The effect of prosodic weakening on the production and perception of trans-consonantal vowel coarticulation in German

Jonathan Harrington,[a)] Felicitas Kleber, and Ulrich Reubold
Institute of Phonetics and Speech Processing, University of Munich, Schellingstraße 3, 80799 München, Germany.

(Received 23 July 2012; revised 11 February 2013; accepted 20 May 2013)

The present study considers whether coarticulation in production and its relationship to categorization could provide a synchronic basis for the prevalence of sound change in unstressed syllables. The size of V_2-on-V_1 coarticulation in the production of /pV_1pV_2l/ non-words (V_1 = /ʊ, ʏ/ and V_2 = /e, o/) produced by German speakers and with stress falling either on the first or second syllable was compared with forced-choice perceptual categorization of resynthesized versions of these non-words. In speech production, /ʏ/ but not /ʊ/ was perturbed by anticipatory V_2-on-V_1 coarticulation. Stress had no influence on coarticulation but caused target undershoot in /ʊ/. The same speakers compensated for coarticulation in perception: however, in the unstressed context the speakers compensated less and their diminished compensatory coarticulation was shown to be linked to /ʊ/-undershoot. Taken together, these results point to a mismatch between coarticulation and categorization that is suggested as a possible source of sound change: whereas de-stressing did not affect V_2-on-V_1 coarticulation in production, it weakened V_2's influence on perceptual /ʊ-ʏ/ categorization. The evidence that this mismatch is indirectly caused by stress-dependent reduction in /ʊ/ that is unrelated to the V_2-source of the coarticulation is also consistent with a model of sound change as non-teleological. *© 2013 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4808328]

PACS number(s): 43.70.Mn, 43.71.Es, 43.71.An [BRM]   Pages: 551–561

I. INTRODUCTION

The variation in speech production is infinite, but it has nevertheless been shown to be lawful (Lindblom, 1963; Lindblom and MacNeilage, 2011) and often subject to constraints in speech production and perception and in the relationship between the two. Although much of this lawful variation is reflected in the types of sound changes that have been documented in many languages of the world (Beddor, 2009), the mechanisms by which phonetic variability becomes phonologized as a categorical change are still far from understood (see Huffman, 2012, and Hualde *et al.*, 2011, for recent discussions). There have broadly been two main approaches to resolving this issue using experimental methods. One of these is due to sociolinguistic (Labov, 1994, 2001) and sociophonetic (Jannedy and Hay, 2006) investigations that are concerned with how sound change spreads across different speaker groups and generations. The other is more closely concerned with the phonetic conditions that give rise to sound change (Ohala, 1981, 1993; Solé, 2012) and the extent to which these same conditions also shape the distribution of the patterns of sounds in the world's languages. The present investigation is rooted in the second of these traditions and is more specifically concerned with the way in which reduction and coarticulation facilitate diachronic change in the vowels of prosodically weak constituents.

In Bybee's (2002) model of lexical diffusion, sound change arises through the articulatory processes of reduction as a result of highly practiced neuromotor activity (see also Mowrey and Pagliuca, 1995) whereas in other models (Beddor, 2009; Lindblom *et al.*, 1995; Kleber *et al.*, 2012; Ohala, 1993), the perceptual evaluation of articulatory reduction and coarticulation is considered to play a much greater role in bringing about sound change. For both Lindblom *et al.* (1995) and Ohala (1993), it is the listener's de-contextualisation of phonetic variation that can be a catalyst for diachronic change. In Lindblom *et al.* (1995), the de-contextualization is brought about when listeners exceptionally pay attention to phonetic variation in semantically predictable and hypoarticulated parts of the utterance that are subject to extensive top-down processing. It is when top-down processing at such semantically redundant points of the signal is disengaged that new pronunciations can suggest themselves as potential variants to be added to the listener's lexicon. This idea foreshadows exemplar theory (Johnson, 1997; Pierrehumbert, 2003) according to which lexical representations are shaped and accrued from auditory traces accumulated in memory. The model of Lindblom *et al.* (1995) is also consistent with the prediction that prosodically weak constituents should be subject to sound change given that these are often semantically redundant sites at which top-down processing is likely to be heavily engaged (and potentially disengaged leading to the potential for sound change). The de-contextualization in Ohala (1993) on the other hand can come about if listeners do not attribute enough coarticulatory variation to the contextual source from which it arises. Thus tonogenesis in certain South East Asian languages may have arisen because listeners incorrectly parsed the coarticulatory intrinsic pitch variation with the vowel instead of with the source (the voicing difference in the preceding stop that was subsequently neutralized) that gave rise to it (Hombert *et al.*, 1979).

[a)]Author to whom correspondence should be addressed. Electronic mail: jmh@phonetik.uni-muenchen.de

It has been shown that prosodically weak constituents are both frequent sites of diachronic change (Beckman *et al.*, 1992) and subject to extensive synchronic variation (Kohler, 2001); but what has so far not been explored is whether prosodic weakening is accompanied by a change in how coarticulation is processed in speech perception that could, in turn, set the conditions for sound change to take place. Coarticulatory influences in production have been shown to be greater in prosodically weak than strong vowels at different levels of the stress hierarchy, including both lexically unstressed vs stressed (Mooshammer and Geng, 2008; van Bergem, 1994) and deaccented vs nuclear accented positions (Cho, 2004; Harrington *et al.*, 2000). If listeners do not compensate for this greater coarticulatory influence, then just such a mismatch between the production and perception of coarticulation would arise that could, according to Ohala's (1993) model, make weak constituents unstable and prone to change. So far, this issue is unresolved if only because listener compensation for coarticulation has been extensively studied in relation to segmental (Mann and Repp, 1980; Fujisaki and Kunisaki, 1976; Ohala and Feder, 1994) but not prosodic variation (although see Fowler, 1981, for an analysis of the perception of coarticulation in unstressed vowels).

In the present paper, the variation in production and perception due to stress is explored with reference to trans-consonantal (VCV) vowel coarticulation. Öhman (1966) was one of the first to show that vowels influence each other across an intervening consonant and subsequent perception experiments (Alfonso and Baer, 1982; Fowler and Smith, 1986; Martin and Bunnell, 1982) have shown that listeners are sensitive to this variation. Languages differ in the extent of VCV coarticulation (Manuel, 1999; Öhman, 1966) and such differences are related to the different degrees with which listeners of different language backgrounds compensate perceptually for this variation (Beddor *et al.*, 2002). As far as stress is concerned, a physiological study by Cho (2004) showed that vowels in prosodically strong positions were influenced less by trans-consonantal vowel coarticulation than vowels in weaker positions while Beddor *et al.* (2002) report that stressed vowels have a strong trans-consonantal coarticulatory influence on weak vowels in English (see also Cole *et al.*, 2010, for a recent review on VCV coarticulation and how it is affected by stress). For the present study, one of the main aims will be to determine whether unstressed vowels are perturbed more by the coarticulatory influences of stressed vowels than vice versa in production and also whether this greater coarticulatory perturbation (if present) is accompanied by corresponding adjustments to listeners' compensation for these effects. To do so, we measured the production and perception of anticipatory trans-consonantal vowel coarticulation in /pVpVl/ non-words produced and perceived by the same 20 subjects.

II. SPEECH PRODUCTION

A. Method

1. Materials, speakers, segmentation, and labeling

The materials included German non-words with a /pV_1pV_2l/ structure where V_1 was lax /ʏ/ or /ʊ/, V_2 was tense /o/ or /e/, and the nuclear accent was either on the first or second syllable (e.g., /ˈpʏpol/, /pʊˈpel/, etc). The phonetic quality of these Standard German monophthongs corresponds approximately with those of the International Phonetic Alphabet, i.e., /e, o/ are monophthongal and phonetically very similar to cardinal vowels 2 and 7, respectively; and /ʏ,ʊ/ are slightly centralized, mid-high, front, and back rounded vowels, respectively. /p/ was chosen as the intermediate consonant because it interferes minimally with tongue articulation, although acoustically it can induce a perceptual retracting effect on vowels due its low labial F2-locus (Stevens and House, 1963). The final lateral is clear in Standard German, i.e., there is much less tongue-body raising than in syllable-final /l/ in many English varieties (Pouplier, 2012). Changing the stress pattern in the way that was done for the present experiment has a minimal effect on vowel quality, i.e., there is much less vowel centralization in non-pitch-accented compared with pitch-accented rhythmically strong vowels in German (Mooshammer and Geng, 2008) than in languages in which full vowels may be reduced to /ə/ such as English.

