer this question, and future research is necessary to answer this question. Related to this issue, a 0.96 correlation between word form frequency and lemma frequency prevents us from making claims about which frequency measure drives our result. Another issue related to morphology is that the duration measurements of our stimuli showed a selective reduction of the prefix, which raises the question whether our stimuli can be considered reasonable approximations of natural reductions. Here, it is important to note that reduction is not a process that affects all parts of speech to the same degree. Corpus studies have shown that, in stressed-time languages such as English, Dutch, and German, reduction affects especially unstressed syllables (Mitterer, 2008; Shockey, 2003, see Table 2.1 on p. 15; Van Bael, Baayen, & Strik, 2007). This makes the prefix more vulnerable than the stem. Moreover, all of our target words were the last content word in the sentence, and usually the very last word and hence were subject to the lengthening effects of focus accent and usually utterance-final lengthening as well, which would have counteracted the smaller shortening effects of intended reduction of the stem.

It is noteworthy that, in our data, it is lexical frequency and not contextual predictability that interacts with the level of reduction. Usually, both contribute to the amount of reduction found in corpus studies (Bell, et al., 2009; Pluymaekers, et al., 2005). In corpora, however, both effects tend to be correlated (see Bell et al., Table 2), making it difficult to distinguish their separate effects. A relevant study here is Scarborough (2010), who orthogonally varied a lexical variable (confusability) and contextual probability and found that vowel reduction is influenced strongly by lexical confusability and only to a small extent by the contextual appropriateness of a word in a given context. This would suggest that, in production, lexical variables might be stronger predictors of phonological reduction than context and discourse factors, and listeners have picked up this correlation and use it in word recognition. However, our design did not experimentally vary predictability, so the current data offers no conclusive evidence on how important predictability is to listeners.

This brings us to the question what type of mechanism allows listeners to take advantage of the correlation between lexical frequency and reduction. There are two main possibilities. First, episodic storage models for the mental lexicon (Connine, Ranbom, & Patterson, 2008; Goldinger, 1998) assume that listeners store phonological variants in the mental lexicon. Because the reduced variants of high-frequency words occur often, high-frequency words will still be strongly activated if the input comes in a reduced form.

It is also possible to account for the results without assuming storage of variant forms (i.e., the standard view in classical models of the mental lexicon). Upon hearing a reduced form, the activation of all high-frequency words may be given an extra boost. This requires lexical representations to be accessible by their frequency, not only, as in listening, by the auditory input, which is difficult to achieve in classical models of spoken word recognition, such as TRACE (McClelland & Elman, 1986) or Shortlist (Norris & McQueen, 2008).

For the more plausible account using multiple phonetic representations, there are a few alternatives. On one end of the continuum, there are models of the mental lexicon in which only acoustic patterns are stored (Goldinger, 1998). On the other end of the continuum, the lexical entry for a given word contains only a set of abstracted pronunciation variants (Connine, et al., 2008; McLennan, Luce, & Charles-Luce, 2003), much like the phonetic transcriptions [vərlift] and [vəlift] of the Dutch word *verliefd* (Engl., 'in love'). As argued elsewhere (Mitterer, 2006), assuming only acoustic patterns is problematic, because it makes it difficult to generalize. If only acoustic patterns are stored, listeners would, for instance, have to learn the likelihood of reduction separately for speakers of Dutch with an alveolar trill (saying [xrokt] for /xərokt/, Engl. 'smoked') and speakers with an uvular trill (saying [xʁokt] for /xərokt/). From a functional point of view, the assumption of storage for multiple abstracted phonetic forms for a given word hence seems to be the more promising account for the current data[2]. In this way, the listener can learn the likelihood of reduction disregarding orthogonal phonetic variation.

The issue of functionality also brings us back to our initial question. Does the ubiquitous nature of phonetic reduction challenge functional theories of phonology and their notion of a balancing act between the needs of speakers and listeners? The answer seems to be "no". Reductions do not occur randomly. Often they follow innate perceptual biases of the listeners. Sometimes they may even help the listener reject competitors and recognize the word intended by the speaker faster than would otherwise be possible.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439. doi: 10.1006/jmla.1997.2558

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America, 119*, 3048-3058. doi: 10.1121/1.2188331

Baayen, H. R., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412. doi: 10.1016/j.jml.2007.12.005

Baayen, H. R., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (release 2 ed.): Linguistic Data Consortium.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92-111. doi: 10.1016/j.jml.2008.06.003

