MAINTAINING CONTRAST IN REDUCED SPEECH IN STANDARD SOUTHERN BRITISH ENGLISH

Emanuela Buizza

University of York
emanuela.buizza@york.ac.uk

ABSTRACT

The aim of this paper is to investigate how contrast is maintained in highly reduced speech in Standard Southern British English. To this end, a carefully controlled elicitation task was designed to collect reduced forms of function words – which are known to be prone to high degrees of reduction – in anacrusis (initial unstressed position). Specifically, the small paradigmatic system of pronoun and contracted auxiliary sequences was recorded and analysed. This paper reports on the acoustic analysis of a few contrasting paradigms (*we'd* vs *we'll* and *you'd* vs *he'd*) focussing on the phonetic features that help maintain the distinction between them. The results indicate that the contrast is maintained by 'apparently missing' phonetic features being temporally re-distributed rather than segmentally realised.

Keywords: Reduction; function words; SSBE; contrast; English auxiliaries.

1. INTRODUCTION

1.1. Reduction

The broad term *reduction* refers to patterns of variation of speech sounds in connected speech in which sounds (and stretches of sounds) are associated with temporal reduction and articulatory undershoot [22, 2]. Compared to their full forms – e.g. the citation form of words found in dictionaries – reduced sounds and words are characterised by shortening, vowel centralisation and monophthongisation, more open consonant articulation, apparent deletion of phonetic features or segments and increase in coarticulatory features [20, 13].

The degree of reduction depends on several factors. The frequency with which a word or phrase recurs in speech is correlated with its degree of reduction [5]. Several studies have shown that sounds and syllables in more frequent words tend to be shorter [24] are more likely to undergo coarticulation [8] and to be deleted [18].

The lexical category of a word affects its degree of reduction too. Function words have a wider range of realisations [21, 23], are more frequent and predictable [3] and therefore can be more reduced than content words. Moreover, function words do not normally receive phrasal stress and can occur in unstressed positions [25]. The position of words in the prosodic structure is crucial in determining their realisation – sounds and words in unstressed position tend to have a shorter duration, lower amplitude, and decreased magnitude of gestures [7, 11].

Despite the pervasiveness of reduction in all speech styles [26], and cases of extreme reduction known to be more common than previously thought [16], reduced speech usually remains intelligible [12]. The aim of this research is to investigate how the contrast between minimal pairs that convey linguistic information is maintained in highly reduced speech. To this end and taking into account all the aspects that affect the degree of reduction listed above, sequences of pronoun and auxiliary were recorded and analysed. The hypothesis is that phonetic features of reduced sounds remain in the signal even in highly reduced speech.

1.2. English pronouns and auxiliaries

Pronoun and auxiliary combinations such as *I'm*, *she's*, *you'd*, *we'll*, *they've* exhibit a wide range of phonetic realisations. Pronouns and auxiliaries are function words, belong to a small paradigmatic system, occur frequently and can occur in unstressed positions. The contracted form of auxiliaries are clitics – grammatical elements that are not independent but need to be attached to another element [19]. According to Kaisse [19], the "host" and the clitic are realised as a phonological unit. They can also undergo grammaticalisation – a process whereby words that are frequently used in combinations become a "storage and processing units" [6]. According to Heine [15], grammaticalised combinations of words undergo "erosion (or 'phonetic reduction'), that is, loss in phonetic substance" [15]. For these reasons sequences of pronoun and auxiliary are the ideal candidate for an investigation on how contrast is maintained in reduced speech.

2. METHODOLOGY

2.1. Data collection

The data collection was carefully designed. Although it is known that spontaneous speech exhibits higher degrees of reduction than read speech, for the purposes of this study, read speech recorded in a laboratory setting was collected. There are several reasons behind this choice. First, all the English pronoun and auxiliary paradigms (henceforth pr+aux) had to be collected for the phonological analysis of the pr+aux system. Second, for the acoustic analysis and comparison of the reduced forms of pr+aux, all the paradigms had to occur in the same phonological context and prosodic structure. It was important to control for the rhythm and stress pattern of each utterance and for the neighbouring sounds. Third, the aim was to trigger high degrees reduction, so it was crucial to place the pr+aux paradigms in a weak position. The position chosen was that of anacrusis – any unstressed syllables before the first stressed syllable in an intonation group [10].