Twenty speakers (6 male, 14 female) with no known speech and hearing difficulties produced ten repetitions of the carrier phrase "ich habe pV_1pV_2l gesagt" (lit: "I have pV_1pV_2l said") with the nuclear accent falling on the target word and with the pitch accent occurring either on the first or the second syllable: The first of these in which V_1 is pitch-accented will be referred to as the stressed condition; and the second in which V_2 is pitch-accented as the unstressed condition. The age range of the speakers was 20–46 yr, and their median age 26 yr, 2 months. All speakers were resident in Munich at the time of the recording; half were students of phonetics; the other half had no prior training in phonetics or speech sciences. All subjects were speakers of Standard German. Recordings were made using the SpeechRecorder software (Draxler and Jänsch, 2004) and a stereo headset (Sennheiser pc 165 USB) in a sound attenuated booth at the Institute of Phonetics and Speech Processing at the University of Munich. Ten repetitions of the eight test sentences (2 target vowels × 2 vocalic contexts × 2 stress conditions) and six filler sentences (of a similar structure but with different consonants and vowels) were presented to the speaker in randomized order on a computer screen. Each sentence was introduced by an orthographically presented question of the type "*Hat Maria PÜPpehl oder PUPpehl gesagt?*" (lit. Did Mary say /ˈpʏpel/ or /ˈpʊpel/) or "*Hat Maria püpPEHL oder püpPOHL gesagt?*" (lit. Did Mary say /pʏpˈol/ or /pʏpˈel/) in order to elicit stress either on the first or second syllable.

The materials were automatically segmented and labeled using the Munich automatic segmentation system (Schiel, 2004). All the boundaries around the target /ʊ,ʏ/ vowel of the first syllable were manually checked and if need be adjusted such that they marked the vowel's periodic onset and offset. The total number of analyzed tokens was 2 (V_1 = /ʊ,ʏ/) × 2 (V_2 = /e, o/) × 2 (stress) × 20 (speakers) × 10 repetitions = 1600.

2. Acoustic parameters

Formant frequencies were calculated (LPC order of 10, a pre-emphasis of 0.95, 25 ms Blackman window with a frame shift of 5 ms) for the target words. Roughly 70% of the formants in the database required manual correction, often because in the back vowel /ʊ/ F3 had been mis-tracked as F2.

In view of these inherent difficulties in tracking formants accurately, an additional acoustic analysis of /ʊ,ʏ/ was carried out based on a two-dimensional space of the spectral slope × curvature. In general, these vowels were expected to be quite well distinguished in this space for two reasons: First because /ʊ/ has a more steeply falling spectrum up to 2 kHz than /ʏ/ (due to the greater F2–F1 proximity in back vs front vowels); and second because /ʏ/ has a more compact (and therefore more curved) spectrum than /ʊ/ in the frequency range 1.5–2.5 kHz in which F2 and F3 of /ʏ/ tend to be located. The spectrum was calculated from the time signal using an FFT with a frame shift of 5 ms and a 1024 point Blackman window giving a frequency resolution of 21.5 Hz. The spectral slope and curvature were derived from the first and second coefficients of the discrete cosine transformation (Harrington *et al.*, 2008; Nossair and Zahorian, 1991; Watson and Harrington, 1999; see also Milner and Shao, 2006, for the relationship to mel-frequency cepstral coefficients) applied to these spectral data between 0.2 and 2.2 kHz after warping the frequency axis to the mel scale (Fant, 1973). The motivation for the upper limit of 2.2 kHz, which mostly excluded F3 of /ʊ/ but was close to the F3 center frequency of /ʏ/ for most speakers, was because it was found to be optimal for separating these vowels. For an N-point dB-spectrum, $s(n)$, whose N points were equally spaced on the mel axis over the frequency range 0.2–2.2 kHz, the mth DCT coefficient, C_m ($m = 1,\ 2$), was calculated from

$$C_m = \frac{2}{N}\sum_{n=0}^{N-1} s(n)\cos\left[\frac{(2n+1)m\pi}{2N}\right]. \tag{1}$$

Thus each vowel was represented by a pair of points (C_1, C_2) that coded the spectral slope and curvature, respectively, every 5 ms between the vowel's acoustic onset and offset.

We then applied principal components analysis to this two-dimensional DCT-space separately to the stressed and unstressed vowels but pooled across speakers. The purpose of doing so was to determine the extent of similarity between this PCA-transformed DCT-space and that of the first two formant frequencies in the same data. The left panel of Fig. 1 shows that /ʊ,ʏ/ were effectively separated based on F2 whereas F1 provided no additional information for their distinction. The middle panel of Fig. 1 shows how the spectral slope (horizontal axis) and curvature (vertical axis) distinguish between these vowel types in the way that was expected: thus, they are distinguished on linear spectral slope because the spectrum falls off more steeply for /ʊ/; and they are distinguished on spectral curvature because of the greater curvature in /ʏ/ compared with /ʊ/. The results of the PCA-transformation of these DCT-data in the middle panel is shown in the right panel of Fig. 1. These data show first that the first PCA-transformed dimension, PCA_1, provides an effective separation between the two vowel types. The same data show the remarkably close correspondence to the formant data in the left panel. Thus, apart from a rotation by 180°, PCA_1 and F2 provide a very similar separation between the two vowel types; while PCA_2 just like F1, provides no effective separation between these vowel types. The very close match ($r = 0.90$) between F2 and PCA_1 derived independently from the DCT-coefficients is a further confirmation of the accuracy of the subsequently manually corrected formant frequencies.

In the following production experiments, results are reported for F2 and PCA_1 extracted at the acoustic temporal midpoint of V_1. Given the similar way in which F2 and PCA_1 distinguished between /ʊ,ʏ/ as well as the high degree of correlation between these parameters, the expectation was that the extent of coarticulation and influence of stress would be similarly manifested on F2 and PCA_1 parameters.

3. Classification

In order to match more closely the production data to the results of the perception experiment in which listeners provided binary forced-choice /ʊ,ʏ/ responses to stimuli (Sec. III), Gaussian classifiers were applied to the acoustic data in the two stress × V_2 contexts. This process included a training stage in which normal distributions were fitted separately to these four contexts from which the posterior probabilities $p(ʊ|A)$ and $p(ʏ|A)$ were derived that any given acoustic value, A, is a member of the category /ʊ/ or /ʏ/ [in all cases $(pʊ|A) + p(ʏ|A) = 1$]. For these data, A was either F2 in mel or PCA_1 extracted at the temporal acoustic midpoint of V_1. The posterior probabilities were then used to derive four /ʊ-ʏ/ decision boundaries separately for the V_1 data in the two stress × V_2 contexts. In order to do so, linear regression was used to estimate the slope, m, and intercept, k separately in these four contexts from (2):

$$\log[p(ʏ|A)/p(ʊ|A)] = mA + k, \tag{2}$$

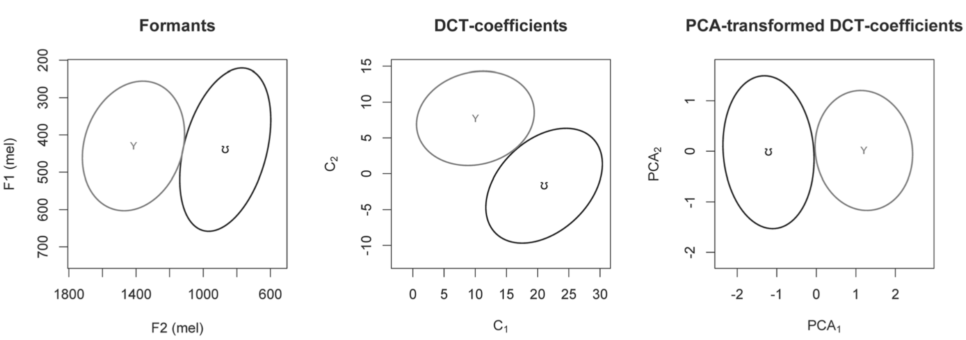

FIG. 1. Shown are the 95% confidence ellipses for data extracted at the temporal midpoint of /ʊ,ʏ/ pooled across all speakers and both prosodic contexts in the plane of (left) the first two formant frequencies in mel, (middle) DCT-derived, mel-scaled spectral slope (C_1) and spectral curvature (C_2), and (right) the first two principal components applied to the data in the middle panel.

from which the sigmoid function was derived in (3):

$$p(\text{ʏ}|A) = \mathrm{e}^{(mA+k)}/[1+\mathrm{e}^{(mA+k)}], \tag{3}$$

using the relationship $p(\text{ʊ}|A) = 1 - p(\text{ʏ}|A)$. The decision boundary was given by $-k/m$ which is the value of A in (3) for which $p(\text{ʏ}|A) = p(\text{ʊ}|A) = 0.5$. The main purpose of (3) was to obtain a probabilistic representation of the production data that could be matched more closely to the (sigmoid-based) psychometric functions derived from the perception experiments (Sec. III). The slope in (3), m, provides information about the probabilistic separation between the two vowel categories: as m tends to 0, the sigmoid function tends towards a straight line, and the degree of separation between the two categories is reduced.

