'Grandpa' or 'opera'? Production and perception of unstressed /a/ and /əʁ/ in German

Tamara Rathcke

Linguistics, University of Konstanz & English Language and Linguistics, University of Kent
tamara.rathcke@uni-konstanz.de

Christine Mooshammer

Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin
christine.mooshammer@hu-berlin.de

In the description of German phonology, two distinct phonetic symbols are currently recommended for the transcription of the vowels [a] (a central low vowel, phonemically /a/) and [ɐ] (phonemically /əʁ/) in word-final, unstressed positions. The present study examines whether differences between these two vowels exist in production and perception of Standard German speakers from the north of Germany. In Experiment 1, six speakers produced a series of minimal pairs that were embedded in meaningful sentences and varied with respect to their accentuation and position within a prosodic phrase. In Experiment 2, the minimal pairs produced by the six speakers of the first experiment were extracted from their respective contexts and tested with 44 native German listeners in a forced-choice identification task. Perceptual results showed a better-than-chance performance for one male speaker of the corpus only. Phonetic analyses also confirmed that only this male speaker produced subtle, but consistent F2/F3 differences between [a] and [ɐ] while the contrast was completely neutralised in the rest of the corpus. We discuss the role of prosody in vowel neutralisation with a specific focus on unstressed vowels and make suggestions for phonetic and phonological accounts of Standard German.

1 Introduction

Unstressed vowels are known to be prone to reduction of phonetic qualities (Lindblom 1963, Mooshammer & Geng 2008) and to be affected by neutralisation of phonemic contrasts (Crosswhite 2004, Padgett & Tabain 2005). While previous research has addressed neutralisation of vowel monophthongs in unstressed syllable positions, the nature of neutralisation processes in the context of a consonant vocalisation – such as coda-/r/ vocalisation (Dittrich & Reibisch 2006) – is still poorly understood. Recently, neutralisation has been documented in pre-rhotic contexts due to the type of /r/ that follows the vowel (Lawson, Scobbie & Stuart-Smith 2013). The neutralisation process affects three originally distinct checked vowels of Scottish English and is likely to have been triggered by coarticulation with the following bunched /r/ that exerts a stronger coarticulatory force over the preceding vowel than a front-tongue raised /r/. In contrast, speakers with the front-tongue raised /r/ often derhoticise, most likely due to delayed anterior tongue gestures (Lawson, Stuart-Smith & Scobbie 2018).

The phonetics of rhotics in the languages of the world varies substantially (Ladefoged & Maddieson 1996: Chapter 7) and can be socially stratified (Lawson et al. 2013, 2018). This is also true for German where the variation is primarily constrained by dialectal, stylistic and allophonic factors and includes taps and trills in different places of articulation [ɾ]/[r]/[ʀ], the uvular fricative [ʁ] and the uvular approximant [ʁ̞] (Kohler 1977/1995). Phonologically, German rhotics are treated as [+sonorant] because of their phonotactic behaviour (see Hall 1993 and Wiese 2003 for discussion). Since uvular [ʁ] and [ʁ̞] are the most widespread variants in Standard German (Kohler 1977/1995, Wiese 2003), an adequate representation of the Standard German rhotic is arguably /ʁ/.[1] Vocalised variants of /ʁ/ are widespread in post-vocalic coda positions of Standard German (e.g. Ulbrich 1973, Kohler 1977/1995, Barry 1995, Meinhold 1989, Simpson 1998). The process of /ʁ/-vocalisation gives rise to many lowering diphthongs in German (e.g. in *hier* [hiːɐ] 'here', *Herr* [hɛɐ] 'Mister', etc.) and also results in an additional central vowel which is known as the 'Lehrer'-schwa or the 'dark' schwa (Barry 1995, 1997). According to existing accounts of German, the unstressed word-final vowels in words like *Opa* 'grandpa' and *Oper* 'opera' are traditionally transcribed with two different IPA symbols, [a] and [ɐ] (Kohler 1977/1995, 1990). This representation implies a perceivable difference in quality between the two unstressed vowels. However, it is not clear if this contrast is indeed produced and perceived by native speakers of German. For example, a common mistake made by children acquiring Standard German orthography is the spelling of <a> for word-final <er>. This misspelling is very frequent (Schmidlin 2002) and considered as one of the typical, phonetically motivated mistakes (Fay 2010).

The aim of the present study is to investigate whether or not speakers of Standard (Northern) German – a variety that displays the post-vocalic /ʁ/-vocalisation (Kohler 1977/1995) – have a systematic acoustic difference when producing the unstressed word-final /a/ and [ɐ],[2] and if this vowel contrast can also be reliably identified by native adult listeners of Standard German. In the following, we will discuss the phonological background and the phonetic underpinnings of this vowel distinction. We further consider how a potential vowel merger might be influenced by sentence-level prosody or orthography, and outline the open questions that our study aims to answer.

1.1 Schwa sounds in Standard German

The status of [ɐ] in the phonology of German is considered marginal, despite the existence of surface minimal pairs like *bitte* [bɪtə] 'please' vs. *bitter* [bɪtɐ] 'bitter' and *Weite* [vaɪtə] 'width' vs. *weiter* [vaɪtɐ] 'further'. In phonological descriptions, the dark schwa [ɐ] is often discussed as the surface representation of the rhotic phoneme /ʁ/ or of the phoneme sequence /əʁ/, derived through a post-lexical vocalisation rule (see e.g. ʀ-coalescence in Hall 1992: 58). Consequently, [ɐ] and [ʁ] are not considered as two separate phonemes, but two allophones of one phoneme in complementary distribution: [ʁ] appears in the onset of a syllable, [ɐ] in all other positions. According to Meinhold (1989), the dark schwa can surface (i) as a postvocalic /ʁ/-allophone in a stressed syllable (e.g. in *hört* [hœɐt] '(s/he) hears'), (ii) as a realisation of the (unstressed) prefixes <er>, <ver>, <zer> (e.g. in *verbringen* 'to spend', with both [fɛɐˈbʁɪŋən] and [fɐˈbʁɪŋən] being possible surface variants), and (iii) as a realisation of the (unstressed) <er> suffix (e.g. in *weiter* [ˈvaɪt̬ɐ] 'further').

---

[1] The symbol /ʁ/ will represent the German rhotic phonologically throughout this paper (given that is it the most frequent variant in the standard variety of German, see Ulbrich 1973, Kohler 1977/1995), except in direct quotations from authors who use different variants. Because the rhotic patterns with the phonological class of sonorants, German phonologists frequently use /r/ or /ʀ/ to represent it (e.g. Hall 1992, 1993; Wiese 2000, 2001, 2003).

[2] Here and henceforth, we refer to the unstressed full vowel in terms of its underlying phonological quality – /a/ – and to the unstressed syllable-final /əʁ/ in terms of its most frequent surface variant in Standard German – [ɐ] (see Kohler 1975/1995, 1990; Meinhold 1989; Hall 1993).

As exemplified in (i) and (ii) above, /ʁ/-vocalisation can lead to a surface diphthong. Such diphthongal realisations seem to be optional in unstressed syllables and in coda positions of stressed syllables that contain an open vowel. A study by Dittrich & Reibisch (2006) compared formant trajectories of F1, F2 and F3 across several vowels followed by /ʁ/, and provided evidence against a postvocalic diphthongisation after /aː/ (e.g. in words like *Paar* [paː] 'a pair') and argued against the existence of the [aːɐ] diphthong in Standard German. Although most phonologists agree on the phonetic quality and use the symbol [ɐ] for (iii), Hall (1993) assumes [ʌ] instead of [ɐ] in his description of German phonology. The argument for this deviation is, however, purely phonological, related to the succinctness of the derivational rule /ʁ/ → [ʌ] that requires a change in only one feature ([consonantal]) instead of two ([consonantal] and [back]).

Phonologically, the two schwas of German – [ɐ] and [ə] – differ in their underlying representation which might lead to different articulatory specifications and the corresponding phonetic realisations. Kohler (1977/1995, 1990) whose account of German is adopted in the present paper, acknowledges the phonological status of /ə/ in contrast to the derived surface character of [ɐ] that is phonologically /əʁ/. Phonetic differences between the two schwas have also been addressed in previous research (e.g. Meinhold 1989, Barry 1995). For example, Meinhold (1989) documented top–down effects in the perception of [ə] vs. [ɐ] (e.g. in *Miete* [miːtə] 'rent' vs. *Mieter* [miːtɐ] 'tenant') and showed that both schwas could only be correctly identified if listeners heard some lexico-syntactic context surrounding the target word with the schwa. If the context was missing and a word had to be judged in isolation, the correct identification of the target vowel declined and was replaced by a bias towards the perception of [ə]. Following the framework of Articulatory Phonology (Browman & Goldstein 1992), Barry (1995) investigated the acoustic contrast between [ə] and [ɐ] in German, and focused primarily on the issue of the contextual influences on the realisation of the two schwas. According to Articulatory Phonology (e.g. Browman & Goldstein 1992), a schwa is a targetless vowel whose production does not involve an active articulatory gesture (only a relaxation movement toward a neutral tongue position) and is therefore highly prone to coarticulatory influences from the neighbouring segments. Given that there are two schwas in German, neither [ə] nor [ɐ] can be completely targetless, though they might still be more variable than the other vowels of the system. Similarly, Moisik (2006) argued that the presence of [ɐ] may lead to a slight displacement of the schwa in the vowel space, as compared to the schwa in English (see Figure 1).
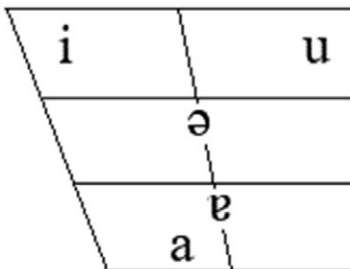

Figure 1 The relative location of [ɐ], /ə/ and /a/ in the German vowel space (after Moisik 2006: 26). Note that German /a/ is a central vowel, though traditionally represented by the front vowel IPA symbol (see Mooshammer & Geng 2008).

Barry's (1995) results showed more contextually induced variability in the production of [ə] than in the production of [ɐ], supporting the idea of a distinction between a functionally defined target of [ɐ] on the one hand, and a hypothetical relaxation target of [ə] on the other. A longer duration found for [ɐ] was interpreted in support of its phonological interpretation as

bi-segmental – i.e. consisting of /ə/ and /ʁ/ – in contrast to the phonologically uni-segmental /ə/. However, the durational difference observed in the data may also be interpreted as a reflex of the different degrees of jaw opening. The articulation of the more open vowel [ɐ] can be expected to take more time to be executed (see Solé & Ohala 2010 for a review on causes of intrinsic vowel duration). A comparison with /a/ is therefore required to shed light on the issue, given that [ɐ] and /a/ can contrast in unstressed word-final positions and produce minimal pairs like *Opa* 'grandpa' and *Oper* 'opera'.