A list of declarative sentences was constructed. Each sentence contained a pr+aux paradigm in sentence-initial unstressed position, followed by the appropriate form of the verb *to burn*. The verb *to burn* was chosen because it has a bilabial consonant in onset and a mid-central vowel in the nucleus. The bilabial stop and the mid-central vowel allow the tongue to be in a neutral position restricting in this way the coarticulatory features due to the tongue position and movements. The main verb was then followed by an article and a noun, except when the verb form *-ing* was used; in this case, only a noun followed. To limit the influence of neighbouring sounds on the pr+aux, the sentences were preceded by silence and the main verb was followed by a restricted set of monosyllabic words. Sounds with long-domain resonances (e.g. syllable-coda /l/ [14]) were excluded from the sentences. All sentences had the same rhythmic structure of alternating weak and strong syllables: W-S-W-**S**. The phrasal stress was placed on the last syllable of the sentence.

Repetition is also known to trigger reduction [1]. For this reason, each paradigm was repeated five times in a row in the same sentence. The only difference between the five sentences was the last word which was different in each sentence so that being ‘new’ information, it would trigger the phrasal stress, while the rest of the sentence was ‘given’ information and would thus be reduced. Example of a sequence of sentences:
You’ve burnt the cake. You’ve burnt the fish. You’ve burnt the chips. You’ve burnt the pie. You’ve burnt the steak.

2.1.1. Speakers and procedures

A Power Point presentation with a sentence on each slide was created. Speakers saw one sentence at a time and read it aloud. The PPT presentation was timed to control for speech rate. On the slides, the last word of each sentence was in bold and highlighted in bright colours, so that the participants would focus on it and place the phrasal stress on it. Eleven female speakers of Standard Southern British English in their 20s were recorded in a quiet Recording Studio. At the beginning of the task, informants were instructed to speak as naturally as possible and to place the stress on the last word of the utterance. The stress pattern they were encouraged to use was played to them before they started the task. Every 40 sentences, the participants had a break of a few minutes in which the rhythmic pattern was played again as a reminder. The stress pattern was recorded by a male speaker using the sounds ta-ta-ta-ta.

2.1.2. Segmentation and measurements

Due to the nature of the material all instances of pr+aux sequences were manually segmented. Instead of delimiting segments, the onset and offset of phonetic events, such as ‘friction’, ‘periodicity’, ‘silence’, were delimited. The acoustic measurements were carried out using scripts in Praat [4].

The acoustic parameters measured were: duration, amplitude and formant dynamics of all sounds. In addition, the first four spectral moments (centre of gravity, standard deviation, skewness and kurtosis) of aperiodic sounds were also measured. For the formant dynamics, the first three formants were measured at 9 equidistant points in time from the first full cycle of periodicity to the last full cycle. The mean amplitude and spectral moments were measured in a temporal window of half the duration of the sound centred at mid-point (from ¼ to ¾ of the duration).

3. RESULTS

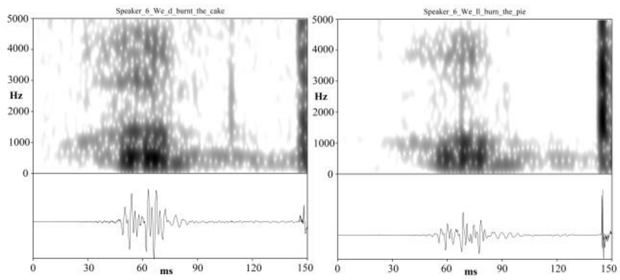
This paper reports the analysis of only a few paradigms focussing on the acoustic differences between them. The paradigms compared here are: *we’d* vs *we’ll*, and *you’d* vs *he’d*.