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Boersma, P. (1998). *Functional Phonology. Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Cho, T., & McQueen, J. M. (2008). Not all sounds in assimilation environment are perceived equally: Evidence from Korean. *Journal of Phonetics, 36*, 239-249. doi: 10.1016/j.wocn.2007.06.001

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics, 70*, 403-411. doi: 10.3758/PP.70.3.403

Corley, M., Macgregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition, 105*, 658-668. doi: 10.1016/j.cognition.2006.10.010

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*, 507-534. doi: 10.1080/01690960143000074

Ernestus, M., Baayen, H. R., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*, 162-173. doi: 10.1006/brln.2001.2514

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes, 2*, 613-656. doi: 10.1080/016909697386646

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279. doi: 10.1037//0033-295X.105.2.251

Grammont, M. (1933). *Traité de phonétique*. Paris: Delagrave.

Hura, S. L., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech, 35*, 59-72.

Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. H. A. Marchal (Ed.), *Speech Production and Speech Modelling* (pp. 69-92). Dordrecht: Kluwer.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht, The Netherlands: Kluwer.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86. doi: 10.1016/0010-0285(86)90015-0

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 539-553. doi: 10.1037/0278-7393.29.4.539

Mitterer, H. (2006). Is vowel normalization independent of lexical processing? *Phonetica, 63*, 209-229. doi: 10.1159/000097306

Mitterer, H. (2008). How are words reduced in spontaneous speech?. In A. Botinis (Ed.), *Proceedings of the ISCA Tutorial and Research Workshop on Experimental Linguistics* . (pp. 165-168). Athens: University of Athens.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology, 59*, 1395-1424. doi: 10.1080/17470210500198726

Mitterer, H., Yoneyama, K., & Ernestus, M. (2008). How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction. *Journal of Memory and Language, 59*, 133-152. doi: 10.1016/j.jml.2008.02.004

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review, 115*, 357-395. doi: 10.1037/0033-295X.115.2.357

Pluymaekers, M., Ernestus, M., & Baayen, H. R. (2005). Leixcal frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America, 118*, 2561-2569. doi: 10.1121/1.2011150

Righi, G., Blumstein, S. E., Mertus, J., & Worden, M. S. (2010). Neural systems underlying lexical competition: An eye tracking and fMRI S\study. *Journal of Cognitive Neuroscience, 22*, 213-224. doi: 10.1162/jocn.2009.21200

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51-89. doi: 10.1016/S0010-0277(03)00139-2

Scarborough, R. A. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In C. Fougeron, B. Kühnert, M. D"Imperio & N. Vallée (Eds.), *Papers in Laboratory Phonology X* (pp. 557-586). Berlin: Mouton.

Shockey, L. (2003). *Sound patterns of spoken English*. Cambridge, MA: Blackwell.

Spinelli, E., & Gros-Balthazard, F. (2007). Phonotactic constraints help to overcome effects of schwa deletion in French. [10.1016/j.cognition.2006.07.002]. *Cognition, 104*, 397-406.

Steriade, D. (2001). Directional asymmetries in place assimilation: a perceptual account. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219-250). New York, NJ: Academic Press.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*, 322–338. doi: 10.1016/j.jml.2004.11.004

Tucker, B. V., & Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: Pirot.

Van Bael, C., Baayen, H. R., & Strik, H. (2007). Segment deletion in spontaneous speech: a corpus study using mixed effects models with crossed random effects *Proceedings of Interspeech* (pp. 2741-2744). Antwerp, Belgium: ISCA.

Zipf, G. K. (1949). *Human behaviour and the principle oleast-effort*. Cambridge MA: Addison-Wesley.

Footnotes

(1) One may wonder why repetition sometimes affects the results and sometimes does not. A crucial difference between the current procedure and that of Mitterer and McQueen (2009) is the use of a constant versus changing sentence frame. In Mitterer and McQueen's experiments, the sentence frame was similar over different trials (e.g., "click on the word cupboard above the square"), while here certain target-competitor pairs are associated with specific carrier sentences (see the Appendix). This may encourage listener to predict targets based on the carrier sentence.

(2) One remaining problem for this model is to account for the special status of canonical forms. Somewhat in contrast with such a model, it has been repeatedly been found that canonical forms, even if less frequent than a reduced form, are better recognized and processed than reduced forms (e.g., Pitt, Dilley, & Tat, 2011).