The IPA symbols used for the (broad) transcription of the two German central open vowels in unstressed positions – /a/ and [ɐ] (Kohler 1975/1995, 1990) – imply a quality difference in vowel height. Note that a narrow transcription of the open stressed vowel is rather [a̲] since its quality is not as front as the traditional use of the symbol suggests (see Heike 1972).

Acoustically, this difference is expected to surface primarily in the frequency of the first formant, with /a/ showing a higher F1 than [ɐ]. However, given that both vowels occur in lexically unstressed positions, the phonetic difference between them can also be expected to be minimised due to an articulatory target undershoot, with /a/ becoming less peripheral (Lindblom 1990). The system of German full (i.e. stressable) vowels is known for more centralised vowel qualities and more closed tongue configurations in unstressed positions (Mooshammer & Geng 2008). Acoustically, centralisation affects particularly the first formant, with an unstressed /a/ measuring a lower F1 frequency as compared to a stressed /a/ (Mooshammer & Geng 2008). As a consequence, such target undershoot is likely to result in vowel neutralisation of the unstressed /a/ and [ɐ]. If this is the case, the two vowels might still differ with regard to their F2 frequency that has been previously found to distinguish between (stressed) /aː/ and /aːʁ/ (Dittrich & Reibisch 2006). The trajectory of F2 is also likely be involved if /əʁ/ has a diphthongal quality (Meinhold 1989, Dittrich & Reibisch 2006).

Apart from a difference in vowel quality, durational differences could also exist. Wiese (2000: 255) argues against an individual sound-level distinction between /a/ and [ɐ] because of their similarity. Since it is, however, possible to distinguish them, he suggests that /a/ and the non-syllabic [ɐ̯] (as in *Bier* [biːɐ̯] 'beer' are separated on the syllabic tier of their representation by association with V and C, respectively. He further assumes that the unstressed vowel /a/ should be 'long, or at least longer' than a vocalic [ɐ] (in e.g. *Oper* 'opera' because the former is underlyingly tense. Since he does not account for short tense vowels in unstressed position and length in phonological representations is conceptualised as binary, it is unclear how the difference could be represented.

1.2 Influences of the orthography

Previous research has suggested that orthography might play a role in preventing vowel mergers specifically (Faber & Di Paolo 1995) or sound neutralisations more generally (e.g. Warner et al. 2004, Ernestus & Baayen 2006). In case of the two vowels in question here, the spelling differs systematically: while the unstressed /a/ is regularly spelled as <a>, the word-final [ɐ] is consistently represented by the digraph <er>. This spelling difference might potentially lead to different pronunciations of word pairs like *Oper* 'opera' and *Opa* 'grandpa', primarily with respect to their duration. Among many reasons that can affect duration of phonological homophones like the English *time* vs. *thyme*, Gahl (2008) discusses the effects of orthography, which are two-fold. First, words with more letters could lead to a longer word duration. Indeed, this explanation has been corroborated by independent evidence provided by Nimz, Immel & Koop (2016) for German phonological homophones like *Wahl* 'choice' and *Wal* 'whale' which show durational patterns as predicted by Gahl (2008). With respect to our research question, the orthography account would predict that words containing [ɐ] (like *Oper* 'opera' should be longer than words with the unstressed /a/ (like *Opa* 'grandpa'. Second, the regularity of grapheme-to-phoneme correspondences might also influence the word duration (Gahl 2008). If the phonemic representation of a word is well reflected in its spelling, the word is likely to be shorter than a word with a more complex orthographic

representation of its phonemes (e.g. _fin_ vs. _thin_ in English or _Wale_ 'whales' vs. _Schale_ 'bowl' in German). Although German has a relatively unambiguous grapheme-to-phoneme mapping (e.g. Katz & Frost 1992), and the <er> spelling of all word-final [ɐ̯] is very regular, we can still predict a shorter duration for words containing an unstressed /a/ due to its one-to-one phoneme-to-grapheme correspondence.

Duration sometimes plays a disambiguating role in spectral mergers. For example, durational differences among merged vowels have repeatedly been reported for stressed vowels (e.g. Labov & Baranowski 2006, Fridland, Kendall & Farrington 2014, Wade 2017). It is unclear if such phonetic cue can also apply to unstressed vowel mergers, though it lends some preliminary support to the prediction of the orthographically driven account and reinforces the idea that duration might distinguish between /a/ and [ɐ̯] in case of a spectral merger.

1.3   Segment–prosody interaction

In terms of their prosodic characteristics, /a/ and [ɐ̯] are known to belong to two different vowel sets: while /a/ is a full vowel of German and can occur in stressed positions and carry a pitch accent, [ɐ̯] is a reduced vowel and is typically unstressed (Vennemann 1991). According to Vennemann's (1991) account of German phonology, unstressed /a/ and reduced [ɐ̯] are subject to different vowel prosodies which have unique phonological characteristics and can therefore be expected to show phonetic differences regardless of the presence of accentuation. For example, in English articulatory gestures are known to show larger amplitude and velocity differences between unstressed and stressed unaccented syllables than between stressed unaccented and accented syllables (Beckman & Edwards 1994). In general, stress and accentuation tend to lead to acoustically more peripheral vowel qualities and larger articulatory movements in German as well as other languages (de Jong, Beckman & Edwards 1993, Hoole & Mooshammer 2002, Mooshammer & Fuchs 2002, Mooshammer & Geng 2008). The amount of coarticulation found in accented words is also much reduced in comparison to unaccented words (de Jong et al. 1993, Harrington, Fletcher & Roberts 1995, Lindblom et al. 2007). We thus can expect the difference between the two vowels to become more salient under accentuation and to be minimised when words are unaccented.

Positions within a prosodic phrase might also condition the variation between the two vowels. Within the π (prosodic) gesture framework (Byrd & Saltzman 1998, 2003; Byrd et al. 2000, 2005), articulatory gestures at phrasal edges are timed further apart as compared to phrase-medial position – a timing property that results in a smaller gestural overlap and could therefore lead to more hyperarticulated and less neutralised phonemic contrasts. To this end, Piroth & Janker (2004) demonstrated that a complete neutralisation of a word-final voicing contrast in German obstruents was revoked in phrase-final positions resulting in subtle durational differences to preserve the underlying contrast, though the finding applied only in southern varieties of German. In contrast, phrase-medial positions tend to increase coarticulatory effects (see e.g. de Jong et al. 1993, Cho 2004). We can thus assume the distinction between unaccented /a/ and [ɐ̯] to diminish in phrase-medial positions and to potentially become manifest in phrase-final positions.

1.4   Perceptual evidence

A disparity of production and perception of a contrast has been previously observed in the context of (near-)mergers (Nycz 2013). Such disparity can go in both directions, i.e. a contrast being perceived but not produced (e.g. Hay, Warren & Drager 2006, Hay, Drager & Thomas 2013) or produced but not perceived (Labov, Yaeger & Steiner 1972). Once spectral cues to a vowel distinction are lost, duration has been frequently shown to play an important role for a perceptual disambiguation (e.g. Labov & Baranowski 2006, Wade 2017). Little is known about the use of duration to distinguish between unstressed /a/ and [ɐ̯]. We argue that it is a likely cue to the vowel contrast in question (see Section 1.2). However, previous work

(Dittrich & Reibisch 2006) did not show a significant effect of duration in the production of vowels with and without a postvocalic /ʁ/, like *Saar* (a German river) vs. *sah* (past tense of the verb 'to see'). There exists very limited perceptual evidence for a correct identification of [ɐ], as opposed to [ə]. Meinhold (1989) showed that the correct perceptual identification was only possible if the target word containing either schwa was presented with some surrounding lexico-syntactic context – that is, the schwa perception seems to rely on a top–down rather than bottom–up processing. In the absence of the relevant context, vowel perception displayed a clear bias towards [ə] (Meinhold 1989).

1.5 Aims and hypotheses

The present study aims to investigate whether or not the contrast between the two unstressed open vowels of German – /a/ and [ɐ] – is maintained in production and perception by speakers of Standard (Northern) German. Following previous research, we hypothesised that the contrast will be enhanced in prosodically strong positions (i.e. under accentuation and phrase-finally) and reduced elsewhere (see de Jong et al. 1993, Harrington et al. 1995, Hoole & Mooshammer 2002, Mooshammer & Fuchs 2002, Lindblom et al. 2007, Mooshammer & Geng 2008).

If the maintenance of this contrast is driven primarily by the underlying phonological representations or orthography, we expected to find durational (maybe in conjunction with spectral) cues that differentiate between the two vowels. To this end, two competing predictions can be put forward: (i) given that [ɐ] is consistently represented by two (in contrast to just one) letters or phonemes, i.e. <er> in orthography and /əʁ/ in phonology, we could predict [ɐ] to be produced as a longer vowel and thus perceived as a distinct vowel, versus (ii) under the phonological assumption of a general prosodic difference between the full and the reduced vowel set (following Vennemann 1991), longer durations could be expected in realisations of the full vowel /a/ that has the potential of carrying lexical stress within the German vowel system. Furthermore, since schwa vowels are also said to be targetless and therefore more variable (Browman & Goldstein 1992, Barry 1995), larger dispersions in the F1/F2 space are predicted for [ɐ] than for /a/.

In contrast, if the difference between the two vowels is primarily phonetic as many accounts of German suggest (Hall 1993; Kohler 1975/1995, 1990; Meinhold 1989), a difference in F1 would be the main acoustic feature to differentiate between the vowels (with a lower frequency for [ɐ]). Formant trajectories might also be expected to carry information about the two vowels, with more diphthongal spectral properties for [ɐ] due to a spelling pronunciation of <er> and the underlying bisegmental representation as /əʁ/ (Dittrich & Reibisch 2006).

Accordingly, we analysed vowel duration, formant frequencies and trajectories in productions of (pseudo)minimal pairs contrasting /a/ and [ɐ] in the word-final position (Experiment 1). We included F3 in the formant analyses, given that F3 is one of the main acoustic correlates of rhoticity in varieties of English (Espy-Wilson et al. 2000, Harrington 2010) though to date, there are no reports that F3 might be of relevance in the German vowel system. Minimal pairs recorded in Experiment 1 were subsequently extracted from the context and tested in a forced-choice identification study (Experiment 2).

2 Method

2.1 Experiment 1: Production

2.1.1 Speakers

We recorded six monolingual native speakers of Standard German from the North of Germany (three male). Four speakers of the sample came from Schleswig-Holstein (Trittau,

Kiel, Flensburg, Lübeck), one speaker each from Hamburg and Niedersachsen (Wolfsburg). The mean age of the speakers was 24 years (SD = 3) at the time of the recording, and all of them were students at the Christian-Albrechts-Universität zu Kiel. None of the speakers reported any speech or hearing impairment. All speakers were naive as to the purpose of this experiment. They were recorded in the sound-proof booth at the Institute of Phonetics and Speech Processing in Kiel.