3.1. We’d and we’ll

In their phonological form, the paradigms *we’d* and *we’ll* differ in the consonant in coda position. However, in the dataset analysed, both paradigms are realised as a short vocoid – in *we’d* the plosion is lost, while in *we’ll* the laterality is lost. In most instances of *we’d*, the plosive is either not articulated or unreleased. If the gesture for the alveolar closure of the stop is in place, only the hold phase is articulated, as the release is masked by the lip closure for the bilabial stop that follows. This is not a feature of reduction, but a known connected speech process: in English, when there are two plosives in a row, the first one is unreleased [9] (all pr+aux are followed by /b/). In *we'll*, the lateral approximant is in most cases vocalised – the tip of the tongue does not make contact with the alveolar ridge. However, the secondary articulation – the movement of the tongue dorsum towards the velum typical of /l/ in coda position in English – is articulated even when the primary articulation is not. This results in the lateral approximant being realised as a back vowel. This process, called L-vocalisation, is known to occur in Southern varieties of English and is not a feature of reduction [17].

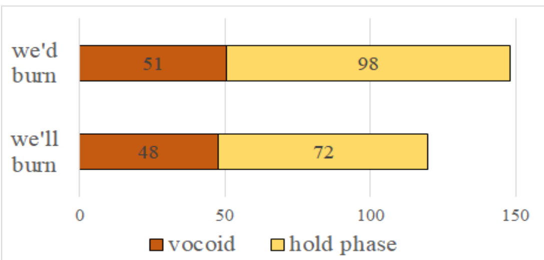
Figure 1 shows an instance of *we'd* (left) and one of *we'll* (right) produced by the same speaker. In both, the pr+aux sequence is realised as a monophthongal vocoid. Only in the voiceless portion at the beginning of *we'd* there is a small F2 movement. The spike in the hold phase of *we'<u>d</u> <u>b</u>urn* is the closure of the lips for the bilabial plosive in *burn*.

Figure 1: reduced instances of *we'd* (left) and *we'll* (right) produced by the same speaker.



These two paradigms differ on two main parameters: duration and formant dynamics. Figure 2 shows the mean duration across speakers and repetitions (N=110) of the vocoid and hold phase in *we'd* and *we'll*.

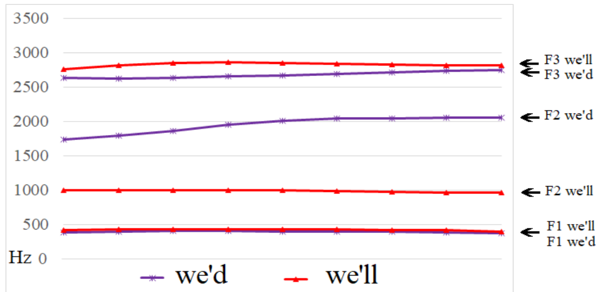
Figure 2: Mean duration of the vocoid (dark) and hold phase (light) of *we'd* and *we'll* in ms.



The mean duration of the hold phase in *we'd* is significantly longer than the mean duration of the hold phase in *we'll* ($\chi^2(1)=74.93$, $p<2.2e-16$). This is due to the instances of *we'd* in which /d/ is articulated but not released – the hold phases of /d/ and /b/ merge in a single long hold phase. Also the mean duration of the whole pr+aux *we'd* is significantly longer than the mean duration of the pr+aux *we'll* ($\chi^2(1)=39.74$, $p=2.901e-10$), while the mean duration of the vocoid alone is not ($\chi^2(1)=0.77$, $p=0.3803$).

The mean formant dynamics of a subset of instances of *we'll* (N=27, 53%) and *we'd* (N=25, 49%) are shown in Figure 3. The subset includes all instances in which the duration of the vocoid was shorter than the mean duration of the vocoid calculated across speakers and repetitions for each paradigm.

Figure 3: Time-normalised formant dynamics of the vocoids in *we'll* (red) and *we'd* (purple).