B. Results

The data in Fig. 2 show F2 and PCA_1 trajectories for both V_1 and V_2 separately for the two stress conditions in which each vowel token was linearly time-normalized between its acoustic onset and offset. The data were then ensemble-averaged across the speakers. For both parameters, Fig. 2 suggests that V_2 exerted a coarticulatory influence on /ʏ/ as shown by the separation of /ʏ/ in the two V_2-contexts. The data in Fig. 2 also show that the size of this anticipatory /ʏ/-on-V_1 influence was very similar in stressed and unstressed contexts. Finally, Fig. 2 also suggests that the /ʊ/ was perturbed only minimally by V_2 (given the similarity of the /ʊ/-trajectories in both V_2-contexts).

The data in Fig. 3 of F2 and PCA_1 extracted at the vowels' acoustic temporal midpoints show similar trends to those in Fig. 2. First, V_2 influenced /ʏ/ as shown by the different

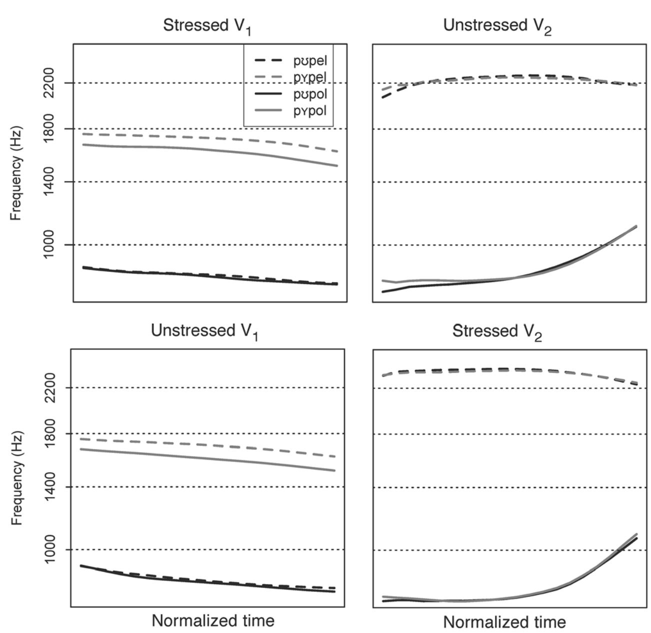

FIG. 2. The second formant frequency in Hz as a function of time aggregated across all speakers and linearly time normalized between the acoustic onset and offset of (left) V_1 and (right) V_2 shown separately for the stress conditions(top) /ˈpV_1pV_2l/ and (bottom) /pV_1ˈpV_2l/. The vowel contexts are /pʊpol/ (solid black), /pʏpol/ (solid gray), /pʊpel/ (dashed black), and /pʏpel/ (dashed gray).

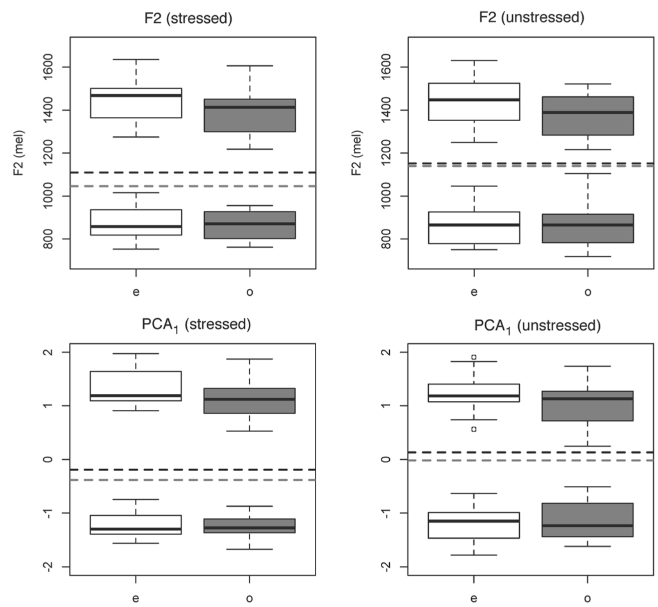

FIG. 3. Boxplots (in which the rectangles span the inter-quartile range) of the distribution of V_1 = /ʊ,ʏ/ in V_2 = /e/ (white) and V_2 = /o/ (gray) contexts when V_1 was (left) stressed and (right) unstressed on (top) F2 and (bottom) PCA_1 at the temporal midpoint of V_1. The 50% /ʊ-ʏ/ decision boundaries in the /e/ (black) and /o/ (gray) contexts are shown as horizontal lines.

location of the distributions for /ʏ/ in the two contexts. Second, Figs. 2 and 3 suggests that the magnitude of this V_2-on-/ʏ/ coarticulatory influence was about the same in stressed and unstressed contexts (since, as shown by Fig. 3, the difference in the distributions of /ʏ/ in the two V_2 contexts was similar between stressed and unstressed conditions). Third, V_2 had a negligible influence on /ʊ/, given the similar location of the distributions for /ʊ/ in these two contexts. Fourthly, there is some evidence from Fig. 3 of a greater variability in the unstressed compared with the stressed context especially for /ʊ/ in the V_2 = /o/ context).

The observations from Fig. 3 were confirmed by the results of a repeated measures analysis of variance (ANOVA) with either F2 or PCA_1 (both extracted at the temporal midpoint of V_1) as the dependent variable and independent factors V_1 (two levels: /ʏ, ʊ/), stress (two levels: stressed, unstressed), and V_2 (two levels: /e, o/). Apart from the predictably large significant influence of V_1 on the acoustic parameters (because /ʊ, ʏ/ occupy different parts of the acoustic space), there was a significant influence of V_2 [for F2: $F(1,19) = 20.0$, $p < 0.001$; for PCA_1: $F(1,19) = 10.2$, $p < 0.01$], no effect of stress, and a significant $V_1 \times V_2$ interaction [for F2: $F(1,19) = 19.5$, $p < 0.001$; for PCA_1: $F(1,19) = 31.5$, $p < 0.001$]. The interaction evidently comes about because, as Fig. 3 shows, there was a coarticulatory influence of V_2 on /ʏ/ but not on /ʊ/. The results of *post hoc* Bonferroni-adjusted *t*-tests confirmed this trend from Fig. 3: there was a significant influence of V_2 on /ʏ/ (F2: $t_{19} = 7.2$, $p_{adj} < 0.001$; PCA_1: $t_{19} = 5.8$, $p_{adj} < 0.001$) but not on /ʊ/.

Although this study is concerned with the V_2 coarticulatory influences on V_1, the results of a repeated measures ANOVA on V_2 with the same factors as for V_1 and with the dependent variable F2 extracted at the temporal midpoint of V_2 are as follows. There was a significantly [$F(1,19) = 144.1$,

$p < 0.001$] predictable influence of V_2 on F2 (since /e,o/ occupy different regions of the vowel space) but no influence of V_1 nor any overall influence of stress on F2 at the midpoint of V_2. However, there was a V2 × stress interaction [$F(1,19) = 7.2$, $p < 0.001$] which comes about because, as Fig. 2 shows, stress expanded the V2 vowel space such that F2 of /e/ and F2 of /o/ were further apart in a stressed than in an unstressed context (compare the top and bottom right panels of Fig. 2): thus the *post hoc* tests showed a significant influence of stress on the temporal midpoint of F2 in both /e/ ($t_{19} = 8.0$, $p_{adj} < 0.001$) and in /o/ ($t_{19} = 6.4$, $p_{adj} < 0.001$). However, this stress-dependent shift of F2 in /e, o/ evidently had no influence on V_2-on-V_1 coarticulation given that, as the results of the previous paragraph show, there was no influence of stress on V_1.