2.1.2   Materials and procedure

The following four minimal pairs contrasting word-final [a] and [ɐ] were chosen for this study:

Dina – Diener	[diːna] female first name – [diːnɐ] 'servant'
Opa – Oper	[ʔoːpʰa] 'grandpa' – [ʔoːpʰɐ] 'opera'
Feta – Väter	[feːtʰa] 'feta (cheese)' – [feːtʰɐ] 'fathers'[3]
Clara – klarer	[klaːʁa] female first name – [klaːʁɐ] 'clearer'

Each word was embedded in meaningful utterances, designed to vary their prosody with respect to the target word's (i) phrasal position (phrase-final vs. phrase-medial) and (ii) presence of accentuation (accented vs. unaccented). For the ease of segmentation, words following the target always started with an initial /z/. Appendix A presents a full list of the test utterances in the four prosodic contexts. The utterances were printed on paper cards, repeated three times, and their serial order was randomised. This results in a total of 48 target tokens per speaker (4 minimal pairs × 2 accent conditions × 2 phrase positions × 3 repetitions). The participants were instructed to read the utterances at their comfortable speech rate and in their most natural tone of voice.

2.1.3   Segmental annotation

Praat (Boersma & Weenick 2016) was used to segment and annotate the target words. Five tiers were generated: Word, Accent, Position, Syllable and Segment. Using default settings of Praat spectrogram and waveform displays, the target words were visually inspected, and all segments of the target word as well as the initial segments of the following word were manually annotated. Vowel boundaries were determined through the presence of a visible second formant in the spectrogram. Audible strong glottalisation of the target vowel in phrase-final positions was annotated as part of the vowel. The following pause, if produced, was labelled between the offset of the final vowel and the visible frication noise of the following, frequently devoiced /z/.

Additionally, syllables of the target word were identified as STR (initial stressed syllables), FUL (unstressed syllables with the full vowel [a]) and RED (unstressed syllables with the reduced vowel [ɐ]).

2.1.4   Prosodic annotation

To check if the speakers produced the intended prosodic structure, all sentences were prosodically annotated following GToBI guidelines (Baumann, Grice & Benzmüller 2005). All analyses below are based on the actual productions of prosodic phrasing and accentuation (which sometimes deviated from the intended prosodic environment as prompted by the sentence context). The intended phrasal emphasis was missing in 11 out of 288 accented target

---

[3] These words constitute a quasi-minimal pair in varieties of German that contrast /eː/ and /ɛː/. These are distinguished in orthography by the graphemes <e> and <ä>, though the sound contrast is neutralised toward /eː/ in most varieties, particularly in Standard Northern German (Kohler 1995: 172f.).

words, and 73 out of the 288 unaccented target words were produced with a pitch accent. One female speaker (F2) in particular showed a strong tendency to produce words with a pitch accent regardless of the context. Approximately 24% of all unintended prosodic realisations were attributable to this speaker. The large majority of all prosodic boundaries were produced as intended in the given context. Only two of the intended phrase-medial target words were produced phrase-finally, with an intermediate phrase boundary and a high boundary tone. These two cases were excluded from the subsequent analyses. All of the intended phrase-final positions were signalled by a low boundary tone and a pause.

2.1.5   Measurements

For acoustic analyses, standard procedures of the *EmuR* library were used (Bombien et al. 2006, Winkelmann et al. 2017). Calculations of formant estimates were carried out by using the procedure *PraatToFormants2AsspDataObj* that transforms Praat estimates into *EmuR* formats. Settings of the largest formant value for the fifth formant were set to 5000 Hz for male speakers and to 5500 Hz for female speakers. Formant frequencies of the first three formants were extracted for the whole duration of the final vowel (see Figure 3 in Section 3.1.1 below). The midpoint of these formant trajectories was taken as the measure of vowel quality in these vowels (see Figure 2 in Section 3.1.1). The values were checked for outliers and manually corrected if necessary. Durations of the target vowels and words were also measured using standard procedures of *EmuR*. Syllable durations were used in order to calculate the log transformed ratios between the initial stressed vowel and the final unstressed vowel (see Sluijter & van Heuven 1996), using the equation: *log duration ratio = log (duration σ₁/duration σ₂)*.

2.1.6   Statistical analyses

Statistical analyses were carried out using R 3.3.0 (R Core Team 2018) with the packages *lme4* (Bates et al. 2015) and *lmerTest* (Kuznetsova, Brockhoff & Christensen 2017). We tested for the effect of *vowel quality* (VOW: [a] vs. [ɐ]), *accentuation* (ACC: accented vs. unaccented), *phrasal position* (PHR: medial vs. final) and SEX. For the model construction in the production part, we follow the recommendation by Harrison et al. (2018) that fixed effects should be selected in a hypothesis-driven way, and retained in the final model.

While prosodic factors of *accentuation* and *phrasal position* cannot be expected to interact with the speaker sex, *vowel quality* might potentially enter the interaction, given that females are known to lead sound changes 'from below' (Labov 2001). Therefore, the tested fixed effects model was specified as VOW, ACC, PHR and SEX as main effects and interactions of VOW with ACC, PHR and SEX (i.e. fixed effects structure in R as VOW*ACC + VOW*PHR + VOW*SEX). *Speaker* and *token* were fitted as random effects.

In order to avoid collinearity between factors and factor levels, the factors were coded and centered by subtracting the grand mean, following suggestions by Gelman & Hill (2006). Centering and coding affects the fixed effects in the following way: the reported intercept does not correspond to a base factor level but to the grand mean of the data. For the individual factors, the β-estimate corresponds to the difference between the two factor levels. As a preliminary, log-likelihood comparisons were used to assess whether the model fit improved by including random slopes by speaker. Relevant models are presented in Appendix B. For significant interactions, the data set was split accordingly. Statistical significances of the fixed effects are presented by the corresponding estimates of the regression coefficients and their associated standard errors, with probabilities of their quotient (a *t*-test) based on the Satterthwaite approximation for denominator degrees of freedom. Dependent variables included F1, F2, F3, duration (raw and normalised, see Section 3.1.2 for detail).

2.2 Experiment 2: Perception

2.2.1 Stimuli

To test the perceptual identification of the two vowels, we selected one instance out of the three target word repetitions from each speaker's production data. All of these tokens were spoken clearly, without errors, stutters or hesitation and could be easily extracted from the carrier sentence. Silences of 100 ms in duration were added before and after each word. A total of 384 experimental stimuli were tested, i.e. the four minimal pairs in two accentual and two phrasal positions ($8 \times 2 \times 2 = 32$) produced by the six speakers of Experiment 1 ($32 \times 6 = 192$) and repeated twice ($192 \times 2 = 384$). The experiment was set up using the speech analysis software *Praat* (Boersma & Weenink 2016).

2.2.2 Listeners

The perception experiment took place at the Humboldt Universität zu Berlin. Forty-four native listeners of Standard German (mean age: 29 years, SD = 5, 19 male) participated in the experiment. None of them reported any hearing impairment or considered themselves bilingual. They came from different regions of Germany, most of them were local (54% came from Berlin/Brandenburg).

2.2.3 Procedure

Laptops running the experimental *Praat* script and a pair of good-quality headphones were used for perceptual testing. Before and after the presentation of an auditory stimulus, participants saw a blank screen for 100 ms. Following a stimulus presentation, they saw a choice of two buttons labelled with the words of the respective minimal pair, and had to answer the question 'Which word did you hear?' by selecting either button. After registering a choice, the experiment proceeded to the next stimulus. The stimuli were presented in a randomised order twice. The location of the correct button on the screen was counterbalanced across the two repetitions of the same stimulus. An experimental session lasted some 20–30 minutes. Individual data (age, gender, place of birth and residence, dialect/standard statement, foreign languages, bilingualism and hearing ability check) were collected at the end of each testing session.

2.2.4 Analyses

To analyse perceptual patterns obtained from the listeners, the sensitivity index *d*-prime (d') was calculated. As applied to our identification data, d' reflects the listeners' ability to correctly identify /a/ when presented with the words *Clara*, *Opa*, *Dinah* and *Feta*, and to correctly reject /a/ when presented with the words *klarer*, *Oper*, *Diener* and *Väter*. The score itself results from *z*-transformed hit and false-alarm rates calculated following the formula: $d' = z(hit) - z(false\ alarm)$ (e.g. Macmillan & Creelman 1991). The higher the resulting score, the better the perceptual separation between two categories, with scores close to 0 being indicative of a performance at chance level. The d'-data were then tested for significance using ANOVA, with *stimulus speaker*, *accentuation* and *phrasal position* (and the interaction of the two prosodic factors) as predictors. In addition, a mixed logistic regression was performed on the raw perception data, with the same fixed factors as above. *Listener*, *item* and *order* of presentation were fitted as random effects. Note that *stimulus speaker* was included as a predictor to test for a potential effect of the substantive inter-speaker variation that we found in the production data among the male speakers of the corpus (see Section 3.1).

3 Results

3.1 Acoustic analyses

3.1.1 Vowel quality

The formant frequencies of F1, F2 and F3 were considered, to measure a potential quality difference between the two vowels. A Lobanov vowel normalisation was applied to the formant data, in order to pool across male and female speakers (Adank, Smits & van Hout 2004). Subsequently, the data were rescaled back to Hertz by multiplying each Lobanov-transformed value by the group standard deviation and then adding the group mean (see Winkelmann et al. 2017).