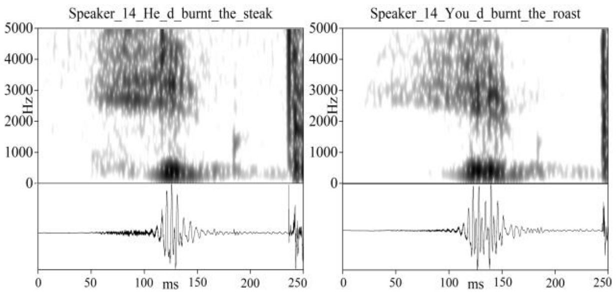


The main difference is in F2. A low F2 is typical of dark-l and back vowels – both of which are articulated with the tongue dorsum further back (or raised) in the oral cavity. A high F2 suggests a front vowel or clear resonance – both of which are articulated with the front of the tongue in a more advanced position in the oral cavity. Therefore, the formant dynamics of *we'll* and *we'd* indicate that *we'll* is characterised by a dark quality or position of the tongue further back than *we'd*, which is characterised by a clear quality or position of the tongue in a more advanced position. Note that the frequencies of all three formants in *we'll* and *we'd* differ from the start of the vocoids, indicating that the whole pr+aux is affected by the consonant in coda position.

3.2. He'd and you'd

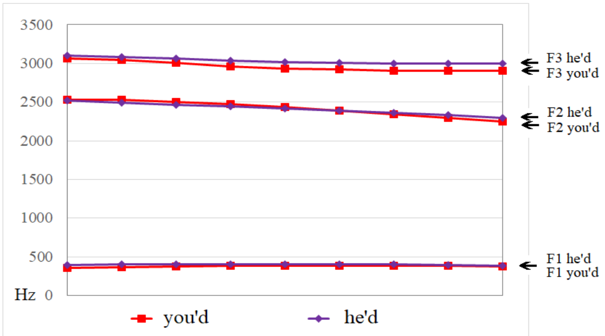
In several instances of *you*, the articulation starts before the vibration of the vocal folds starts. This feature can produce a portion of voiceless friction before voicing begins. Although in some cases the friction is too weak or short to be audible, it occurs in 70% (N=144/206) of instances of *you*. When it is audible, it has a palatal quality. This feature raises the question of whether the contrast with *he* is maintained, as the glottal friction in onset of *he* has a similar close front quality. Figure 4 shows an example of *he'd* (left) and *you'd* (right) produced by the same speaker.

Figure 4: instance of *he'd* (left) and *you'd* (right) produced by the same speaker.



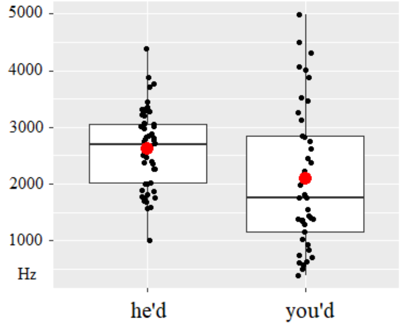
Surprisingly, the mean formant dynamics (calculated across speakers and repetitions) of the vocoid in *you'd* and *he'd* are rather similar (Figure 5).

Figure 5: Time-normalised formant dynamics of the vocoid in *he'd* (purple) and *you'd* (red).



The most noticeable difference between the formant dynamics in Figure 5 is the lower F3 in *you'd*, which might indicate that there is a residue of lip-rounding in the production of *you*. However, the main difference between the two paradigms is in the spectral qualities of the initial friction. The centre of gravity (Figure 6) of the friction in *you'd* is significantly lower than that in *he'd* ($\chi^2(1)$=13.07, p=0.0003).

Figure 6: Centre of gravity of the voiceless friction at the beginning of *he'd* and *you'd*.



The standard deviation and skewness are significantly different too, but kurtosis is not (Standard Deviation: $\chi^2(1)$=3.9, p=0.048; skewness: $\chi^2(1)$=6.48, p=0.011). This suggests that, despite the initial portion of friction possibly creating confusion between *you* and *he*, its spectral properties actually help maintain the contrast. When there is no friction at the beginning of *you*, the contrast is maintained by the presence of glottal friction in onset of *he*. The next step of this research is a perception experiment to test the perceptual roles of these spectral cues in maintaining the contrast between the two paradigms.