As far as the decision boundaries between /ʊ/ and /ʏ/ derived from (3) are concerned (and shown by the horizontal dashed lines in Fig. 3 and the vertical dashed lines in Fig. 4), these were influenced by V_2-context: That is, the decision boundaries in a V_2 = /o/ context were shifted towards lower F2 or PCA_1 values than those in a V_2 = /e/ context. They were also influenced by stress: The first obvious effect is that they were shifted for both V_2-contexts into the /ʏ/-space. This stress-induced shift is likely to come about because of the greater variability particularly of /ʊ/ in the unstressed context: As a result /ʊ/ encroaches to a greater extent on /ʏ/ in the unstressed context causing a corresponding shift in the decision boundaries towards /ʏ/. The greater variability of the data in the unstressed context particularly for /ʊ/ was confirmed by a Levene test for the homogeneity of variance with F2 or PCA_1 of the vowel tokens as the dependent variable and stress as the independent factor applied separately to the /ʊ/ and /ʏ/ data. These results show that the variance was significantly greater in the unstressed context for /ʊ/ [F2: $F(1,785) = 11.1$, $p < 0.001$; PCA_1: $F(1,785) = 53.4$, $p < 0.001$]. For /ʏ/, the variance was significantly greater in the unstressed than in the stressed context for PCA_1 [$F(1,790) = 7.8$, $p < 0.01$] but not for F2. Thus in general, the results are consistent with the evidence in Fig. 3 that the variability was greater in the unstressed than in the stressed context, and especially so for V_1 = /ʊ/.

Figure 4 also shows that the sigmoid functions derived from (3) were closer together in the unstressed than the stressed contexts: Thus the area between the sigmoids was smaller in the unstressed than in the stressed context for greater probabilistic classifications of the data as /ʏ/ as opposed to /ʊ/, especially so for F2. The reason for their greater proximity in the unstressed context is because of the increase in the variability of /ʊ/, especially for V_2 = /o/. Consequently, there was greater uncertainty in the probabilistic /ʊ-ʏ/ categorizations that is manifested as a flattening of the sigmoid function in the unstressed context especially for V_2 = /o/ (compare the gray-colored sigmoids in the stressed and unstressed contexts in Fig. 4): It is this flattening that caused a greater approximation between the two sigmoids, with the result that they were closer together in just the part of the vowel space in which V_2 exerted its greatest coarticulatory influence on /ʏ/.

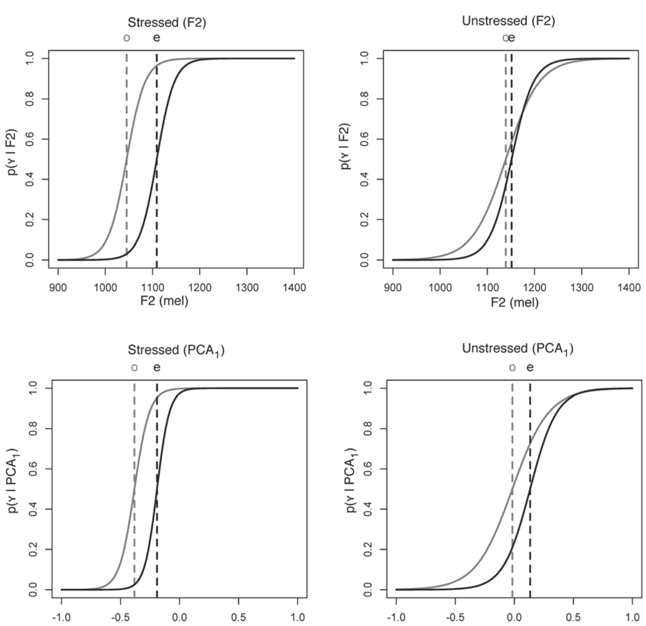

FIG. 4. Sigmoid functions of /ʊ-ʏ/ categorizations based on (top) F2 and (bottom) PCA_1 extracted at the temporal midpoint of V_1 in the context of V_2 = /o/ (gray) and V_2 = /e/ (black) shown separately for (left) stressed and (right) unstressed V_1. The 50% decision boundaries in the two V_2 contexts are shown as vertical dashed lines. The *y* axis is the posterior probability of /ʏ/ (given F2 or PCA_1), i.e., the proportion of F2 or PCA_1 categorized as /ʏ/ as opposed to /ʊ/.

C. Discussion

The results have shown anticipatory V_2-on-V_1 coarticulation in the front /ʏ/ but not in the back /ʊ/ vowel. However, there was no evidence that the magnitude of coarticulation was greater in unstressed than in stressed syllables. There were, however, three influences of stress on the probabilistic categorization of the production data. First, the /ʊ-ʏ/ decision boundaries in the two V_2 = /e, o/ contexts were shifted into the /ʏ/ space in the unstressed context, i.e., a greater part of the vowel space was given over to categorizations as /ʊ/. Second, the derived sigmoid functions in the two vowel contexts were flatter for the unstressed compared with stressed contexts, especially for V_2 = /o/. Third, the sigmoid functions were separated to a lesser degree for increasing proportional classifications as /ʏ/ in the unstressed context. All of these three changes in the categorizations can be related both to the shift in the production data of /ʊ/ towards /ʏ/ and to the greater variability of the data in the unstressed compared with the stressed context, especially for /ʊ/.

The main issue in the next section was to consider the extent to which the influence of stress on the perception of coarticulation was matched by these findings from speech production. If listeners' perceptual categorizations matched those from production, then their /ʊ-ʏ/ cross-over boundaries in the unstressed compared with the stressed condition should be shifted into the /ʏ/ space; they should also be flatter, and somewhat closer together for increasing /ʏ/-categorizations, just as they were found to be in speech production.

III. SPEECH PERCEPTION

A. Method

A male speaker of Standard German with slight South German regional characteristics (the third author of this paper) produced utterances containing /ˈpʏpel/, i.e., a front V_1 in a stressed syllable in the context of a front V_2, and /ˈpʊpol/, i.e., a back V_1 in a stressed syllable in the context of a back V_2, embedded in the carrier phrase "ich habe ___ gesagt" (literally: "I have ___ said") with nuclear accent on the target word. An 11-step continuum was created between /pʊp/ (taken from /ˈpʊpol/) and /pʏp/ (taken from /ˈpʏpel/) by using the spectral morphing method of STRAIGHT (Kawahara *et al.*, 2009); to ensure a continuous variation of F2 in equal Hz steps in the continuum, so-called frequency anchors were placed in the spectra of the vowels' temporal midpoints of the two original /pʊp/ and /pʏp/ tokens at 826 and 1441 Hz, respectively. Continuously varying transitions toward this midpoint values were introduced by setting frequency anchors at equally distributed temporal anchors throughout the syllables. The segmental durations were set to the arithmetic mean of the originally produced phonetic segments by means of STRAIGHT's temporal morphing in order to ensure that segment durations remained identical throughout all stimuli (which was important for the pitch manipulation described below).

The initial syllable of /ˈpʊpol/ and /ˈpʏpel/, respectively, was replaced with this 11-step continuum and the words were spliced in the same "ich habe ___ gesagt" utterance, thereby creating an /ʊ-ʏ/ continuum in V_2-fronting ('ich habe /ˈpVpel/ gesagt') and V_2-backing ('ich habe /ˈpVpol/ gesagt') contexts. For both of these stressed continua, /ˈpV_1pV_2l/ was nuclear-accented with a trochaic stress pattern, i.e., with a pitch-accent on the first syllable. The unstressed continua were derived from these by shifting the pitch-accent to the second syllable (Fig. 5) with Praat's (Boersma and Weenink, 2011) implementation of the Pitch-Synchronous-Overlap-and-Add-algorithm (Moulines and Charpentier, 1990) such that /pV_1ˈpV_2l/ was also nuclear-accented but with an iambic stress pattern. Other than these pitch differences, the target words were acoustically identical across the stressed and unstressed conditions. In all cases, V_1 and V_2 were full vowels, and there was no vowel centralization in the unstressed continua. The 44 sentences [11 (steps) × 2 (V_2-context) × 2 (stress)] were repeated 10 times, and presented in a randomized order in a two-alternative forced choice identification test in which subjects classified each stimulus depending on V_2 either as /pʊpel/ or /pʏpel/ and /pʊpol/ or /pʏpol/, respectively, by clicking on the corresponding orthographical representation, namely, "puppehl," "püppehl," "puppohl," and "püppohl." The listeners were the same subjects who had participated in the speech production experiment. The stimuli were presented to listeners over headphones in a quiet room. There was no time-pressure to complete the response.