Figure 2 shows 95%-dispersion ellipses of the Lobanov-normalised F1 and F2 frequencies measured at vowel midpoint, comparing reduced [ɐ] (in green, shown as 6), unstressed [a] (in brown) and stressed long [aː] (in black). The latter was extracted from the first syllable in *Clara* and *klarer*. Except for the speaker M2, all speakers show a considerable overlap between an unstressed [a] and a reduced [ɐ] whereas their stressed [aː] is distinct due to a higher F1 frequency, in comparison to the two unstressed vowels. Since F1 is known to inversely correlate with the degree of jaw opening and tongue lowering (e.g. Lindblom & Sundberg 1971), this result suggests that both unstressed vowels were produced with a less open jaw and a higher tongue position than the stressed [aː].
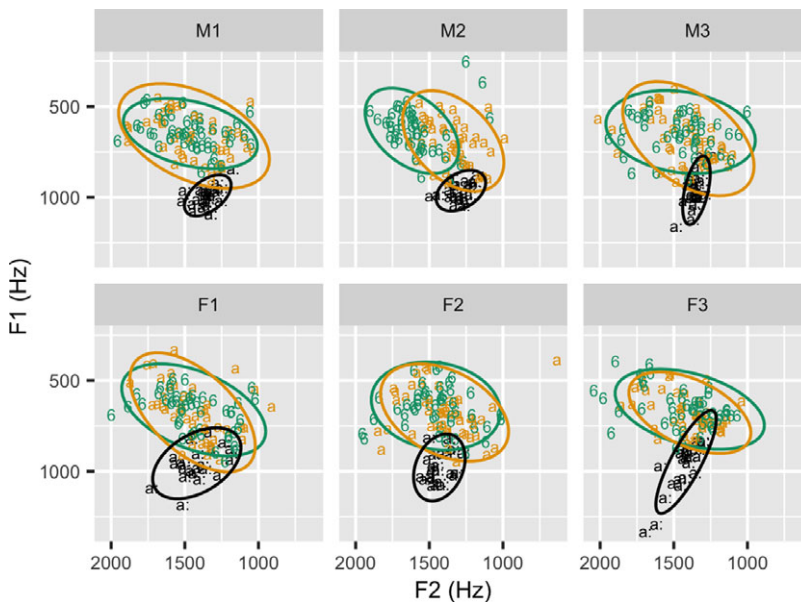

Figure 2 (Colour online) Dispersion ellipses of F1/F2 frequencies for [ɐ] (in green, annotated as 6, in line with the SAMPA notation of German, see Wells 1996) and unstressed /a/ (in brown). F1/F2 of the stressed /aː/ (in black) is given for comparison (taken from the speakers' productions of the target words *Clara* ([klaːʁa]) and *klarer* ([klaːʁɐ]). Male speaker data are plotted in the top panels, and female speaker data in the bottom panels.

Results from linear mixed-effects models are shown in Appendix B (see Table B1). They revealed no main effect of *vowel quality* ([a] vs. [ɐ]) for any of the three formants measured at vowel midpoint, and no interactions with either prosodic factor. In contrast, both accentuation ($t = 6.3$, $p < .001$) and phrasal position ($t = 4.1$, $p < .01$) showed an effect on

F1, with higher F1 frequencies measured under accentuation (an increase of 66 Hz) and in phrase-final positions (an increase of 98 Hz). Phrase-final positions led to a significant lowering of the higher formants, both F2 (by 187 Hz, $t = 9.3$, $p < .001$) and F3 (by 73 Hz, $t = 2.5$, $p < .05$). The frequencies of F3 further displayed a raising by 58 Hz under accentuation ($t = 2.3$, $p < .05$). For the frequencies of F2 and F3 there were also significant interactions between SEX and VOW that can be attributed to speaker M2 since the interaction lost significance once this speaker had been excluded from the dataset.

As hypothesised above, the dark schwa can be expected to be more variable in its target configuration than full vowels. Figure 2 indeed shows smaller dispersion ellipses for the stressed [aː] (they are, however, based on a smaller number of observations taken from two out of eight target words), though the dispersion of [ɐ] and unstressed [a] does not vary consistently.

Apart from the static differences in quality, we analysed formant trajectories of the two vowels, expecting a more diphthongal realisation for [ɐ]. For this, formant trajectories were extracted using *EmuR* library (Winkelmann et al. 2017), displayed and checked visually. Figure 3 shows averaged and time-normalised formant trajectories of F1, F2, and F3 split by speaker sex and accentuation. The graph indicates no differences between the averaged trajectories that would be suggestive of a more diphthongal realisation of [ɐ] (displayed in green) as compared to [a] (displayed in brown). To represent a diphthongal pronunciation of [ɐ] as [əɐ], the F1 trajectory would be expected to display a lower F1 at the onset of the vowel and a slight rise towards the offset, which is not the case here. Moreover, M2 productions of [a] and [ɐ] (shown on the right) differ exclusively with respect to F2 height but not the predicted F1 dynamics. That is, the speaker does not display a tendency towards a diphthongal realisation of [əɐ].
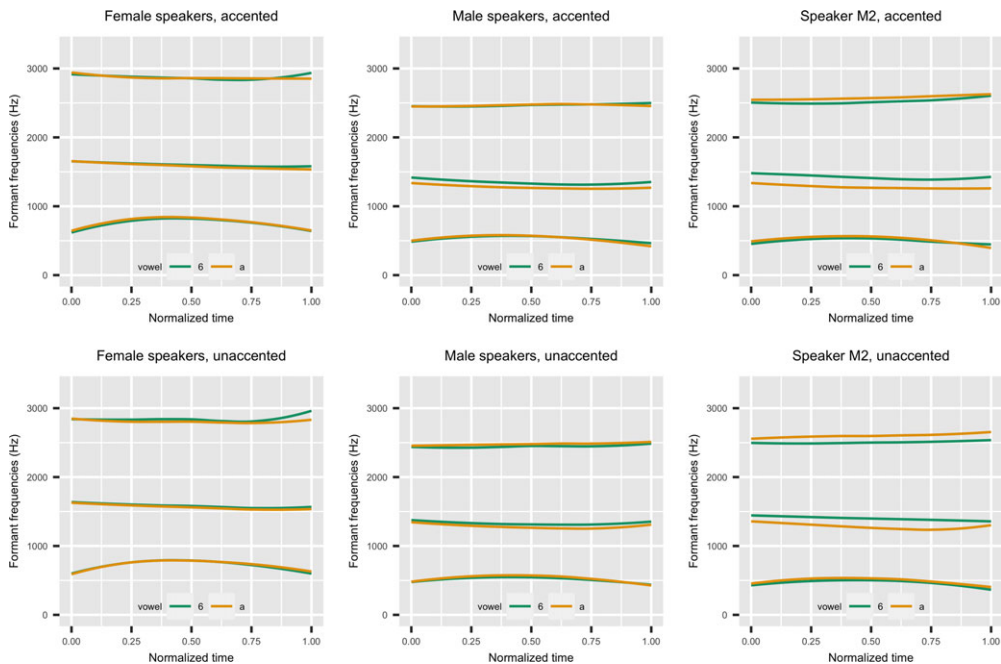

Figure 3 (Colour online) Time-normalised formant trajectories of unstressed /a/ (in brown) and [ɐ] (in green) averaged across accented and unaccented tokens produced by all female speakers (left panels), all male speakers (middle panels) versus speaker M2 only (right panels).

3.1.2 Durations of vowels, syllables and words

In terms of timing, both absolute and relative durations (calculated as ratios between the stressed and unstressed syllable duration in the same word) were considered. Figure 4 shows absolute durations of the two vowels in phrase-medial (on the left) and phrase-final (on the right) positions, in accented ([+acc]) and unaccented ([−acc]) words. Statistical results with inclusion of ACC slope for the random subject term are presented in Appendix B (see Table B2). All vowels were significantly longer in phrase-final as compared to phrase-medial positions ($t = 23.91$, $p < .001$). A visual inspection of the boxplots also suggested longer durations of [a] than [ɐ], but this numerical difference did not reach significance, either as a main effect or in interaction with accentuation.
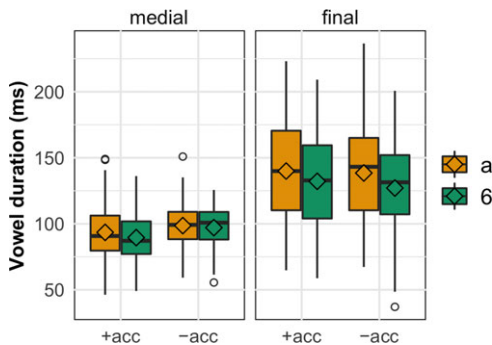

Figure 4 (Colour online) Absolute durations in milliseconds of unstressed /a/ (in brown) and [ɐ] (in green) measured in different prosodic contexts: phrase-medially (on the left) and phrase-finally (on the right), under accentuation ([+acc]) and in unaccented positions ([−acc]). Straight lines across the boxplots indicate the median, diamonds indicate the mean.

In order to normalise for speech rate differences that might affect absolute durations, a syllable timing metric was calculated as the logarithm of the durational ratio between the stressed and the unstressed syllable (see Sluijter & van Heuven 1996). This ratio indicates whether or not an unstressed [a] displays a larger effect of word stress (relative to the stressed syllable) than a reduced [ɐ], as predicted by Vennemann (1991) and Wiese (2000). If the first syllable is longer than the second, the ratios take negative values. The resulting log ratios are shown in Figure 5, and are broken down by phrasal position and by word-pair.

As can be seen in Figure 5, the ratios are highly dependent on the word pair since the segmental composition of the first syllable varied from a very long duration (as in syllables with a consonant cluster and a long vowel, e.g. *Clara/klarer*) to a relatively short duration (as in syllables with a singleton or empty onsets and a long vowel, e.g. *Opa/Oper)*. The difference between the two target vowels was thus tested for each word pair separately with speaker (but not token) as random effects. To avoid the type-1 error due to multiple pairwise comparisons, the alpha level was set to 0.0125 (deriving from 0.05/4). *Phrasal position* yielded significant main effects for all four word-pairs (see Appendix B, Table B3) whereas *accentuation* decreased the ratio only for the pair *Feta/Väter*. The word pair *Clara/klarer* showed a significant interaction between vowel quality and phrasal position which was lost once speaker M2 was excluded from the data set. For the word pair *Dina/Diener*, the interaction between vowel quality and accentuation was significant. After exclusion of M2, this interaction lost significance, though the model yielded a significant main effect of vowel quality, with larger ratios for *Dina* compared to *Diener*. One of the possible factors that might lead to a distinction between the two syllable durations in these words is the spelling difference. As was pointed out in the introduction, longer durations for the words ending in <er> could be expected
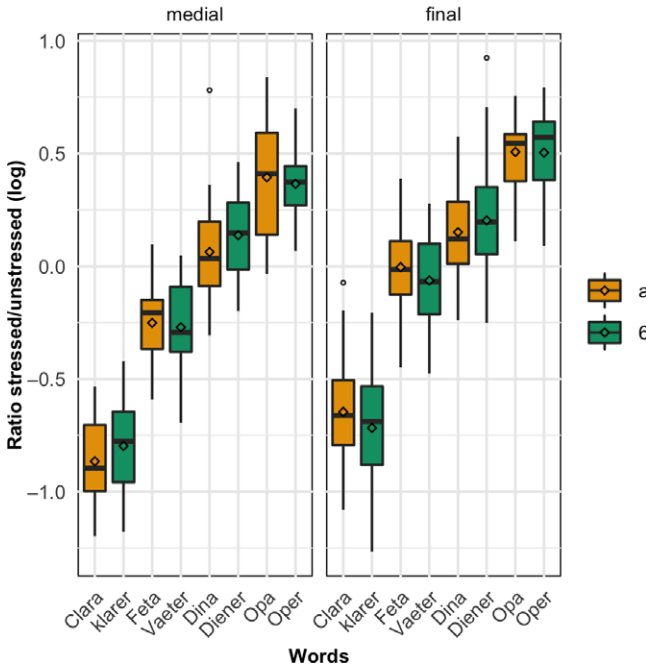
Figure 5 (Colour online) Normalised log-ratio durations of the stressed/unstressed syllables, displayed by target word pair and contrasting words with the full vowel /a/ (in brown) and words with the reduced vowel [ɐ] (annotated as 6, in green). Left: words in medial position; right: words in final position. 0 indicates that the stressed and the unstressed syllable are equal in duration; positive values indicate that the unstressed syllable containing /a/ or [ɐ] is longer. Straight lines across the boxplots indicate the median, diamonds indicate the mean.

because (i) there are more letters to represent one sound and (ii) the spelling is rather opaque with regards to its actual pronunciation.