4. CONCLUSION

The analysis and comparisons reported in this paper suggest that even when minimal pairs of pr+aux are highly reduced the contrasting features between them are still present in the acoustic signal. In the case of the paradigms *we'd* and *we'll* the consonants in coda position are apparently missing. However, their acoustic correlates spread on the remaining sound material and are still available in the signal. As for the lateral, its primary articulation is lost (the apical gesture), but its secondary articulation (velarisation) affects the entire paradigm even when *we'll* is reduced to a very short vocoid. This contrasts with the 'clear' resonance in *we'd*. In the case of *you'd*, the temporal delay between articulation and vibration of the vocal folds that creates friction at the beginning of the pr+aux does not neutralise the contrast with *he'd*. This is because the spectral properties of the friction in *you* differ from those of the glottal friction in *he*. However, we do not know yet if these acoustic cues are available to perception. The next step of this research is a perception experiment to determine the role of the remaining acoustic features of reduced sounds in the correct interpretation of reduced speech.

5. ACKNOWLEDGEMENTS

This work was supported by the Economic and Social Research Council [ES/J500215/1].

6. REFERENCES

[1] Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., Newlands, A. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 42, 1–22.

[2] Barry, W., Andreeva, B. 2001. Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association* 31/1, 51–66.

[3] 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.

[4] Boersma, P., Weenink, D. 2016. Praat: doing phonetics by computer [Computer program]. Version 6.0.18, retrieved 24 May 2016 from http://www.praat.org/

[5] Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

[6] Bybee, J., Scheibman, J. 1999. The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37/4, 575–596.

[7] Cho, T. 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32, 141–176.

[8] Cooper, W. E., Paccia-Cooper, J. 1980. *Syntax and Speech*. Cambridge, Mass: Harvard University Press.

[9] Cruttenden, A. 2008. *Gimson's Pronunciation of English*. London: Hodder Education.

[10] Cruz-Ferreira, M. 1998. Intonation in European Portuguese. In Hirst, D., Di Cristo, A. (eds), *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press, 167–178.

[11] de Jong, K., Beckman, M. E., Edwards, J. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36, 197–212.

[12] Ernestus, M., Warner, N. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics* 39, 253–260.

[13] Gahl, S., Yao, Y., Johnson, K. 2012. Why reduce? Phonological neighbourhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66, 789–806.

[14] Heid, S., Hawkins, S. 2000. An acoustical study of long domain /r/ and /l/ coarticulation. *Proceedings of the fifth seminar on Speech Production: Models and Data*, *ISCA*, 77–80.

[15] Heine, B. 2003. Grammaticalization. In Joseph, B. D., Janda, R. D. (eds), *The handbook of historical linguistics*. Oxford: Blackwell, 575–601.

[16] Johnson, K. 2004. Massive reduction in conversational American English. In Yoneyama, K., Maekawa, K. (eds), *Proceedings of the first session of the 10th International Symposium on Spontaneous Speech: Data and analysis*, 29–54.

[17] Johnson, W., Britain, D. 2007. L-vocalisation as a natural phenomenon: explorations in sociophonology. *Language Sciences* 29, 294–315.

[18] Jurafsky, D., Bell, A., Gregory, M., Raymond, W. 2001. The effect of language model probability on pronunciation reduction. *Proceedings of ICASSP-01 II*, 801–804.

[19] Kaisse, E. M. 1985. *Connected speech: The interaction of syntax and phonology*. New York: Academic Press.

[20] Kohler, K. J. 1991. The phonetics/phonology issue in the study of articulatory reduction. *Phonetica* 48, 180–192.

[21] Lavoie, L. M. 2002. Some influences on the realisation of *for* and *four* in American English. *Journal of the International Phonetic Association* 32/2, 175–202.

[22] Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773–1781.

[23] Local, J. 2003. Variable domains and variable relevance: Interpreting phonetic exponents. *Journal of Phonetics* 31, 321–339.

[24] Pluymaekers, M., Ernestus, M., Baayen, R. H. 2005. Lexical frequency and acoustical reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118/4, 2561–2569.

[25] Selkirk, E. O. 1996. The prosodic structure of function words. In Demuth, K., Morgan, J. L. (eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, 187–213.

[26] Warner, N., Tucker, B. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America* 130, 1606–1617.