Psychometric (sigmoid) functions of the form in (4) were fitted separately to the perceptual classifications in each of the four contexts for the V_2 = /e, o/ × stressed/unstressed contexts:

$$p_ʏ = e^{(mx+k)}/[1 + e^{(mx+k)}] \quad (4)$$

in which $p_ʏ$ was the proportion of /ʏ/ responses ($p_ʏ + p_ʊ = 1$), x an acoustic parameter on the 11-step continuum, and m and k the slope and intercept, respectively, of the sigmoid that were estimated using logistic regression, as described below. In order to provide as close a match as possible to the acoustic data, x was calculated using the same DCT and PCA transformations that had been applied to the production data in Sec. II A 2. More specifically, a spectral slice was extracted separately at the temporal midpoint of each of the 11 V_1 stimuli in the continuum and from these the DCT coefficients were calculated over the mel-scaled spectrum in the frequency range 0.2–2.2 kHz using (1). As a result, each of the 11 stimuli was represented by a pair of points that coded the spectral slope and curvature to which principal component analysis was applied: The 11 values of x in (4) were those of the PCA_1, i.e., the first rotated dimension that was obtained after applying principal component analysis to the slope and curvature parameters (exactly analogous to the procedure used in speech production).

The intercept, k, and slope, m, in (4) were estimated using logistic regression and a generalized linear mixed model with the listener entered into the model in such a way to calculate a listener-specific intercept, k_S, and a listener specific slope m_S (i.e., one intercept and one slope for each listener).[1] These parameters k, m, k_S, m_S were calculated separately for the four V_2× stressed/unstressed contexts. The data displayed for all listeners together in Fig. 6 below were based on k and m calculated across all listeners; the statistics were applied in all cases to the listener-specific k_S, m_S values. Given that the results from the production data in Sec. II had shown an exceptionally close match between PCA_1 and F2, the results for speech perception and their relationship to

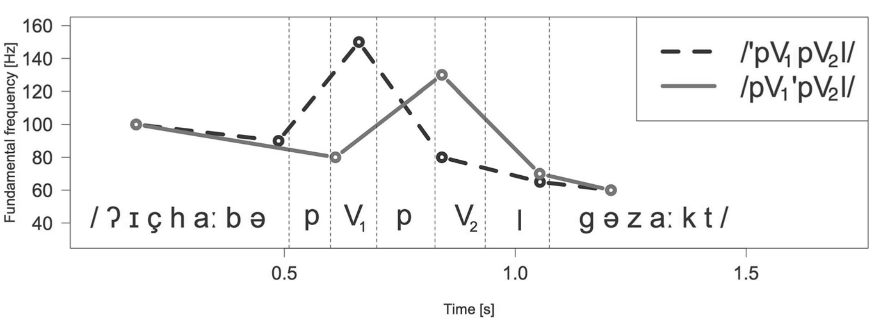

FIG. 5. Stylized fundamental frequency contours for the utterances "Ich habe pV_1pV_2l gesagt" with trochaic stress pattern (/ˈpV_1pV_2l/, black, dashed) and with iambic stress pattern (/pV_1ˈpV_2l/, gray, solid) in which V_1 was a token from an /ʊ-ʏ/ continuum and in which V_2 was /e/ or /o/. The dashed vertical lines are the acoustic segment boundaries, which remained unchanged under all conditions.

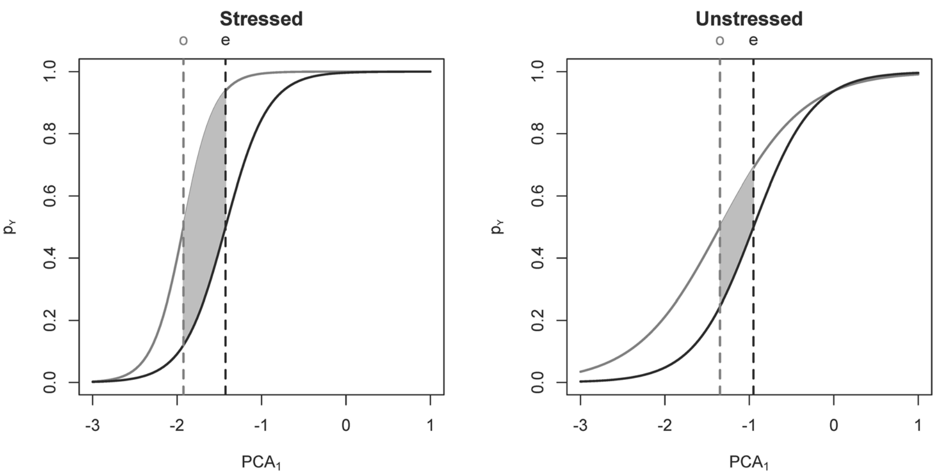

FIG. 6. Psychometric functions fitted to listeners' /ʊ-ʏ/ categorizations in the context of V_2 = /o/ (gray) and V_2 = /e/ (black) shown separately for (left) stressed and (right) unstressed V_1 as a function of PCA_1 calculated at each of the 11 stimulus steps. The 50% category boundaries in the two V_2 contexts are shown as vertical dashed lines. The gray shaded region is the area extending vertically between the sigmoids and horizontally between the two decision boundaries in the V_2 contexts.

the production data below were obtained for x in (4) based on PCA-transformed data only.

A calculation was made of the magnitude of the V_2-on-V_1 coarticulatory influence in speech production and its relationship to the size of the V_2-on-V_1 effect in speech perception. This calculation was carried out separately for each subject, S, and separately for each of the two prosodic contexts, $P = \{\text{stressed, unstressed}\}$. There were three steps in this calculation. First, the magnitude of the V_2-on-V_1 coarticulatory influence in speech production, $prod_{S.P}$ was calculated from

$$prod_{S.P} = \bar{u}_{S.P.e} - \bar{u}_{S.P.o} + \bar{y}_{S.P.e} - \bar{y}_{S.P.o}, \tag{5}$$

in which $\bar{u}_{S.P.e}$ is the mean of PCA_1 (extracted at the vowel's acoustic temporal midpoint) across all of subject S's repetitions of V_1 = /ʊ/ in prosodic context P and in the context V_2 = /e/; and in which $\bar{u}_{S.P.o}$ is the mean PCA_1 calculated across repetitions by the same subject in the same prosodic context but in the context V_2 = /o/ ($\bar{y}_{S.P.e}$ and $\bar{y}_{S.P.o}$ are analogous quantities but for V_1 = /ʏ/). If V_2 has no influence on V_1 then $prod_{S.P}$ is zero; larger positive shifts of V_1 in a V_2 = /e/ context and larger negative shifts of V_1 in a V_2 = /o/ context (see Fig. 3) result in larger values on $prod_{S.P}$. Second, the size of the V_2-on-V_1 effect in speech perception was calculated from $perc_{S.P}$:

$$\begin{aligned} perc_{S.P} = &\int_{b_{S.P.o}}^{b_{S.P.e}} \frac{e^{m_{S.P.o}x + k_{S.P.o}}}{1 + e^{m_{S.P.o}x + k_{S.P.o}}}\,dx \\ &- \int_{b_{S.P.o}}^{b_{S.P.e}} \frac{e^{m_{S.P.e}x + k_{S.P.e}}}{1 + e^{m_{S.P.e}x + k_{S.P.e}}}\,dx, \end{aligned} \tag{6}$$

in which $b_{S.P.e}$, $m_{S.P.e}$, and $k_{S.P.e}$ are, respectively, the decision boundary, slope, and intercept for subject S and prosodic context P in the V_2 = /e/ context and in which x has the same definition as in (4) ($b_{S.P.o}$, $m_{S.P.o}$, and $k_{S.P.o}$ are analogous variables in the V_2 = /o/ context). The calculation in (6) is of the shaded areas shown in Fig. 6 for the stressed and unstressed contexts, but carried out separately for each subject. Finally, the ratio between these quantities in production and perception was calculated from

$$R_{S.P} = perc_{S.P} / prod_{S.P} \tag{7}$$

If a subject normalizes minimally for the contextual effects of V_2 in perception in relation to the coarticulatory V_2-on-V_1 influence in the same subject's speech production (and same prosodic context), then $R_{S.P}$ has a low value. The hypothesis to be tested was that a subject's perceptual normalization for context in relation to the same subject's magnitude of coarticulation in production was less in the unstressed than in the stressed context, i.e., that $R_{S.\text{Unstressed}} < R_{S.\text{Stressed}}$.