So far, neither the raw vowel durations nor the ratios confirm these predictions for all word pairs. Since the differential spelling could also affect the duration of the whole word, Figure 6 presents these durations for medial and final positions. The effect of vowel identity was statistically tested for each word pair with the Bonferroni adjustment for multiple pairwise comparisons. All words lengthened significantly in phrasal position (see Appendix B, Table B4). Additionally, significant accentual lengthening was observed for word pairs *Feta/Väter* and *Opa/Oper*. The word pair that showed a significant effect of main interest here was *Dina/Diener*, with significantly shorter duration of the latter word. However, this finding is at odds with the hypothesis assuming a longer duration for a word with a higher number of graphemes. Given that the spelling of *Diener* has two letters more than the spelling of *Dina* (all other pairs differ in only one letter), the hypothesised effect is contrary to the present finding.

3.2 Perceptual identification

The sensitivity index *d′* approached 0 across the overall dataset, though the subsequent analysis of variance showed a significant effect of *stimulus speaker* ($F(1,5) = 2.8$, $p < .05$). Figure 7 shows the percentage of correct responses broken down by *speaker* (Figure 7A) in comparison to *accentuation* (Figure 7B) and *phrasal position* (Figure 7C). For most speakers, the number of correct responses was close to chance (i.e. 50%) while the perceptual identification of M2's tokens at 60% was slightly (but significantly) better than chance. A by-item *d′*-analysis did not show any significant effects for any of the word pairs (see Figure 8).
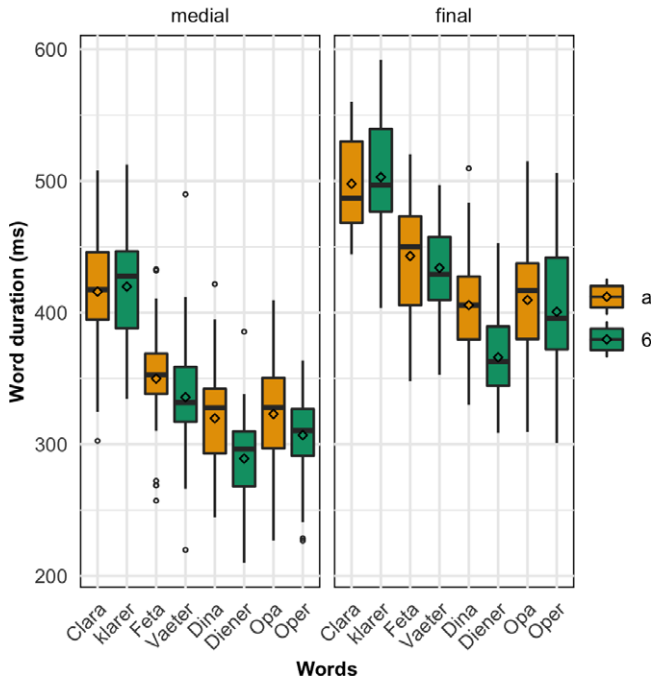
Figure 6 (Colour online) Word durations in milliseconds, displayed by target word pair and contrasting words with the full vowel /a/ (in brown) and words with the reduced vowel [ɐ] (annotated as 6, in green). Left: words in medial position; right: words in final position. Straight lines across the boxplots indicate the median, diamonds indicate the mean.

Logistic mixed-effects models fit by maximum likelihood revealed a significant effect of *phrasal position* (df = 1, LRT = 7.2, $p < .01$) and *stimulus speaker* (df = 5, LRT = 53.7, $p < .001$). The effect of phrasal position indicated that tokens extracted from phrase-final positions were more often correctly identified than tokens taken from phrase-medial positions (see Figure 7C). The speaker effect was driven exclusively by the male speaker M2 whose stimuli were recognised correctly more often than tokens produced by any other speakers (see Figure 7A). Once the data from speaker M2 were excluded from the set, none of the tested factors reached significance. For the remaining speakers, there was a trend toward significance for accentuation (df = 1, LRT = 2.8, $p = .0950$), indicating that accented tokens were slightly more often correctly identified than the unaccented ones.

4 Discussion and conclusion

4.1 Neutralisation of unstressed vowels

The results of both production and perception experiments presented here go hand in hand. While our acoustic analyses of the production data in Section 2.1 above show little difference between temporal measures, dynamic and static formant values of [a] versus [ɐ], results from the perception test with around 53.5% correct identification rates also suggest that the difference between [a] and [ɐ] is very subtle and therefore difficult to detect. Accentuation slightly enhanced the listener ability to distinguish between the two vowels, but the dataset did not provide the necessary statistical support for the associated hypothesis. Contrary to our hypothesis, identification scores were not consistently improved in phrase-final positions. Phrasal position had an effect on the perception of tokens produced by the outlier speaker
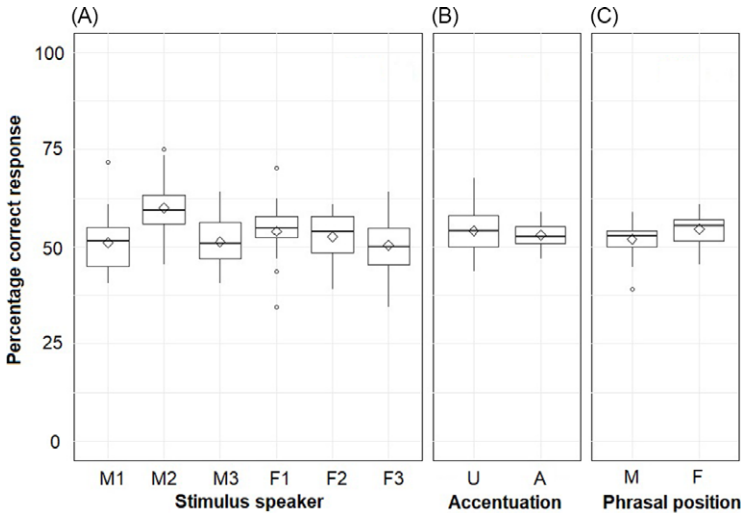
Figure 7 Proportion of correct answers for six stimulus speakers (A), two accentuation levels (B: unaccented, accented) and two phrasal positions (C: medial, final). The chance level is 50%. Straight lines across the boxplots indicate the median, diamonds indicate the mean.
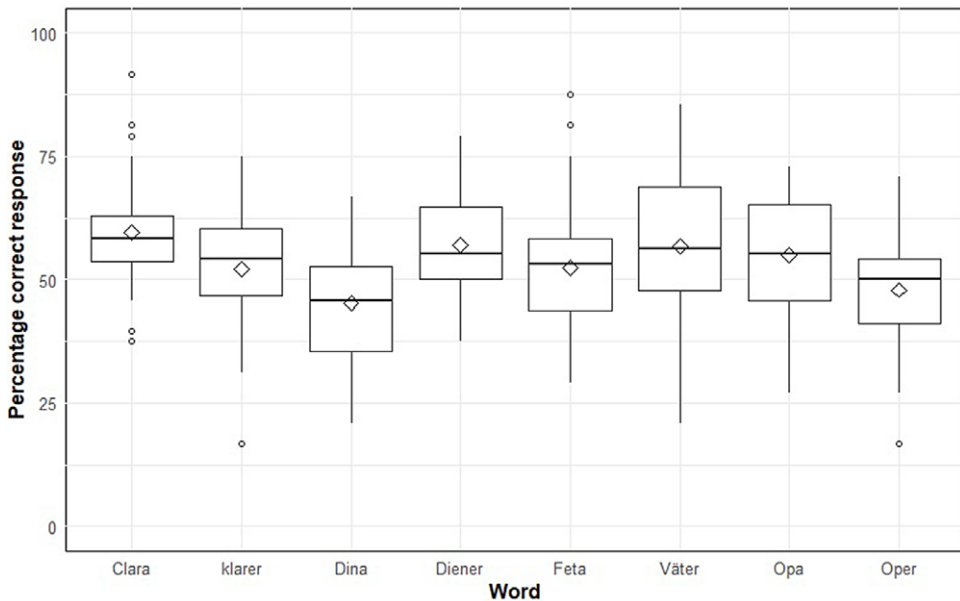
Figure 8 Proportion of correct answers for the eight target words of the study. The chance level is 50%. Straight lines across the boxplots indicate the median, diamonds indicate the mean.

though. M2's vowels [a] and [ɐ] could be significantly better distinguished in phrase-final than in phrase-medial positions, as predicted (de Jong et al. 1993, Piroth & Janker 2004, Cho 2004; see Section 1.3 above). Speaker M2 was the only speaker out of the six speakers of the present corpus who produced consistently different vowel qualities as measured by F2 and F3, although no difference in F1 could be measured even in this speaker's data. Given

existing impressionistic descriptions of German (e.g. Hall 1993; Kohler 1990, 1975/1995; Meinhold 1989), a lower F1 for unstressed /a/ was expected. Thus, the difference between unstressed /a/ and [ɐ] is subtle, if at all present, in these speakers' productions.

The finding that the vowels were not distinguished by duration is somewhat surprising, given how frequently this acoustic feature is exploited for a distinction in spectral vowel mergers (Labov & Baranowski 2006, Fridland et al. 2014, Wade 2017). Duration also plays a role in distinguishing between /a/ and vocalised /aɐ/ in stressed syllables that do not show any spectral differences due to /ɐ/-vocalisation (Ulbrich & Ulbrich 2007). Neither the phonological nor the orthographic accounts helped to adequately predict vowel duration patterns in these data. This finding suggests that in the context of a merger, the systems of stressed versus unstressed vowels obey different principles. While non-spectral cues like duration are often exploited in stressed vowel mergers, unstressed vowels with their shorter durations do not seem to benefit from trading cue relations in a comparable way. Indeed, neutralisation of unstressed vowel contrasts has often been attributed to their short duration given that articulatory targets are more likely to be undershot if the articulators have insufficient time to execute the required movements, with open vowels being particularly affected (e.g. Lindblom 1963, Flemming 2004). Vowel mergers in unstressed syllables are quite common in the languages of the world (for an overview see Crosswhite 2004), though short duration cannot be assumed to be the sole factor driving such effects (see Mooshammer & Geng 2008).

The results presented here generally showed a high level of agreement between the acoustic patterns found in the production data and their perception by a group of native German listeners, though there was also a discrepancy. More specifically, phrase-final tokens of the less-neutralising speaker M2 were better distinguished by the listeners than his phrase-medial realisations of the two vowels, though F2/F3 differed only slightly in his production of the two vowels across all prosodic contexts. While such discrepancies are not unprecedented in the context of ongoing mergers (Hay et al. 2006), our study pointed out how such discrepancies may be moderated by prosody. The effect might be, at least partly, attributable to longer durations of target words due to the finality. Given that the vowel quality differences of the speaker's productions were rather subtle across the board, and increase only slightly under accentuation, his phrase-final, longer vowels could be more salient in perception than his phrase-medial, shorter vowels. Moreover, phrase-final positions evoke less coarticulation than phrase-medial positions (see e.g. de Jong et al. 1993, Byrd & Saltzman 2003, Cho 2004) and can thus lead to less centralised vowel qualities.

Overall, these results do not support the interpretation of Vennemann's (1991) account of German phonology put forward in Section 1.5 above, according to which phonologically different 'vowel prosodies' of the unstressed, full vowel /a/ on the one hand and the derived, reduced [ɐ] on the other would lead to different surface realisations regardless of their accentuation. The same holds for Wiese's (2000: 255) assumption that an unstressed /a/ should be 'long, or at least longer' than [ɐ]. An expected durational contrast due to an underlying phonological difference in the vowels' affiliation with the syllabic tier cannot be corroborated by the data, though the results support Wiese's (2000) featural specification of unstressed /a/ and [ɐ] that does not distinguish between the two vowels.

4.2 Implications for description of Standard (Northern) German

The results of this study indicate that the difference between unstressed /a/ and [ɐ] tends to be neutralised in speakers of Standard (Northern) German. Clear acoustic-phonetic differences between the two vowels are absent in productions of minimal pairs, and listeners are not able to reliably identify tokens containing /a/ versus [ɐ]. Accordingly, word pairs such as *Oper* 'opera' and *Opa* 'grandpa' can be considered homophones rather than minimal pairs. Following the principles of the IPA (first published in 1888), there should be a one-to-one correspondence between a sound and its symbol, i.e. only detectably different sounds are to be represented by different symbols. This implies that at least in Standard (Northern) German,

one symbol should be used for both unstressed /a/ and [ɐ]. Based on the visual inspection of the formant charts and trajectories in Figures 2 and 3, the unstressed vowel [a] is produced with a lower F1 than its long, stressed counterpart [aː], indicating a more centralised and closed articulation in the unstressed vowel production. Given that the dispersion ellipses of the unstressed [a] and [ɐ] overlap almost completely, their location in the F1/F2 formant space suggests a slightly less open quality than that of a stressed [aː] in these data, the most likely candidate for this merged quality would be the symbol for the dark schwa [ɐ], leading to the transcription [ʔoːpʰɐ] for both, *Opa* 'grandpa' and *Oper* 'opera'. The advantage of this transcription seems not only empirically valid but also quite obvious in the context of foreign language teaching where a phonological approach to the vowel representation might be particularly misleading.

There is, however, at least one argument against this recommendation, namely the fact that a vocalic merger of this kind stems from a general articulatory process of vowel reduction due to target undershoot that affects all German vowels in a similar way (see Mooshammer & Geng 2008). Moreover, vowels in many languages undergo such target undershoot in unstressed or unaccented positions (see e.g. Beckman, de Jong & Lee 1992, Crosswhite 2004). The undershoot in German does not depend on shorter vowel duration in the absence of stress or accent *per se*, but is caused by an increase in coarticulation with the neighbouring consonants (Mooshammer & Geng 2008). Traditionally, this kind of vowel reduction is not marked in broad transcriptions of German since it is a direct consequence of prosodically weak positions and can thus be derived by a rule. Therefore, it seems debatable whether unstressed /a/ should constitute an exception.

Moreover, the present findings are necessarily restricted to the Northern variety of Standard German and are unlikely to generalise to other dialects. The phonetic dispersion of possible cross-dialectal realisations of the dark schwa is shown in Figure 9 (after Kohler 1977/1995), and can include [ə], [ɐ], [ɑ], [ɜ] (Russ 1990). According to Ulbrich & Ulbrich (2007), /ʁ/-vocalisation is a relatively recent phenomenon in German. In the 1960s and the 1970s, postvocalic coda-/ʁ/ was still mostly produced as a trill or a fricative (at least after short vowels), while 1990s saw the rise of 'rhotacised vowels' in all postvocalic contexts (Ulbrich 2002). These predecessors of modern vocalised rhotics were first observed in the context of a preceding long vowel in the early 20th century (Viëtor 1901). Work by Ulbrich & Ulbrich (2007) demonstrates that the process of /ʁ/-vocalisation applies across all dialectal areas of German-speaking Europe and has been spreading from the north to the south. Accordingly, we predict that vowel neutralisations due to /ʁ/-vocalisation might be particularly advanced in the variety studied here while other, especially southern, varieties are likely to show some phonetic differences between unstressed /a/ and /əʁ/ in the same linguistic materials as used here.
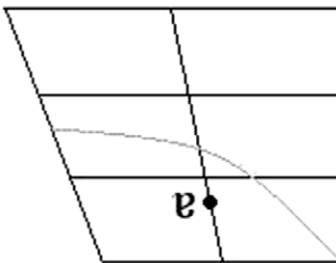

Figure 9 Cross-dialectal dispersion of [ɐ] in the vowel quadrilateral (after Kohler 1977/1995: 165).

4.3   Conclusions and outlook

The present study found little evidence for a distinction between unstressed /a/ and [ɐ] in a small corpus of Standard Northern German, and we conclude that the vowel contrast is neutralised in this variety (see Ulbrich & Ulbrich 2007). Our corpus included laboratory speech recordings rather than an unscripted speech database. If unstressed /a/ and [ɐ] were distinct vowel qualities, we would expect to find such different realisations primarily in the present type of data, but this was generally not the case. It is possible that we would find more hyperarticulated realisations of unstressed /a/ and [ɐ] under a corrective focus, e.g. in a "mishearing" scenario. Acoustic cues to a phonetic contrast in such explicit hyperarticulated contexts would benefit from further detailed analyses with respect to their spectral and durational cues. Additionally, the F0/F1 distance (see Traunmüller 1981) might be a promising phonetic cue to explore for these two vowels in different prosodic contexts. Moreover, the forced-choice identification task employed in the present study might have lacked the sensitivity that is required to tap into the perceptual processing of vowel neutralisations. More sophisticated online methods (such as e.g. eye tracking) might be more appropriate for observing the disambiguating effects of fine-phonetic detail in the acoustic signal of neutralised vowel contrasts.

The present study was based on a relatively small sample of six speakers and a geographically heterogenous group of listeners, and would benefit from a follow-up with larger, more homogenous groups of participants. Specifically, 44 listeners of our perception study came from diverse regional backgrounds, but were asked to respond to samples taken from the Standard Northern variety. We assumed that regional differences across German Standards would be negligible in read speech, though an effect of accent familiarity cannot be completely ruled out. However, such effects are known to affect online processing only temporarily (Floccia et al. 2006), and were accounted for in the statistical model of the perception data in the present study.

Acknowledgements

Appendix A. Experimental materials (target words only)

Phrase positions *Target words*	Accented positions	Unaccented positions
Phrase-medial *Väter*	Im Kindergarten ist Elternabend und die Erzieherin hat ausdrücklich gefordert: Die **Väter** sollen auch kommen. Sie will über die Vater-Kind-Beziehung reden.	Ich habe den besten Vater der Welt. Die **anderen** Väter sind weder so großzügig noch so liebenswürdig wie meiner. Da bin ich mir sicher.
Phrase-final *Väter*	Die Politiker haben es beschlossen: Mehr Rechte für **Väter**. Sie sollen nicht nur zahlen, sondern ihre Kinder auch öfter sehen dürfen.	Meine Eltern haben sich früh getrennt und meine Mutter hat wieder geheiratet. Man kann sagen, ich habe **zwei** Väter. Sicher ist das merkwürdig, aber ich finde es toll.

Phrase positions _Target words_	Accented positions	Unaccented positions
Phrase-medial _Feta_	Beim Einkaufen ist es immer das Selbe. Limburger habe ich schnell gefunden, aber **Feta** suche ich ewig. Jedes Mal wird der umgeräumt.	Bei Feta bin ich sehr wählerisch. Ich mag nicht rohen, sondern nur **eingelegten** Feta sehr gerne. Den esse ich dann am liebsten mit Oliven.
Phrase-final _Feta_	Es gibt viele Käsesorten, aber am besten finde ich **Feta**. Super schmeckt er überbacken und mit Preiselbeeren serviert.	Deshalb nehme ich immer **extra viel** Feta. Sonst schmeckt der Salat viel zu fade.
Phrase-medial _Diener_	Alle Kinder träumen davon, in einem Schloss leben zu können. Ein **Diener** servierte ihnen das Essen und räumte ihre Zimmer auf. Eine schöne Vorstellung.	Bei der Schlossführung haben wir einiges über das Dienstpersonal gelernt. Ein **guter** Diener sagt nichts. Er macht still seine Arbeit und wird am besten nicht bemerkt.
Phrase-final _Diener_	Von meiner Mutter hab ich oft gehört: „Ich bin nicht dein **Diener**! Sieh zu, dass du deinen Müll alleine wegräumst."	In einem Palast wird überall Dienstpersonal gebraucht. Es gibt nicht nur ein paar, sondern **hunderte** Diener. Sie werden in der Küche gebraucht, zum Ankleiden oder zum sauber machen.
Phrase-medial _Dina_	Ich gehe heute zum Schulkonzert. Meine Freundin **Dina** singt im Chor und hat ein Solo. Das werde ich mir nicht entgehen lassen.	Unsere Katze wurde seit Tagen nicht mehr gesehen. Deshalb müssen wir nicht erst morgen, sondern **heute** nach Dina suchen. Nicht das ihr was passiert ist.
Phrase-final _Dina_	Ich habe viele Freunde zu meinem Geburtstag eingeladen. Besonders freue ich mich auf **Dina**. Sie ist gerade erst aus Amerika zurückgekommen.	Muss man dir denn alles zweimal sagen? Ich meine nicht die dünne, sondern die **dicke** Dina. Sie ist viel netter als die andere.
Phrase-medial _klarer_	Meine Musiklehrerin verlangte immer von mir, ich sollte **klarer** singen. Ich weiß bis heute nicht, was sie damit gemeint hat.	Nach langem Probieren habe ich nun eine Brille mit der ich **endlich** klarer sehen kann. Sie hat genau die richtige Stärke.
Phrase-final _klarer_	Das hättest du auch gleich sagen können. Jetzt sehe ich die Dinge **klarer**. Samstags bist du also nicht zu Hause.	Wasserenthärter in der Spülmaschine macht die Gläser nur **bedingt** klarer. Schließlich sollte man nicht immer auf die Werbung hören.
Phrase-medial _Clara_	Ich weiß, es ist kaum zu glauben, aber nicht ihr Bruder, sondern **Clara** sammelt Modellautos. Hannes geht lieber ins Ballett.	Da hast du etwas falsch verstanden. Nicht in München, sondern in **Kiel** sucht Clara seit Langem eine Wohnung Ansonsten hätten wir ihr helfen können.
Phrase-final _Clara_	Darf ich dir meine Tochter vorstellen? Das ist **Clara**. Sie hat gerade ihr Abitur gemacht und will nun Medizin studieren.	Clara ist eine schlechte Sängerin. Im Chor kennen sie alle nur als **schreiende** Clara. Sie ist wirklich nicht zu beneiden.