B. Results

The data from 2/20 listeners who responded with /ʏ/ to almost all stimuli were discarded; the results reported below are based on those of the remaining 18 subjects.

The psychometric functions in Fig. 6 show that listeners' categorizations were affected by V_2-context: that is, there were more /ʏ/ responses in a V_2 = /o/ than in a V_2 = /e/ context. This result is consistent with the finding of a V_2-dependent difference in the decision boundaries in speech production (Figs. 3 and 4). The category boundaries in perception were also shifted to the right in the unstressed vs stressed context, i.e., towards stimuli associated with a higher F2: Thus, even though the acoustic stimuli presented to the listeners were identical across the two stress conditions, listeners nevertheless labeled a greater proportion of these unstressed stimuli as /ʊ/ commensurate with the shift of the decision boundaries in speech production into the /ʏ/-space (Figs. 3 and 4). The results of a repeated measures ANOVA with the categorical /ʊ-ʏ/ boundaries (the vertical lines of Fig. 6) as the dependent variable, and with Stress (2 levels) and V_2-context (2 levels) as independent factors showed a significant effect of V_2 context [$F(1,17) = 50.1$, $p < 0.001$] and a significant shift of the category boundaries towards the /ʏ/-end of the continuum in the unstressed compared with the stressed context [$F(1,17) = 28.6$, $p < 0.001$]. There was no significant interaction between these factors which shows that the perceptual category boundaries in the two V_2-contexts were separated roughly to the same extent in the unstressed and stressed conditions.

Figure 6 shows that the psychometric function in the V_2 = /o/ context was flatter in the unstressed than in the stressed condition just as the fitted sigmoids had been in speech production (Fig. 4). A repeated measures ANOVA with the slope of the psychometric function as the dependent variable and with stress (2 levels) and V_2-context (2 levels) as independent factors showed a significant effect on slope of stress [$F(1,17) = 60.4$, $p < 0.001$], but no overall main effect of V_2 and a significant interaction between these factors [$F(1,17) = 17.7$, $p < 0.001$]. The *post hoc*, Bonferroni corrected *t*-tests showed steeper slopes in the stressed than in the unstressed context in both the V2 = /e/ ($t_{17} = 3.6$, $p_{adj} < 0.05$) and in the V2 = /o/ ($t_{17} = 7.4$, $p_{adj} < 0.001$) contexts. The interaction comes about because, whereas the slope was significantly steeper in the stressed /o/ vs stressed /e/ context ($t_{17} = 3.2$, $p_{adj} < 0.05$), it was significantly shallower in the unstressed /o/ vs unstressed /e/ context ($t_{17} = 3.3$, $p_{adj} < 0.05$). As far as the effects on slope of stress are concerned, these results from speech perception were quite well matched with speech production which, for the PCA data (lower panel Fig. 4) shows shallower slopes for unstressed vs stressed in both V_2 contexts.

The flatter slope in the V_2 = /o/ contexts was one of the main factors that contributed to the increase in the overlap between the psychometric functions in the unstressed context: this finding also matches the results from speech production in which the derived sigmoid functions were closer together in the /ʏ/-vowel space in the unstressed context (Fig. 3). The overlap between the psychometric functions was quantified by calculating the shaded areas shown in Fig. 6 but separately for each subject using (6). The results in Fig. 7 show that this area was greater for stressed than unstressed contexts. The results of a paired-sample Wilcoxon signed rank test with the area as the dependent variable and with stress (2 levels) as the independent factor showed that the area between the psychometric curves in the two V_2 = /e, o/ contexts between the decision boundaries was smaller in the unstressed than in the stressed context ($V = 168$, $p < 0.001$).

Finally, the areas derived from (6) in perception were compared with coarticulatory influence of V_2-on-V_1 influence in speech production. The hypothesis to be tested was that $R_{S.P}$, the ratio of the V_2-on-V_1 influence in perception to the coarticulatory V_2-on-V_1 in production, was less in the unstressed than in the stressed context. The results in Fig. 8 of the calculation of $R_{S.P}$ from (7) shows that this was the case. Compatibly, a paired sample Wilcoxon signed rank test showed that this ratio was significantly smaller in the unstressed than the stressed context ($V = 145$, $p < 0.01$). This result shows, therefore, that the size of the V_2-on-V_1 influence in perception in relation to the magnitude of the V_2-on-V_1 in production was less in unstressed than in stressed syllables.

IV. GENERAL DISCUSSION

The present study has been concerned with the influence of prosodic weakening on the size of anticipatory V_2-on-V_1 coarticulation from two different, but related perspectives: The coarticulatory V_2 influences that cause a shift in the position of the tongue resulting in acoustic differences in V_1; and phonological categorization. As far as the first of these is concerned, the V_2 = /o/ context exerted a coarticulatory backing influence on /ʏ/; but there was no such V_2-dependent coarticulatory shift in /ʊ/ and the magnitude of V_2-on-/ʏ/ coarticulation was found to be unaffected by stress. Phonological categorization was measured in this study both by fitting Gaussian models to the production data and from forced-choice perception experiments carried out on the same listeners who had participated as speakers in the production study. These categorizations whether obtained from Gaussian models fitted to the production data or from listener classifications were shown to be context-dependent: in both cases, the probability of classifying V_1 as /ʏ/ as opposed to /ʊ/ was greater in the context of V_2 = /o/ than it was for V_2 = /e/. For perception, this finding demonstrates compensation for coarticulation: Listeners factor out the coarticulatory variation from the signal and attribute it to the source from which it originates, in this case V_2.

In the prosodically weak context, /ʊ/ was closer in production to the acoustically defined /ʏ/-target and was more

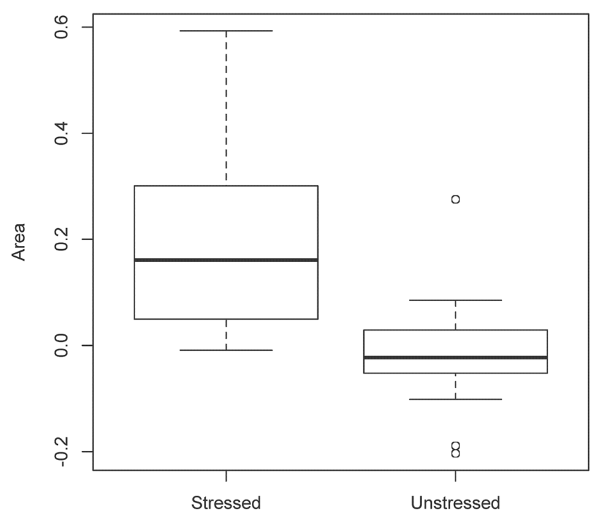

FIG. 7. Distribution of the area between the psychometric functions in the two V_2 contexts calculated between the two decision boundaries [cf. the formula in (6) and the shaded area in Fig. 6] shown separately for stressed and unstressed contexts (each boxplot includes one data point per subject).

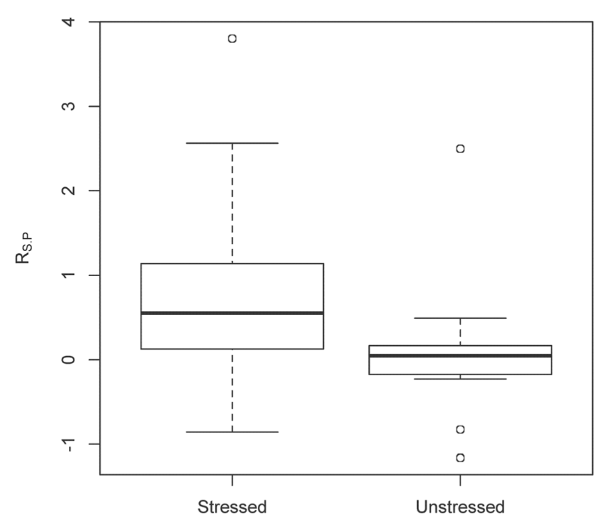

FIG. 8. Distribution of $R_{S.P}$, the ratio of the V_2-on-V_1 influence in perceptual categorization to coarticulation in production [cf. formula (7)] shown separately for stressed and unstressed contexts (each boxplot includes one data point per subject).

variable. These stress-dependent differences had three effects on phonological /ʊ-ʏ/ categorization derived by fitting Gaussian models to the production data: The decision boundaries were shifted into the /ʏ/ space in the unstressed context; the /ʊ-ʏ/ decisions were probabilistically more ambiguous, as shown by the flatter sigmoid functions (Fig. 4) compared with those from the stressed context; and finally the influence of V_2-context on the categorization in production was weaker as shown by the smaller areas between the sigmoid functions, than it was for stressed syllables in the /ʏ/ space, i.e., in just the region in which V_2-on-V_1 coarticulation was most marked in the production data. Listeners were evidently sensitive to these prosodic influences because these categorization differences derived from fitting Gaussians to their production data were also in evidence in their perceptual responses: thus, in the unstressed context in which V_1 had no pitch-accent, listeners' category boundaries were shifted towards /ʏ/ (i.e., there were more /ʊ/ responses), their psychometric curves were flatter, and the influence of V_2 on their categorizations towards the /ʏ/-end of the continuum was significantly diminished compared with their responses in the stressed context. In general, these results provide further evidence for the considerable amount of phonetic variation to which listeners are sensitive (Local, 2003): Listeners evidently have knowledge not only of segmental context effects, as numerous studies on the compensation for coarticulation have shown (Lindblom and Studdert-Kennedy, 1967; Mann and Repp, 1980; Ohala and Feder, 1994), but in addition and as the results of the present study show, listeners' responses are also biased by the expected influences of prosodic weakening on vowels in speech production. These results demonstrate furthermore the very close connection between speech production and perception (Fowler *et al.*, 2003; Fowler, 2005) as far as contextual influences, both segmental and prosodic, are concerned.

At the same time, the results point to a different association between coarticulation and categorization across the two prosodic contexts: Although the coarticulatory influence of V_2-on-/ʏ/ was unchanged from stressed to unstressed syllables, the influence of V_2-context on /ʊ-ʏ/ categorization was reduced in the unstressed context. The smaller difference in the /ʊ-ʏ/ perceptual categorizations between the two V_2 contexts in the unstressed condition suggests, in turn, that listeners compensated less, or were more uncertain about compensating for, V_2-dependent coarticulation in an unstressed context, even though the magnitude of V_2-on-V_1 coarticulation in production was unaffected by stress. This weakening in listeners' compensation for coarticulation is suggested as a potential source of sound change: If compensation for coarticulation is weakened in an unstressed context but the magnitude of actual coarticulation is unchanged, then some of the coarticulatory variation could be parsed with V_1 and not with the V_2 coarticulatory source from which it originates thereby leading to sound change (Beddor, 2009; Ohala, 1993), potentially resulting in umlaut.

The cause of the weakening of V_2-context on phonological categorization in /ʊ-ʏ/ categorizations was /ʊ/-undershoot: More specifically, the greater variability of /ʊ/ in unstressed syllables caused a flattening of the psychometric functions in the two V_2 contexts which in turn diminished their separation in just the part of the /ʏ/ vowel space in which coarticulation was shown to occur in production. This result establishes a relationship between models in which sound change is driven by reduction (Bybee, 2002; Mowrey and Pagliuca, 1995) and those such as Ohala's (1993) in which the conditions for sound change to take place are explained in terms of the perceptual (mis)parsing of coarticulation. This result also shows that phonetic reduction in one part of the vowel system (i.e., /ʊ/ undershoot) can have consequences for how much coarticulation is attributed to the source in a different part of the vowel space (i.e., how much V_2-dependent coarticulation is parsed from /ʏ/): That is, sound change in one part of the vowel space may develop indirectly as an emergent property of phonetically driven undershoot in another. This conclusion is consistent with one of the tenets of Ohala's (1993) model that much of sound change is non-teleological: It seems unlikely that /ʊ/-undershoot would be planned in order to set the conditions for a V_2-dependent sound change to take place in /ʏ/. At the same time, these results suggest a further generalization beyond Ohala's (1993) model and the apparent-time studies in Harrington *et al.* (2008) and Kleber *et al.* (2012) that are based on it. In those models, the conditions for sound change are given when coarticulation in production is mis-parsed in perception. The more general interpretation suggested by the present results is that the origin of sound change is due to a mismatch between coarticulation in speech production and context-dependent phonological categorization. In a stable system, coarticulation in production and categorization are related in that the effect of context on category boundaries (e.g., on different cut-off points between /sV − ʃV/ depending on rounding of the following vowel—Mann and Repp, 1980) and the size of the coarticulatory influence in production are closely linked. However, it is when other phonetic processes such as vowel undershoot leave coarticulatory relationships in production unchanged but nevertheless bring about a change in the category boundaries that the system can become unstable and prone to change.

The final issue to be considered is why V_2-on-V_1 coarticulation should affect $V_1 =$ /ʏ/ but not $V_1 =$ /ʊ/ and whether this finding can be related to the so-called secondary umlaut by which V_1 back vowels fronted diachronically in the context of a front V_2 in West Germanic (e.g., Old High German *turi*, *wurfil*, *oli* > Middle High German *tür*, *würfel*, *öl* —see, e.g., Howell and Salmons, 1997; Iverson *et al.*, 1994; Schulze, 2010). As far as the first of these is concerned, there is evidence (see, e.g., Cole *et al.*, 2010) that trans-consonantal vowel coarticulation can be masked by the place of articulation of the intervening consonant: thus the low F2 labial locus and the low F2 /ʊ/-target might in combination be sufficient to oppose the potential fronting effect induced by the high F2 target in /e/. But on the other hand the /ʏ/'s high F2 might be more easily shifted due to the combined F2-lowering effects of the labial consonant and of /o/'s low F2. These synchronic differences in which /ʏ/ is affected by V_2 to a greater extent than /ʊ/ across an intervening labial consonant can be related diachronically to the greater resistance to umlaut when V_1 and the intervening consonant share several features: that is, there may be an association between the synchronically marginal $V_2 = /e, o/$ influences on $V_1 = /ʊ/$ and the evidence discussed in Howell and Salmons (1997) that secondary umlaut in /u, i/ contexts is often blocked across intervening labial geminates (e.g., Standard German *hüpfen*, "to jump," with umlaut, but Alemannic *hupfe* without).

In conclusion, prosodic weakening can bring about vowel undershoot in one part of the vowel system that can indirectly diminish the influence of V_2-context on vowel categorization in another part of the vowel space. If, as for the present results, the influence of context on categorization is reduced in relation to its influence on coarticulation in speech production, then the conditions for sound change to take place may be met. The present results therefore shift an explanation for the origin of sound change away from a mismatch between coarticulation in production and perception to one in which coarticulation in production may become misaligned with phonological categorization. Finally, the present results may also explain why phonetic reduction (due in this case to stress) may precipitate sound change, even though the outcome of the sound change itself is not necessarily reductive (in the sense that V_2-on-V_1 coarticulation does not entail vowel centralization).

At the same time, caution is warranted in generalizing from these results given that the speech material is limited and because the disyllabic words containing a sequence of a lax followed by a tense vowel are undoubtedly quite rare in German. Further studies of how different types of coarticulation vary in production and perception in different languages are needed to strengthen these conclusions about the relationship between prosodic weakening and sound change.

ACKNOWLEDGMENTS

The authors thank Carolin Sabath and Jessica Siddins for recruiting and recording the subjects, for segmentation and labeling of the audio data and for the preparation of the stimuli used in the perception experiment. This research was supported by German Research Council Grant No. HA3512/9-1 "The relationship between coarticulation, prosodic weakening, and sound change."

[1]The model in R that was used for this purpose was model = lmer(response ~ x + (1 + x | L), family = binomial) of the lme4 package in which response was the listener's forced choice response (/ʊ/ or /ʏ/), x one of the 11 PCA-transformed values of the synthetic continuum in (4), and in which L was the listener. The population intercept, k, and population slope, m, are given in R by fixef(model); the listener-specific intercept, k_S, and listener-specific slope, m_S, are given by coef(model).

Alfonso, P. J., and Baer, T. (**1982**). "Dynamics of vowel articulation," Lang. Speech **25**, 151–173.

Beckman, M. E., De Jong, K., Jun, S. A., and Lee, S. H. (**1992**). "The interaction of coarticulation and prosody in sound change," Lang. Speech **35**, 45–58.

Beddor, P. S. (**2009**). "A coarticulatory path to sound change," Language **85**, 785–821.

Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (**2002**). "Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates," J. Phonetics **30**, 591–627.

Boersma, P., and Weenink, D. (**2011**). *Praat: Doing phonetics by computer* [Computer program], version 5.2.40.

Bybee, J. (**2002**). "Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change," Lang. Variation Change **14**, 261–290.

Cho, T. (**2004**). "Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English," J. Phonetics **32**, 141–176.

Cole, J., Linebaugh, G., Munson, C., and McMurray, B. (**2010**). "Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach," J. Phonetics **38**, 167–184.