Appendix B. Model summaries of the production data analyses

The factors included in the analyses and presented in Tables B1–B4 are as follows: VOW = vowel quality, ACCE = accentuation, PHR = phrase position, SEX = speaker sex, VOW*ACCE = interaction of vowel quality and accentuation, VOW*PHR = interaction of vowel quality and phrase position, VOW*SEX = interaction of vowel quality and speaker sex (see also Section 2.1.6 in the text).

Table B1 Results of linear mixed effects models for vowel quality (F1, F2 and F3).

F1	Beta	Std. Error	df	t	p
Intercept	686.679	20.307	7.974	33.815	< .001
VOW	27.222	40.612	7.974	0.670	n.s.
ACCE	66.021	10.423	552.599	6.334	< .001
PHR	98.251	24.286	7.964	4.046	< .01
SEX	9.021	9.931	548.211	0.908	n.s.
VOW*ACCE	15.918	20.854	552.652	0.763	n.s.
VOW*PHR	24.467	48.569	7.964	0.504	n.s.
VOW*SEX	−21.424	19.864	548.213	−1.079	n.s.
F2					
Intercept	1439.310	47.070	7.989	30.578	< .001
VOW	−61.714	94.135	7.989	−0.656	n.s.
ACCE	26.238	19.836	8.245	1.323	n.s.
PHR	−187.141	20.022	8.026	−9.347	< .001
SEX	0.907	9.103	546.473	0.100	n.s.
VOW*ACCE	−13.681	39.677	8.250	−0.345	n.s.
VOW*PHR	53.688	40.042	8.027	1.341	n.s.
VOW*SEX	76.720	18.208	546.49	4.214	< .001
F3					
Intercept	2659.827	26.792	7.912	99.279	< .001
VOW	8.791	53.581	7.913	0.164	n.s.
ACCE	58.061	25.234	14.986	2.301	< .05
PHR	−73.058	29.641	8.968	−2.465	< .05
SEX	6.469	20.812	548.956	0.311	n.s.
VOW*ACCE	−33.330	50.482	15.001	−0.660	n.s.
VOW*PHR	−0.793	59.280	8.969	−0.013	n.s.
VOW*SEX	−90.636	41.628	548.960	−2.177	< .05

Table B2 Results of linear mixed effects models for raw vowel duration (ms).

	Beta	Std. Error	df	t	p
Intercept	113.829	8.304	9.748	13.707	< .001
VOW	5.674	8.683	7.628	0.653	n.s.
ACC	5.969	2.936	5.431	2.033	n.s.
PHR	41.503	1.736	550.540	23.909	< .001
SEX	7.493	10.414	5.926	0.719	n.s.
VOW*ACC	0.570	3.659	546.141	0.156	n.s.
VOW*PHR	5.189	3.466	549.739	1.497	n.s.
VOW*SEX	−6.530	3.483	550.441	−1.875	n.s.

Table B3 Results of linear mixed effects models for vowel duration ratios comparing [aː] and [ɐ] in each word pair separately.

Clara – klarer (all speakers)	Beta	Std. Error	df	t	p
Intercept	−0.759	0.060	6.098	−12.571	< .001
VOW	−0.009	0.028	134.037	−0.306	n.s.
ACC	−0.014	0.034	134.304	−0.409	n.s.
PHR	0.141	0.027	134.030	5.258	< .001
SEX	0.085	0.121	6.108	0.703	n.s.

Table B3 Continued

Clara – klarer (all speakers)	Beta	Std. Error	df	t	p
VOW*ACC	0.077	0.068	134.172	1.135	n.s.
VOW*PHR	0.146	0.054	134.019	2.736	< .01
VOW*SEX	−0.007	0.057	134.057	−0.115	n.s.
Clara – klarer (without M2)					
Intercept	−0.719	0.062	5.067	−11.508	< .001
VOW	−0.029	0.030	112.022	−0.966	n.s.
ACC	−0.036	0.034	112.294	−1.042	n.s.
PHR	0.159	0.028	112.019	5.658	< .001
SEX	0.001	0.125	5.080	0.009	n.s.
VOW*ACC	0.117	0.068	112.130	1.717	n.s.
VOW*PHR	0.105	0.056	112.016	1.860	n.s.
VOW*SEX	0.041	0.061	112.032	0.681	n.s.
Feta – Väter	Beta	Std. Error	df	t	p
Intercept	−0.148	0.047	5.992	−3.125	n.s.
VOW	0.041	0.024	134.991	1.755	n.s.
ACC	−0.074	0.024	135.280	−3.040	< .01
PHR	0.226	0.024	135.018	9.589	< .001
SEX	0.013	0.095	5.998	0.141	n.s.
VOW*ACC	−0.022	0.049	135.765	−0.440	n.s.
VOW*PHR	0.053	0.047	134.991	1.115	n.s.
VOW*SEX	−0.047	0.047	135.001	−0.992	n.s.
Dina – Diener (all speakers)					
Intercept	0.136	0.060	6.007	2.261	n.s.
VOW	−0.062	0.026	135.009	−2.415	n.s.
ACC	0.054	0.027	135.223	2.037	n.s.
PHR	0.094	0.026	135.021	3.608	< .001
SEX	0.033	0.121	6.007	0.274	n.s.
VOW*ACC	−0.160	0.054	135.358	−2.986	< .01
VOW*PHR	−0.002	0.052	135.009	−0.037	n.s.
VOW*SEX	0.064	0.052	135.012	1.235	n.s.
Dina – Diener (without M2)					
Intercept	0.185	0.056	4.997	3.332	
VOW	−0.080	0.030	111.993	−2.687	< .01
ACC	0.042	0.030	112.252	1.408	n.s.
PHR	0.081	0.029	112.002	2.760	< .01
SEX	−0.065	0.111	4.993	−0.588	n.s.
VOW*ACC	−0.134	0.060	112.485	−2.230	n.s.
VOW*PHR	0.001	0.059	111.987	0.022	n.s.
VOW∗SEX	0.102	0.060	111.988	1.706	n.s.
Opa – Oper					
Intercept	0.440	0.029	6.101	15.236	< .001
VOW	0.018	0.029	136.126	0.609	n.s.
ACC	−0.049	0.029	137.41	−1.687	n.s.
PHR	0.121	0.029	136.114	4.211	< .001
SEX	0.054	0.058	6.044	0.936	n.s.
VOW*ACC	−0.032	0.058	136.778	−0.559	n.s.
VOW*PHR	−0.036	0.057	136.047	−0.631	n.s.
VOW*SEX	−0.083	0.057	136.111	−1.451	n.s.

Table B4 Results of linear mixed effects models for raw word duration comparing word pairs separately.

Clara – klarer	Beta	Std. Error	df	*t*	*p*
Intercept	456.51	7.826	6.350	58.331	< .001
VOW	−9.264	7.148	134.111	−1.296	n.s.
ACC	16.228	8.613	135.110	1.884	n.s.
PHR	82.354	6.762	134.082	12.180	< .001
SEX	7.032	15.675	6.385	0.449	n.s.
VOW*ACC	22.659	17.194	134.620	1.318	n.s.
VOW*PHR	−3.916	13.518	134.043	−0.290	n.s.
VOW*SEX	−3.399	14.371	134.188	−0.236	n.s.
Feta – Väter					
Intercept	391.666	8.714	5.960	44.948	< .001
VOW	11.540	5.783	134.958	1.996	n.s.
ACC	29.508	5.964	135.471	4.947	< .001
PHR	97.940	5.798	135.007	16.892	< .001
SEX	11.546	17.436	5.972	0.662	n.s.
VOW*ACC	−8.691	12.088	136.313	−0.719	n.s.
VOW*PHR	−8.177	11.589	134.958	−0.706	n.s.
VOW*SEX	−25.994	11.619	134.975	−2.237	n.s.
Dina – Diener (all speakers)					
Intercept	344.669	8.064	5.962	42.744	< .001
VOW	34.649	5.299	134.967	6.538	< .001
ACC	−5.330	5.495	135.472	−0.970	n.s.
PHR	81.428	5.348	134.996	15.226	< .001
SEX	10.029	16.127	5.961	0.622	n.s.
VOW*ACC	3.942	11.037	135.783	0.357	n.s.
VOW*PHR	11.244	10.688	134.966	1.052	n.s.
VOW*SEX	−13.708	10.592	134.974	−1.294	n.s.
Dina – Diener (without M2)					
Intercept	342.408	9.669	4.984	35.413	< .001
VOW	34.339	5.657	111.985	6.070	< .001
ACC	−8.667	5.643	112.289	−1.536	n.s.
PHR	87.013	5.542	111.995	15.700	< .001
SEX	14.202	19.347	4.979	0.734	n.s.
VOW*ACC	−0.617	11.360	112.564	−0.054	n.s.
VOW*PHR	20.525	11.078	111.977	1.853	n.s.
VOW*SEX	−13.659	11.285	111.979	−1.210	n.s.
Opa – Oper					
Intercept	361.332	10.421	6.022	34.672	< .001
VOW	11.370	5.784	136.030	1.966	n.s.
ACC	20.541	5.857	136.449	3.507	< .01
PHR	92.830	5.750	136.027	16.145	< .001
SEX	12.920	20.828	6.005	0.620	n.s.
VOW*ACC	13.306	11.662	136.236	1.141	n.s.
VOW*PHR	−3.126	11.495	136.005	−0.272	n.s.
VOW*SEX	10.813	11.461	136.026	0.943	n.s.

References

Adank, Patti, Roel Smits & Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America* 116, 3099–3107.

Barry, William J. 1995. Schwa vs. schwa + /r/ in German. *Phonetica* 52, 228–235.

Barry, William J. 1997. Another R-tickle. *Journal of the International Phonetic Association* 27(12), 35–45.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67(1), 1–48.

Baumann, Stefan, Martine Grice & Ralf Benzmüller. 2005. German intonation in Autosegmental-Metrical Phonology. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 55–83. Oxford: Oxford University Press.