Draxler, C., and Jänsch, K. (**2004**). "SpeechRecorder – A universal platform independent multi-channel audio recording software," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Lisbon, Portugal (European Language Resources Association, Paris, France), pp. 559–562.

Fant, G. (**1973**). *Speech Sounds and Features* (MIT Press, Cambridge, MA), Chap. 3, pp. 32–84.

Fowler, C. A. (**1981**). "Production and perception of coarticulation among stressed and unstressed vowels," J. Speech Hear. Res. **24**, 127–139.

Fowler, C. A. (**2005**). "Parsing coarticulated speech in perception: Effects of coarticulation resistance," J. Phonetics **33**, 199–213.

Fowler, C. A., Brown, J. M., Sabadini, L., and Weihing, J. (**2003**). "Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks," J. Memory Lang. **49**, 396–413.

Fowler, C. A., and Smith, M. R. (**1986**). "Speech perception as 'vector analysis': An approach to the problems of invariance and segmentation," in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. Klatt (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 123–139.

Fujisaka, H., and Kunisaki, O. (**1976**). "Analysis, recognition and perception of voiceless fricative consonants in Japanese," in *Annual Bulletin Research Institute of Logopedics and Phoniatrics* (Faculty of Medicine, University of Tokyo, Tokyo), Vol. 10, pp. 145–156.

Harrington, J., Fletcher, J., and Beckman, M. E. (**2000**). "Manner and place conflicts in the articulation of accent in Australian English," in *Papers in Laboratory Phonology, Acquisition and the Lexicon*, edited by M. B. Broe and J. B. Pierrehumbert (Cambridge University Press, Cambridge, UK), Vol. V, pp. 40–51.

Harrington, J., Kleber, F., and Reubold, U. (**2008**). "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," J. Acoust. Soc. Am. **123**, 2825–2835.

Hombert, J. M., Ohala, J. J., and Ewan, W. G. (**1979**). "Phonetic explanations for the development of tones," Language **55**, 37–58.

Howell, R. B., and Salmons, J. C. (**1997**). "Umlautless residues in Germanic," J. Germanic Ling. **9**, 83–111.

Hualde, J. I., Simonet, M., and Nadeu, M. (**2011**). "Consonant lenition and phonological recategorization," Lab. Phonol. **2**, 301–329.

Huffman, M. K. (**2012**). "Modulation of speech gestures through prosody or sound change: A commentary," Lab. Phonol. **3**, 27–36.

Iverson, G. K., Davis, G. W., and Salmons, J. C. (**1994**). "Blocking environments in Old High German umlaut," Folia Ling. Hist. **15**, 131–148.

Jannedy, S., and Hay, J. (**2006**). "Modelling sociophonetic variation," J. Phonetics **34**, 405–408.

Johnson, K. (**1997**). "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Morgan Kaufmann Publishers, San Francisco, CA), pp. 145–165.

Kawahara, H., Takahashi, T., Morise, M., and Banno, H. (**2009**). "Development of exploratory research tools based on TANDEM-STRAIGHT," in *Proceedings of the Annual Summit and Conference of the Asia-Pacific Signal and Information Processing Association 2009* (Asia-Pacific Signal and Information Processing Association, Hong Kong), pp. 111–120.

Kleber, F., Harrington, J., and Reubold, U. (**2012**). "The relationship between the perception and production of coarticulation during a sound change in progress," Lang. Speech **55**, 383–405.

Kohler, K. J. (**2001**). "Articulatory dynamics of vowels and consonants in speech communication," J. Int. Phonetic Assoc. **31**, 1–16.

Labov, W. (**1994**). *Principles of Linguistic Change, Internal Factors* (Blackwell Publishing, Oxford, UK), Vol. 1, pp. 1–384.

Labov, W. (**2001**). *Principles of linguistic change, Social Factors* (Blackwell Publishing, Oxford, UK), Vol. 2, pp. 1–596.

Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.

Lindblom, B., Guion, S., Hura, S., Moon, S. J., and Willerman, R. (**1995**). "Is sound change adaptive?," Riv. Ling. **7**, 5–36.

Lindblom, B., and MacNeilage, P. (**2011**). "Coarticulation: A universal phonetic phenomenon with roots in deep time," *Speech, Music and Hearing—Quarterly Progress and Status Report* (Kungliga Tekniska högskolan, Stockholm, Sweden), Vol. 51, pp. 41–44.

Lindblom, B., and Studdert-Kennedy, M. (**1967**). "On the role of formant transitions in vowel recognition," J. Acoust. Soc. Am. **42**, 830–843.

Local, J. (**2003**). "Variable domains and variable relevance: interpreting phonetic exponents," J. Phonetics **31**, 321–339.

Mann, V. A., and Repp, B. H. (**1980**). "Influence of vocalic context on perception of the [ʃ]-[s] distinction," Percept. Psychophys. **28**, 213–228.

Manuel, S. (**1999**). "Cross-language studies: Relating language-particular coarticulation patterns to other language-particular facts," in *Coarticulation: Theory, Data and Techniques*, edited by W. J. Hardcastle and N. Hewlett (Cambridge University Press, Cambridge, UK), pp. 179–198.

Martin, J. G., and Bunnell, H. T. (**1982**). "Perception of anticipatory coarticulation effects in vowel–stop consonant–vowel sequences," J. Exp. Psych. **8**, 473–488.

Milner, B., and Shao, X. (**2006**). "Clean speech reconstruction from MFCC vectors and fundamental frequency using an integrated front-end," Speech Commun. **48**, 697–715.

Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**, 453–467.

Mooshammer, C., and Geng, C. (**2008**). "Acoustic and articulatory manifestations of vowel reduction in German," J. Int. Phonetic Assoc. **38**, 117–136.

Mowrey, R., and Pagliuca, W. (**1995**). "The reductive character of articulatory evolution," Riv. Ling. **7**, 37–124.

Nossair, Z. B., and Zahorian, S. A. (**1991**). "Dynamic spectral shape features as acoustic correlates for initial stop consonants," J. Acoust. Soc. Am. **89**, 2978–2991.

Ohala, J. J. (**1981**). "The listener as a source of sound change," in *Papers from the Parasession on Language and Behavior*, edited by C. S. Masek, R. A. Hendrick, and M. F. Miller (Chicago Linguistic Society, Chicago, IL), pp. 178–203.

Ohala, J. J. (**1993**). "Sound change as nature's speech perception experiment," Speech Communication **13**, 155–161.

Ohala, J. J., and Feder, D. (**1994**). "Listeners' normalization of vowel quality is influenced by restored consonantal context," Phonetica **51**, 111–118.

Öhman, S. E. G. (**1966**). "Coarticulation in VCV utterances: spectrographic measurements," J. Acoust. Soc. Am. **39**, 151–168.

Pierrehumbert, J. B. (**2003**). "Phonetic diversity, statistical learning, and acquisition of phonology," Lang. Speech **46**, 115–154.

Pouplier, M. (**2012**). "The gestural approach to syllable structure: Universal, language- and cluster-specific aspects," in *Speech Planning and Dynamics, Speech Production and Perception*, edited by S. Fuchs, M. Weihrich, D. Pape, and P. Perrier (Peter Lang GmbH, Frankfurt am Main), Vol. 1, pp. 63–96.

Schiel, F. (**2004**). "MAuS goes iterative," *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Lisbon, Portugal (European Language Resources Association, Paris, France), pp. 1015–1018.

Schulze, J. H. (**2010**). "Der i-Umlaut im Althochdeutschen. Theorie, Phonetik und Typologie sowie eine optimalitätstheoretische Analyse," in *Bamberger Beiträge zur Linguistik* ("The i-umlaut in Old High German. Theory, phonetics and typology as well as an optimality theoretic analysis," in *Bamberg Contributions to Linguistics*), edited by T. Becker, M. Haase, S. Kempgen, M. Krug, and P. Noel Aziz Hanna (University of Bamberg Press, Bamberg, Germany), Vol. 3, pp. 7–208.

Solé, M.-J. (**2012**). "Natural and unnatural patterns of sound change?," in *The Initiation of Sound Change: Perception, Production, and Social Factors*, edited by M.-J. Solé and D. Recasens (John Benjamins Publishing, London), pp. 123–146.

Stevens, K. N., and House, A. S. (**1963**). "Perturbation of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**, 111–128.

van Bergem, D. R. (**1994**). "A model of coarticulatory effects on the schwa," Speech Commun. **14**, 143–162.

Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.