Beckman, Mary E. & Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In Patricia Keating (ed.), *Phonological structure and phonetic form: Papers in Laboratory Phonology III*, 1–33. Cambridge: Cambridge University Press.

Beckman, Mary E., Kenneth Jong, Sun-Ah Jun & Sook-Hyang Lee. 1992. The interaction of coarticulation and prosody in sound change. *Language and Speech* 35(1/2), 45–58.

Boersma, Maarten & Don Weenink. 2016. Praat: Doing phonetics by computer. https://www.fon.hum.uva.nl/praat/.

Bombien, Lasse, Steve Cassidy, Jonathan Harrington, Tina John & Sallyanne Palethorpe. 2006. Recent developments in the Emu speech database system. *Proceedings of the 11th SST Conference Auckland*, 313316.

Browman, Catherine & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49, 155–180.

Byrd, Dani, Abigail Kaun, Shrikanth Narayanan & Elliot Saltzman. 2000. Phrasal signatures in articulation. In Michael Broe & Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon*, 70–87. Cambridge: Cambridge University Press.

Byrd, Dani, Sungbok Lee, Daylen Riggs & Jason Adams. 2005. Interacting effects of syllable and phrase position on consonant articulation. *The Journal of Acoustical Society of America* 118, 3860–3873.

Byrd, Dani & Elliot Saltzman. 1998. Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics* 26, 173–199.

Byrd, Dani & Elliot Saltzman. 2003. The elastic phrase: Modelling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31, 149–180.

Cho, Taehong. 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32(2), 141–176.

Crosswhite, Katherine. 2004. Vowel reduction. In Hayes et al. (eds.), 191–231.

Dittrich, Ricarda & Geske Reibisch. 2006. An acoustic study of /r/-vocalization in word-final position. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*, 19–26.

Ernestus, Mirjam & Harald Baayen. 2006. The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In Louis Goldstein, Douglas Whalen & Catherine Best (eds.), *Laboratory Phonology VIII*, 27–49. Berlin: Mouton de Gruyter.

Espy-Wilson, Carol, Suzanne Boyce, Michel Jackson, Shrikanth Narayanan & Abeer Alwan. 2000. Acoustic modelling of American English /r/. *The Journal of Acoustical Society of America* 108, 343–356.

Faber, Alice & Marianna Di Paolo. 1995. The discriminability of nearly merged sounds. *Language Variation and Change* 7(1), 35–78.

Fay, Johanna. 2010. *Die Entwicklung der Rechtschreibkompetenz beim Textschreiben: Eine empirische Untersuchung in Klasse 1 bis 4*. Frankfurt am Main: Peter Lang.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Hayes et al. (eds.), 233–276.

Floccia, Caroline, Jeremy Goslin, Frédérique Girard & Gabrielle Konopczynski. 2006. Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance* 32(5), 1276–1293.

Fridland, Valerie, Tyler Kendall & Charlie Farrington. 2014. Durational and spectral differences in American English vowels: Dialect variation within and across regions. *The Journal of the Acoustical Society of America* 136(1), 341–349.

Gahl, Susanne. 2008. "Thyme" and "Time" are not homophones: Word durations in spontaneous speech. *Language* 84(3), 474–496.

Gelman, Andrew & Joseph Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Hall, Tracy A. 1992. *Syllable structure and syllable-related processes in German* (Linguistische Arbeiten 276). Berlin: Walter de Gruyter.

Hall, Tracy A. 1993. The phonology of German /ʀ/. *Phonology* 10(1), 83–105.

Harrington, Jonathan. 2010. Acoustic phonetics. In William Hardcastle, John Laver & Fiona Gibbon (eds.), *The handbook of phonetic science*, 2nd edn., 81–129. Oxford: Blackwell.

Harrington, Jonathan, Janet Fletcher & Corinne Roberts. 1995. Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics* 23, 305–322.

Harrison, Xavier, Lynda Donaldson, Maru Correa, Julian Evans, David Fisher, Cecily Goodwin, Beth Robinson, David Hodgson & Richard Inger. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 4794.

Hay, Jen, Katie Drager & Brynmor Thomas. 2013. Using nonsense words to investigate vowel merger. *English Language and Linguistics* 17(2), 241–269

Hay, Jen, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4), 458–484.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.). 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

Heike, Georg. 1972. Quantitative und qualitative Differenzen von /a. . ./-Realisationen im Deutschen. *The 7th International Congress of Phonetic Sciences* (ICPhS VII), 725–729.

Hoole, Philip & Christine Mooshammer. 2002. Articulatory analysis of the German vowel system. In Peter Auer, Peter Gilles & Helmut Spiekermann (eds.), *Silbenschnitt und Tonakzente*, 129–152. Tübingen: Niemeyer.

Jong, Kenneth, Mary E. Beckman & Jan Edwards. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36(2–3), 197–212.

Katz, Leonard & Ram Frost. 1992. The reading process is different for different orthographies: The orthographic depth hypothesis. In Ram Frost & Leonard Katz (eds.), *Orthography, phonology, morphology, and meaning*, 67–84. Amsterdam: Elsevier.

Kohler, Klaus. 1990. Illustrations of the IPA: German. *Journal of the International Phonetic Association* 20(1), 48–60.

Kohler, Klaus. 1995. *Einfuehrung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.

Kuznetsova Alexandra, Per B. Brockhoff & Rune H. Christensen. 2017. *LmerTest*: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13), 1–26.

Labov, William. 2001. *Principles of linguistic change*. Oxford: Blackwell.

Labov, William & Maciej Baranowski. 2006. 50 msec. *Language Variation and Change* 18, 1–18.

Labov, William, Malcah Yaeger & Richard Steiner. 1972. *A quantitative study of sound change in progress*. Philadelphia, PA: The U.S. Regional Survey.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Lawson, Eleanor, James Scobbie & Jane Stuart-Smith. 2013. Bunched /r/ promotes vowel merger to schwa: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics* 41, 198–210.

Lawson, Eleanor, Jane Stuart-Smith & James Scobbie. 2018. The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study. *The Journal of the Acoustical Society of America* 143, 1646.

Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35, 1773–1781.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer.

Lindblom, Björn, Augustine Agwuele, Harvey Sussman & Eir E. Cortes. 2007. The effect of emphatic stress on consonant vowel coarticulation. *The Journal of the Acoustical Society of America* 121(6), 38023813.

Lindblom, Björn & Johan Sundberg. 1971. Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America* 50(4B), 1166–1179.

Macmillan, Neil A. & Douglas C. Creelman. 1991. *Detection theory: A user's guide.* Cambridge: Cambridge University Press.

Meinhold, Gottfried. 1989. Das problematische [ɐ]. In Edith Slembek (ed.), *Von Lauten und Leuten: Zeitschrift für Peter Martens zum 70. Geburtstag*. Frankfurt am Main.

Moisik, Scott. 2006. *The L2 phonetic acquisition of nuclear-/r/: The case of English and German*. Calgary: University of Calgary Press.

Mooshammer, Christine & Susanne Fuchs. 2002. Stress distinction in German: Simulating kinematic parameters of tongue tip gestures. *Journal of Phonetics* 30, 337–355.

Mooshammer, Christine & Christian Geng. 2008. Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association* 38(2), 117–136.

Nimz, Katharina, Katharina Immel & Kai O. Koop. 2016. Wer die Qual hat, hat keinen Wal. Presented at the 12. Phonetik und Phonologie im deutschsprachigen Raum, LMU/Munich.

Nycz, Jennifer. 2013. New contrast acquisition: Methodological issues and theoretical implications. *English Language and Linguistics* 17(2), 325–357.

Padgett, Jaye & Marija Tabain. 2005. Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica* 62(1), 14–54.

Piroth, Hans G. & Peter M. Janker. 2004. Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics* 32, 81–109.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.r-project.org/index.html.

Russ, Charles V. J. 1990. *The dialects of modern German*. London: Routledge.

Schmidlin, Regula. 2002. Zum Orthographieerwerb bei norddeutschen, südwestdeutschen und schweizerischen Kindern: Schärfungsmarkierung und r-Schreibung. Ergebnisse einer empirischen Untersuchung. In Doris Tophinke & Christa Röber-Siekmeyer (eds.), *Schärfungsschreibung im Fokus*, 170–185. Baltmannsweiler.

Simpson, Adrian P. 1998. Accounting for the phonetics of German r without processes. *ZAS Papers in Linguistics* 11, 91–104.

Sluijter, Agaath M. & Vincent J. van Heuven. 1996. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Fourth International Conference on Spoken Language Processing* (ICSLP).

Solé, Maria J. & John J. Ohala. 2010. What is and what is not under the control of the speaker: Intrinsic vowel duration. In Cecile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.), *Papers in Laboratory Phonology 10*, 607–655. Amsterdam: De Gruyter Mouton.

Traunmüller, Hartmut. 1981. Perceptual dimension of openness in vowels. *The Journal of the Acoustical Society of America* 69, 1465–1475.

Ulbrich, Christiane & Hobst Ulbrich. 2007. The realisation of /r/ in Swiss German and Austrian German. *Proceedings of the 16th International Congress of Phonetic Sciences* (ICPhS XVI), 1761–1764.

Ulbrich, Hobst. 1973. Zur Kodifizierung der R-Aussprache im Siebs. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 26, 120–133.

Ulbrich, Hobst. 2002. R-Aussprache 1966 und 1996 – stabile und instabile Realisierungsmodi. *Sprache und Sprechen* 34, 143–152.

Vennemann, Theo. 1991. Syllable structure and syllable cut prosodies in Modern Standard German. In Pier Marco Bertinetto, Michael Kenstowicz & Michele Loporcaro (eds.), *Cartamen Phonologicum II: Papers from the 1990 Cortona Phonology Meeting*, 211–243. Torino: Rosenberg and Sellier.

Viëtor, Wilhelm. 1901. *Kleine Phonetik des Deutschen, Englischen, und Französischen*. Leipzig: O. R. Reisland.

Wade, Lacey. 2017. The role of duration in the perception of vowel merger. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1), 30.

Warner, Natasha, Allard Jongman, Joan A. Sereno & Rachèl Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32, 251–276.

Wells, John. 1996. German SAMPA. https://www.phon.ucl.ac.uk/home/sampa/german.htm.

Wiese, Richard. 2000. *The phonology of German*. Oxford: Oxford University Press.

Wiese, Richard. 2001. The phonology of /r/. In Tracy Alan Hall (ed.), *Distinctive feature theory*, 335–368. Berlin: de Gruyter.

Wiese, Richard. 2003. The unity and variation of (German) /r/. *Zeitschrift für Dialektologie und Linguistik* 70, 25–43.

Winkelmann, Raphael, Klaus Jänsch, Steve Cassidy & Jonathan Harrington. 2017. *EmuR*: Main Package of the EMU Speech Database Management System R package version 0.2.3.